Psychology of Language and Communication 2024, Vol. 28, No. 1

DOI: 10.58734/plc-2024-0001

Adriana Biedroń[1], Sławomir Pasikowski[2], and Mirosław Pawlak[3]
[1] English studies, Pomeranian University in Słupsk, Poland
[2] Faculty of Educational Sciences, University of Lodz, Poland
[3] English studies, Adam Mickiewicz University, Kalisz, Poland

Working Memory Capacity, TL Grammar Attainment and Length of Study as Predictors of Explicit and Implicit (Automatized) Knowledge of English Passive Voice

This study investigated the relationship between phonological short-term memory (PSTM), working memory capacity (WMC), and receptive and productive dimensions of explicit and implicit (automatized) knowledge of English passive voice, also taking into account the effect of grammar attainment and self-reported length of study. Participants were 152 Polish university students majoring in English. Two measures of PSTM and WMC were applied. Receptive and productive explicit knowledge were measured by means of an untimed grammaticality judgment test and a test requiring the provision of correct verb forms, respectively. Receptive implicit (automatized) knowledge was assessed with a timed grammaticality judgment test while its productive dimension was assessed through a focused communication task. Canonical correlation for the entire model was rather high, which means that the original variables were strongly related to each other. However, finer-grained analyses showed that it was primarily overall grammar attainment, and, to a lesser extent, WMC that determined the levels of explicit and implicit (automatized) grammar knowledge.

Key words: phonological short-term memory; working memory capacity; implicit (automatized) knowledge; explicit knowledge; grammar attainment

Address for correspondence: Adriana Biedroń

Pomeranian University in Słupsk, English Studies, Arciszewskiego 22a, Słupsk 76-200 Poland.

E-mail: adriana.biedron@upsl.edu.pl

Predictive studies into the relationship between working memory (WM) and grammar learning in a second or foreign language (L2) have demonstrated that both phonological short-term memory (PSTM) and working memory capacity (WMC) are involved in this process. However, it is still not clear which of these components facilitates the development of explicit and implicit (automatized) L2 knowledge. This is partly because many studies fail to distinguish the two types of knowledge, which makes this issue difficult to resolve. For example, Wen and Li (2019) claim that knowledge learned through WM, whether PSTM or WMC, is perhaps partly explicit and partly implicit, although PSTM is more likely to result in implicit knowledge because it involves the ability to internalize linguistic input as unanalyzed chunks. Sagarra (2017) underlined the impact of WM on both explicit and implicit tasks, but only when WM tests involve the processing component, that is, time pressure. Most of the studies examining the relationship between WM and L2 grammar are (quasi-)experimental and include some kind of pedagogical intervention. Despite their pedagogical value, the problem of which WM subsystem affects which type of grammar knowledge remains unresolved. The current study contributes to this line of inquiry by examining the role of PSTM and WMC in the development of productive and receptive dimensions of explicit and implicit (automatized) knowledge of the English passive voice (DeKeyser, 2009; Ellis, 2009). It should also be stressed at the outset that the labels explicit and implicit are not employed in the strict psycholinguistic sense manifesting awareness of the target language features being used, but in a less absolute sense of being able to access them in real-time processing (Ellis, 2005). In other words, following DeKeyser (2017), implicit knowledge is equated here with highly automatized knowledge from which it is functionally indistinguishable. Thus, such knowledge can be drawn upon under time pressure and other conditions which make conscious reliance on rules difficult. For this reason, the term implicit is modified by the parenthesized term automatized in connection with the empirical investigation reported below.

Explicit And Implicit L2 Knowledge

The distinction between explicit and implicit L2 knowledge is crucial to theorizing about and researching second language acquisition (SLA). Following Ellis (2004, 2006, 2009), explicit knowledge is seen as conscious, declarative, and unconstrained by developmental readiness (Pienemann & Lenzing, 2015) or the learner's age. It lends itself to verbalization, but at the same time, it can only be accessed when sufficient time is provided, it lacks precision, and it is often inaccurate. Implicit knowledge can be characterized as tacit, intuitive and procedural. It can be easily employed in real-time processing and is indispensable in spontaneous communication. It is variable and systematic, but it is not available for self-report and its development is subject to constraints stemming from processing operations and age. The distinct nature of explicit and implicit knowledge finds support in empirical evidence coming from neurological research, according to which they are stored in different areas of the human brain (e.g., Paradis, 2009; Ullman, 2015). Nonetheless, there is no consensus regarding the true nature of the two types of linguistic representation. DeKeyser (2009) argues that explicit and implicit knowledge should not be seen as equivalents of declarative knowledge and procedural knowledge because, for instance, procedural knowledge that has not been sufficiently automatized may still require application of rules. He also points out that entirely implicit knowledge is unlikely to develop in learners who are past the critical period and have scarce access to the target language, a situation that is the norm in most foreign language contexts (DeKeyser, 2009; DeKeyser & Juffs, 2005). In such conditions, even advanced learners who have no difficulty engaging in spontaneous communication can easily invoke the relevant rules if needed. For this reason, even though the two constructs are clearly distinct (cf. Suzuki & DeKeyser, 2017), it makes more sense in such cases to opt for the term highly automatized explicit knowledge, which is "(…) unintentional, uncontrollable, unconscious, efficient, and fast" (DeKeyser, 2017, p. 17), thereby allowing accurate and fluent L2 performance. Another highly controversial but also consequential issue is the relationship between explicit and implicit knowledge. While some scholars (e.g., Krashen, 1981) embrace a noninterface position denying any contribution of conscious awareness to the growth of implicit knowledge, others, in particular DeKeyser (2017), take the opposite stance, claiming that declarative knowledge can enable parallel development of procedural knowledge which can be subject to gradual automatization through practice involving real operating conditions. This position seems to be fully warranted if we consider the fact that L2 instruction in most contexts starts with some kind of conscious knowledge of how the target language works and such practices have proven effective for many thousands of learners (cf. Pawlak, 2019).

Given the importance of the constructs of explicit and implicit L2 knowledge from a theoretical perspective, it is of utmost significance to develop valid and reliable measures of them. One such attempt was made by Ellis et al. (2009), who adopted seven criteria when designing tests of explicit and implicit L2 knowledge: degree of awareness, time available, focus of attention, systematicity, certainty, metalanguage, and learnability. Drawing on these criteria, they measured explicit knowledge by means of (a) an untimed grammaticality judgment test, which required learners to decide on the correctness of a sentence, indicate how certain they were about their decision, and state whether they relied on intuition or rule knowledge, and (b) an untimed metalinguistic knowledge test, where participants were asked to explain errors in sentences on the basis of rules, find L2 grammar structures, and label different parts of a sentence. They tapped into implicit knowledge using three measures: a timed grammaticality judgment test with the amount of time adjusted to sentence length, an oral elicited imitation test in which learners were instructed to agree or disagree with a sentence they heard and subsequently repeat it using correct English, and an oral narrative task where participants had to retell a story on reading it twice. The measures included in the battery manifested high levels of validity and reliability, and they have also been validated for languages other than English (Gutiérrez, 2013; Zhang, 2015). However, these measures are not free from limitations. These concern, for example, the potential role of moderating variables, such as WM or the complexity of the targeted features, the difficulty involved in administering some tests, in particular elicited oral imitation, or challenges involved in developing pure measures of explicit and implicit knowledge that would not be affected by other types of representation (cf. Bielak & Pawlak, 2013; DeKeyser, 2009; Suzuki & DeKeyser, 2017). One solution to such problems could involve devising increasingly sensitive and sophisticated tests of the two types of L2 representation, a good case in point being the study by Suzuki and DeKeyser (2017) in which eye-tracking and reaction time were used to tease apart implicit knowledge and highly automatized knowledge. An equally viable option, though, would be to admit that such fine-grained distinctions, while insightful for theorizing about SLA, may be of limited relevance for pedagogically-oriented research. This is because such research typically aims to determine whether learners can use the targeted structures not only when they have ample time to resort to pertinent rules (e.g., in exercises involving paraphrasing), which is largely reflective of explicit knowledge, but also in real-time processing needed in interactive communication which offers insights into implicit or at least highly automatized knowledge. In line with this rationale, carefully constructed traditional grammar tests may well serve the purpose of tapping into explicit knowledge since they allow ample time to fall back on rules. In the same vein, a sufficient measure of implicit or highly automatized knowledge could be performance on focused communication tasks (Ellis, 2003) which require the use of a given grammar feature for their completion. Although designing such tasks poses daunting challenges in the case of some grammar structures, such problems are more easily overcome than trying to use sophisticated hardware in the classroom, and they are also characterized by higher ecological validity (cf. Pawlak, 2019). This was the approach that was embraced in the current study.

WM And Implicit And Explicit Grammar Knowledge

Working memory is defined as the limited memory capacity to temporarily store, process and maintain a restricted amount of information while performing mentally demanding tasks (Wen & Li, 2019). Two components of WM, namely, the phonological loop and the central executive are most relevant to L2 learning. The former, also named phonological short-term memory (PSTM), is responsible for momentarily holding sound-based information through the process of articulatory rehearsal, which helps in the learning of the phonological forms of new words (Baddeley et al., 1998). It is usually measured with simple span tasks, such as the digit span or the nonword span. The latter, also termed as working memory capacity (WMC), performs executive functions and is usually measured with complex span tasks, such as reading, listening, speaking, or operation span, which involve simultaneous processing of information. In line with Kane et al. (2004), WMC tasks reflect a domain-general factor which results in the broad predictive utility of WM span measures whereas STM tasks, based on the same stimuli as the WMC tasks, are much more domain specific. Second language acquisition research (e.g., Li et al., 2019; Linck et al., 2014; Wen, 2019; Wen & Li, 2019) has provided evidence that both verbal components of WM, namely, PSTM and WMC, affect various aspects of L2 learning at different ages and proficiency levels. Thus far, two meta-analyses have been conducted on associations between WM and aspects of the acquisition and processing of an additional language. Linck et al. (2014) analyzed 748 correlation coefficients reported in 79 studies and found a weak correlation between both simple and complex measures of WM and L2 proficiency (r = .25). In addition, WMC was found to be a better predictor in this respect than PSTM (r = .27 and r = .17, respectively) whereas verbal measures demonstrated stronger predictive power (r = .26) than nonverbal measures (r = .18). The meta-analysis by Li (2019) corroborated the overall weak predictive power of WM in SLA. Li took into consideration the process and product aspects of interaction in L2. The correlation with the process aspect, which refers to such factors as noticing of corrective feedback and producing modified output, turned out to be variable and inconsistent. On the other hand, the relationship with the product of interaction, that is, the effects of instructional treatments measured on pretests and post-tests, was weak. It was also revealed that WM was more strongly correlated with the effects of explicit rather than implicit treatments. While WM seems to be a much weaker predictor of SLA than foreign language aptitude (Li, 2019), this factor has come to be equated with language aptitude (Wen et al., 2016) and has been confirmed to be implicated in L2 grammar learning (e.g., Engel de Abreu & Gathercole, 2012; Martin & Ellis, 2012; Li et.al., 2019; Sagarra, 2017). Many studies have convincingly shown that both PSTM and WMC have distinctive effects on some aspects of L2 acquisition, processing, and use, with such effects being themselves moderated by the cognitive demands of the processing component of WM tests, proficiency, and task demands (Sagarra, 2017; Wen, 2016; Wen & Li, 2019). Working memory is also involved in the development of L2 proficiency (Gagné et al., 2022; French, 2006). It seems that the effects of WM are evident on tests of both explicit (Goo, 2012; Li, 2013) and implicit knowledge (Kim et al., 2015; Li, 2013; Li et. al., 2019, Sanz et al., 2014). Révész (2012) suggests that PSTM is important for the acquisition of implicit knowledge while WMC is facilitative of the development of explicit knowledge. However, there are also studies which have failed to detect any impact of WM on grammar learning (e.g,, Antoniou et al., 2016; Ettlinger et al., 2014). This section provides a brief overview of studies that have examined the link between PSTM and WMC on the one hand, and implicit and explicit knowledge of L2 grammar, both in terms of comprehension and production, on the other. For reason of space, the synthesis will be confined to most recent empirical investigations that have directly addressed this issue.

Williams and Lovatt (2003) conducted two small-scale laboratory-based experiments with the use of semiartificial microlanguages in which they investigated the relationship between PSTM and the capacity to generalize grammatical gender. They found that PSTM predicted grammatical generalization abilities at r = .60 and hypothesized that grammar rules represent generalizations of patterns across sequences of words. They also claimed that the ability to store morpheme combinations is ultimately related to rule learning because it reflects the ability to successfully encode input sequences. Their general conclusion was that there is a correlation between PSTM and rule learning, even if the targeted rules are quite abstract. Importantly, given the focus of this study, the researchers assumed that the process of rule learning that occurred was explicit because participants who scored 100% on the final generalization test were able to clearly articulate the rules applied. Similar results were reported by French and O'Brien (2008), who examined the role of PSTM in grammar learning of native French-speaking children benefitting from a five-month intensive English program. The researchers tapped into PSTM with a nonword span in the first and last month of instruction. Grammar was measured through a written test focused on grammatical structures that received explicit instruction. They found that PSTM accounted for 28% of variance in grammar scores at the second test, even after controlling for vocabulary knowledge. O'Brien et al. (2006) found significant influences of PSTM on measures of adult L2 learning, including the use of free grammatical morphemes and subordinate clauses in a narrative task, both at the beginning and at the end of a semester-long course in Spanish. In this case, L2 learning was both explicit and implicit, as the learners studied formally at university and communicated with native speakers during study abroad. However, a narrative task requires in most part implicit or highly automatized explicit knowledge. In both studies, the relationships between PSTM and L2 grammar learning were found only for more advanced L2 learners and not beginners. Verhagen et al. (2015) tested the relationships between PSTM and the production of grammatical structures in a narrative task by Turkish child learners of Dutch. Similar to French and O'Brien, the results revealed significant correlations, even after controlling for vocabulary knowledge. However, in contrast to French and O'Brien's study in which the children received explicit instruction, in Verhagen et al.'s study, the Turkish children acquired the L2 in a naturalistic setting, that is, in most part implicitly.

Examining the relationship between WMC and grammar comprehension and production has a long history, although evidence for the existence of such a link is far from clear-cut. In such investigations complex span tests requiring simultaneous processing of information are typically employed. Martin and Ellis (2012) focused on the distinction between PSTM and WMC, and their correlation with vocabulary and grammar learning in an artificial foreign language. Participants were 40 monolingual native English speakers from a large American University. They were not provided with any explicit instruction or explanations. PSTM was measured through nonword repetition and nonword recognition, and WMC was tapped with a listening span task. A number of significant correlations were found between PSTM measures and the final grammar scores, which corroborates prior findings regarding the relationship between PSTM and L2 grammar knowledge. Moreover, WMC was also related to final grammar scores, with the correlations being consistently stronger than for PSTM. The results demonstrated significant effects of PSTM and WMC on grammar learning, some of which were mediated by vocabulary and some of which were independent. It was suggested that grammar learning, contrary to vocabulary learning, requires more processing capacities as well as processing of relevant information both from the input and long-term memory. Therefore, it seems evident that WMC plays a greater role in grammar learning and production than PSTM. In Martin and Ellis's study, the learning process was implicit. However, the researchers assumed that the participants might have attempted to learn explicitly and thus they tested how much explicit knowledge was acquired in the implicit learning condition. Correlations were found between explicit knowledge of rules and final grammar scores, which was interpreted as evidence that PSTM and WM are involved in the emergence of explicit grammar knowledge in implicit learning condition.

Empirical evidence also suggests that the effect of both PSTM and WMC on implicit and explicit grammatical knowledge is moderated by different types of instruction, learning conditions, task demands, cognitive demands of the processing component of WM tests, and L2 proficiency (Sagarra, 2017, Santamaria & Sunderman, 2015; Sanz et al., 2014; Tagarelli et al., 2015; Suzuki & DeKeyser, 2017). For example, Sanz et al. (2014) carried out two experimental studies in which they attempted to determine how WMC affected learning a morphological system in miniature Latin in two conditions in the case of 57 young learners. In the first, input-based practice and explicit feedback were provided, whereas in the second, practice was preceded by a grammar lesson. Listening span was employed as a measure of WMC. The results revealed that WMC played a role in practice without grammar instruction, but did not moderate the learning effects in the more traditional approach with a direct focus on grammar. The researchers concluded that providing prepractice grammar explanation lessened the effect of individual differences in WMC, but this impact was more pronounced when the instruction was confined exclusively to explicit feedback. In other words, learning with limited instruction places greater demands on WMC. When it comes to task demands, Santamaria and Sunderman (2015) examined the relationship between WMC and L2 beginners' ability to benefit from instruction on comprehension and production tests. Participants were 52 students of French from three American universities. The structure tested was target language direct object pronoun and WM was assessed through a reading span task. An association was found between WMC and beginners' performance on a cognitively demanding explicit production task (i.e., fill-in-the-blanks), but not on a simpler implicit interpretation task (i.e., listening to sentences and choosing the best English translation).

Another study which addressed the role of WM in shaping explicit and implicit L2 grammar knowledge was conducted by Tagarelli et al. (2015) who investigated the relationship between WMC and the acquisition of L2 German syntax under incidental and intentional learning conditions. Participants were 62 native speakers of English with no experience in learning the target language. A grammaticality judgement test measured explicit and implicit knowledge through confidence ratings (i.e., how certain participants were in their choices) and source attributions (i.e., whether the choices were attributed to rule knowledge, memory, guessing, or intuition). Working memory capacity was tapped by means of two nonverbal complex tests: operation-word span and a letter-number ordering task. The researchers found that in the intentional group, the confidence ratings were higher and judgments were based most often on the knowledge of rules. Working memory measures did not correlate with the grammaticality judgment test in either condition. Still, WMC correlated with performance only on grammatical items in the intentional group, which indicates that the impact of WM on grammar learning is multifaceted and manifests itself mainly under explicit conditions.

In contrast to studies conducted by O'Brien et al. (2006) and French and O'Brien (2008), in which the relationships between PSTM and L2 grammar learning were found only for more advanced L2 learners, studies by Grey et al. (2015), Sagarra (2017), and Serafini and Sanz (2016) suggest that WM correlates with grammar learning at lower, but not higher proficiency levels. For example, Grey et al. (2015) did not reveal PSTM and WMC effects with advanced study abroad learners in an implicit (time-constrained) grammaticality judgment task over the period of 5 weeks. Their sample included 26 learners of Spanish with both explicit (i.e., classroom setting) and implicit (i.e., study abroad) experience. The researchers suggest that intensive L2 exposure abroad is a favorable context for advanced L2 learners, which appears not to be constrained by WM limitations. Serafini and Sanz (2016) tested beginning, intermediate, and advanced university L2 Spanish learners. Two nonverbal measures were applied to tap WMC and PSTM: operation span and digit span tasks, respectively. L2 knowledge was measured through two tests conducted three times during one semester. A connection was reported between WMC and PSTM and performance on both the cognitively simple explicit grammaticality judgment task and the cognitively complex implicit elicited oral imitation task in lower, but not higher, proficiency learners.

Suzuki and DeKeyser (2017) examined the interaction between automatized explicit knowledge and implicit knowledge and PSTM in naturalistic L2 acquisition. The sample included 100 advanced L2 speakers of Japanese residing in Japan. Phonological short-term memory was tapped into with a letter span task. Moreover, they applied tests of explicit and implicit aptitude. It was found that the aptitude for explicit learning predicted acquisition of automatized explicit knowledge, which, in turn, predicted acquisition of implicit knowledge, but there was limited effect of implicit learning aptitude and PSTM on the acquisition of both explicit and implicit knowledge. The researchers hypothesized that other higher-order cognitive aptitudes play a more important role in grammar learning than PSTM. In line with previously described studies, the lack of a PSTM effect can be attributed to the high level of advancement of the sample.

More recent studies also place emphasis on cognitive demands of a WM task as a factor determining the relationship between WM and grammar learning. Sagarra (2017) examined longitudinal effects of WMC on grammar learning among 330 beginning English learners of Spanish and found that only WM tests with a processing component yielded evidence for such effects. She attributed the variability in WM effects on processing of morphosyntax and syntax, as well as grammar learning to the effect of the WM test's cognitive load, L2 proficiency, and task demands. In her estimation, these factors, rather than task explicitness result in WMC-grammar learning nexus. Li et al. (2019) conducted a study which is particularly relevant because not only does it provide insights into the role of WMC in the development of explicit and implicit knowledge, but it also involved the grammar feature targeted in the current study. Li et al. examined the effect of WMC in five instructional conditions, diversified according to the presence and timing of instruction focused on English passive voice. Participants were 150 eighth-grade English language learners. Working memory capacity was measured by means of an operation span task while learning outcomes were assessed through an untimed grammaticality judgement test and elicited oral imitation, which represented measures of explicit and implicit knowledge. The feedback was in the form of a prompt and a recast, which required retrieval of information from long-term memory and processing the input. Therefore, WM was heavily loaded. Working memory capacity was correlated with learning outcomes in terms of both explicit and implicit knowledge only in the groups which received within-task feedback, and not in pretest or posttest feedback, which was interpreted as evidence that WM is implicated when there is a heavy processing burden imposed by within-task corrective interventions. A plausible explanation is that learners with greater WM capacity excel at memorizing unanalyzed target language chunks which can be effortlessly accessed in spontaneous production. These results suggest that both explicit and implicit knowledge are affected by WMC but only in on-line processing. Finally, the studies by Pawlak and Biedroń (2019, 2021) focused on the correlations between PSTM, assessed through a nonword span task, and WMC, measured through a listening span task and explicit and implicit grammar knowledge. Both WM tests were cognitively demanding. In Pawlak and Biedroń (2021) the sample included 171 English philology students from three Polish universities. The results showed overall weak correlations between grammar and WM. Specifically, WMC contributed to explicit productive knowledge and PSTM to implicit productive knowledge.

Following Li et al.'s (2019; see also Wen & Li, 2019) recommendations, there is an urgent need for probing further into the relationship between WMC, PSTM, and different types of L2 knowledge. Such research is necessary because, in light of the literature review, the empirical evidence is scant and inconclusive. First of all, there is lack of consistency in the measurement of WMC and PSTM. In the majority of studies, researchers used a single measure of either WMC or PSTM (e.g., French & O'Brian, 2008; Suzuki & DeKeyser, 2017). What is more, only verbal instruments were applied in some studies (Martin & Ellis, 2012; Santamaria & Sunderman, 2015) and only nonverbal ones in others (Serafini & Sanz, 2016; Tagarelli et al., 2015), which significantly limits the interpretation of results. Our study used two tests of PSTM (verbal and nonverbal) and two verbal tests of WMC because verbal measures demonstrate stronger predictive power (Linck et. al., 2014). This not only holds the promise of more fine-grained results but also has the potential to offer insights into the effect of measurement type on these results. Moreover, as most studies indicate that WM effects are observable only in the case of cognitively demanding tasks (Li et al., 2019; Sagarra, 2017), all of our WM tests fulfilled this requirement. Consistently, we decided to apply four measures of grammar knowledge, that is, explicit productive and receptive, as well as implicit (automatized) productive and receptive. When it comes to proficiency, the results of previous studies are contradictory, as in some of them, the effect of WM was visible only for advanced learners (French & O'Brien, 2008; O'Brien et al., 2006), whereas in others, only for beginners (Sagarra, 2017). Since little empirical evidence is available concerning L2 learners representing intermediate levels of proficiency, the majority of our sample were intermediate and upper-intermediate students, although it did contain students who could be characterized as advanced. The current study seeks to fill these research gaps by examining the predictive role of PSTM and WMC in affecting the productive and receptive dimensions of explicit and implicit (automatized) knowledge of passive voice in L2 English.

Research Methodology

Aims and Research Questions

The current study builds on preliminary studies by Pawlak and Biedroń (2019) as well as Pawlak and Biedroń (2021) which applied verbal and nonverbal tests of PSTM and WMC, and involved slightly different samples (not all data were available for all the measures used). It is also part of a larger-scale research project gauging the predictive role of several individual difference (ID) factors in relation to L2 grammar knowledge (cf. Pawlak & Biedroń, 2019, for more details). The specific aim was to shed light on the predictive effect of PSTM and WMC on the development of explicit and implicit (automatized) knowledge of the English passive in comprehension and production. The current study also took into account two mediating variables: overall target language grammar attainment and length of study.

The following two research questions were formulated:

RQ1: What is the relationship between verbal WM and the productive and receptive dimensions of explicit and implicit (automatized) knowledge of the target structure?

RQ2: What is the relationship between length of study, overall attainment in target language grammar, and the productive and receptive dimensions of explicit and implicit knowledge of the target structure?

Our preliminary hypothesis was that both PSTM and WMC would contribute to explicit and implicit (automatized) grammar knowledge, with the effects of WMC being stronger. We also assumed that WM would contribute mainly to the productive dimension of grammar knowledge. We also expected that overall attainment in grammar as well as the length of study might explain some variance in grammar tests results.

Participants

Convenience sampling was applied to gather participants. Participants were 152 English majors, 103 females and 49 males, from three regional Polish universities, enrolled in years 1, 2 and 3 of a three-year BA program in English. English philology studies require a high level of mastery of the target language because they educate professionals prepared to work as teachers, translators, or business representatives. Thus, the program incorporated an intensive course in English, focused on components such as grammar, pronunciation, speaking, or writing, as well as content classes including courses on linguistics, literature, history, culture, or L2 teaching methodology, for the most part taught in the target language. The average length of learning English was 12 years and students' proficiency was assessed to be in the range from intermediate to advanced, that is, between B2 and C1 depending on the level in the program. Nonetheless, there were considerable differences among participants with respect to overall target language mastery. Even though the majority of the students reported regular contact with the target language, this mostly took place through media, and few had an opportunity to stay in an English-speaking country or communicate with native speakers regularly.

Targeted Structure

When deciding on the targeted structure, care needed to be taken to choose a feature that was known to participants to some extent but still posed a certain degree of challenge, both in terms of accurate application of rules with no pressure and their use in spontaneous communication. Passive voice was selected as a feature meeting these requirements. This is because it had been taught to the students on many occasions and they also had to use it on a regular basis. At the same time, since its use involves a wide array of tenses, aspects, and past participle forms, it was expected to guarantee considerable variation in performance. It should be added that the passive can be considered a difficult structure to master in terms of both explicit and implicit knowledge (Ellis, 2006). On the one hand, even with sufficient time allowing access to rules, the ability to use this feature accurately, meaningfully, and appropriately (cf. Larsen-Freeman, 2003) in the myriad of contexts in which it appears poses a formidable task, with consequences for subsequent proceduralization and automatization of such conscious knowledge (DeKeyser, 2015). On the other hand, the development of implicit knowledge necessitates access to complex processing operations (cf. Pienemann & Lenzing, 2015), and even the automatization of explicit knowledge calls for a high degree of coordination of different elements in real time.

Data Collection

Three sources of data were gathered for the purpose of this study: (a) tests of PSTM and WMC, (b) measures of explicit and implicit (automatized) knowledge in production and comprehension, and (c) information about the length of study and grammar attainment. It should be emphasized that the measures of WM as well as L2 knowledge were administered only once, with the effect that the study was cross-sectional in nature.

Measures of PSTM and WMC

Digit Span (PSTM). Wechsler Adult Intelligence Scale − WAIS-R, adapted for use with the Polish population by Brzeziński et al. (1996). This test includes sets of digits to repeat initially forwards then backwards (e.g. “Repeat the numbers 8, 1, 3 in reverse order”). Generally, the digit span is considered to be a nonverbal test of PSTM, but backward repetition is likely to engage the central executive, as this task places a greater load on WM than forward repetition. Split-half reliabilities for the WAIS-R (Pl) were .88-.93 for the full scale. The validity coefficients (i.e., correlations with other intelligence tests, e.g., Raven’s Matrices) ranged between .39-.60.

Nonword repetition (PSTM). Polish Nonword Span (PNWSPAN, Zychowicz et al., 2018), is a test of PSTM, which includes sequences of Polish nonwords (i.e., wordlike forms that do not exist in a given language but that correspond to the language’s phonological rules). All the nonwords are two-syllable, phonologically possible sequences of five Polish sounds in a CVCVC order (e.g., zaduk, homil, julet). The nonwords are prerecorded and displayed in sets of two, three, four, five, and six, with three trials per stage. The set sizes gradually increase from two to six, with a total of 60 nonwords. Participants are asked to repeat the nonwords in the correct order. Partial scoring is applied, with each item being assigned from 0 to 3 points, depending on the quality of its recall. The Cronbach’s α value of the test was .68.

Polish Reading Span (WMC). Polish Reading Span (PRSPAN, Biedroń & Szczepaniak, 2012) is a Polish adaptation of the American Reading Span (Waters & Caplan, 1996). This test comprises a dual task that requires participants to read a series of sentences and, simultaneously, keep track of the last word displayed, so that the words can be recollected later. The test includes eight sets of sentences, which contain from three to 10 sentences in Polish, some of which make sense in everyday life and others do not. The length of each sentence is approximately 10 words and there is an unrelated word at the end, which is a two-syllable noun (e.g., A frog said that it is a nice day today. Tree). The PRSPAN score is the cumulative number of words recalled perfectly in all trials. The maximum score is 52. Test-retest reliability was .89. The Cronbach's α for the internal consistency for the PRSPAN is .69 and the standardized item alpha was .76.

Polish Listening Span (WMC). Polish Listening Span (PLSPAN, Zychowicz et al., 2017) is a test of WMC constructed for the Polish population. It comprises nine sets of sentences which increase in size from two to 10, with a total of 54. Analogically to PRSPAN, some of the sentences make sense in the context of everyday life and others do not (The goat quickly said that it surely preferred the microphone). Test-takers are asked to indicate whether a sentence makes sense to guarantee processing of input, and memorize the last word. The maximum score is 54 points. Test-retest reliability of the tool was established in a pilot study and was .91. PLSPAN was characterized by adequate internal consistency reliability with the Kuder Richardson α at .76.

Measures of L2 Knowledge

Explicit productive knowledge. A pen-and-paper test requesting participants to provide the correct form of 15 verbs placed in a continuous text without time limitations ("When the city (3) (plan) ____________ as the nation capital, the whole process (4) (observe) ____________ by President Washington"). 0, 0.5 or 1 points are given for each answer depending on the severity of the error. 20% of the tests were coded by another researcher, and interrater reliability was .92. Internal consistency, determined by calculating Cronbach's α, equaled .92.

Explicit receptive knowledge. An untimed grammaticality judgment test in which participants decided on the correctness of 15 sentences and offered a justification in cases where they indicate that a given sentence was incorrect. An example item from the test is: "This new blue car has fitted with an alarm." 0, 0.5 or 1 points are given for each response based on the presence of the justification and its quality. 20% of the tests were assessed by another researcher, with interrater reliability of 0.88. The internal consistency reliability of this measure was .62, as established by Cronbach's α.

Implicit (automatized) productive knowledge. A communicative task which required participants to describe a place on the basis of 15 prompts which necessitated the use of the passive. An example prompt is: "locate the suburbs near a beautiful lake and park, surround by a garden, can see from the highway

(3)." The participants had two minutes to go over the prompts and then were given 8 minutes to describe the location. Their oral performance was audio-recorded. A total of 15 points could be scored, with each sentence resulting from a prompt being awarded 0, 0.5 or 1 points using the same criteria as on the test of explicit productive knowledge. 20% of the recordings were coded by another researcher, with interrater reliability of .79. Internal consistency of the test, determined by Cronbach's α, was .87.

Implicit (automatized) receptive knowledge. A timed grammaticality judgment where the participants had to decide on the correctness of 15 sentences embedded in a PowerPoint presentation. An example item is as follows: "Tom deeply regretted not being selected for the team." Each sentence was displayed for an average of seven seconds, but the time was adjusted as a function of its length. The items were included in separate slides and separated from the preceding/following items with an empty slide. Each response could receive 0 or 1 points, with the binary choice eliminating the need to calculate interrater reliability. Cronbach's αfor this measure was at .58.

Target Language Grammar Attainment and Length of Study

Information about overall attainment in English grammar as well as the length of experience in learning the target language were self-reported by participants in questionnaires tapping ID factors that were not the focus of the current study. In the former case, both objective and subjective measures were used in the form of the final semester grammar course grade and self-assessment of this subsystem. In both cases, the choices ranged from 2.0 (lowest or fail), through 3.0, 3.5., 4.0, 4.5, and 5.0 (highest), reflecting institutional grading policies.

Data Analysis

Quantitative analysis was used to analyze the data from the three sources. This involved calculating descriptive statistics (i.e., means and SDs) for each test as well as canonical analysis (CA). Canonical analysis represents a family of general linear model (GLM) methods and is a generalization of multiple regression analysis for many explained variables (Tabachnick & Fidell, 2013). The choice of this procedure was motivated by the fact that the use of regression analysis to assess the model for each of the dependent variables not only distorts the image of the studied phenomenon, but also raises the value of type I error (α). It is referred to as alpha inflation (α*), which means that the value of αincreases correspondingly to the number of executions of the same test (c): $\alpha^* = 1-(1-\alpha)^c$. In order to avoid this, an alternative analytical solution was employed to build a model including all the assumed independent (explanatory) and dependent (response) variables included in the research project.

Canonical analysis examines the latent variables' relationships which are formed by summing the weights of the standardized original independent variables (IV) and summing the weights of the standardized original dependent variables (DV). Thus, the sums of these weights are weighted too. The set of weighted sums of independent (U) and the set of weighted sums of dependent (V) variables are called canonical variables or variates. Paired together, they form orthogonal canonical roots (CRs). The weighted sums in each of the two variates, set U and set V, are selected so as to maximize the correlation between these variates. In the next step of the canonical analysis, the subsequent pairs of variables with maximized correlation are searched to explain the remaining part of the variance of the original variables not explained by the previous canonical root. The number of canonical roots is the same as the number of variables in the smaller set of IV and DV. The canonical correlation decreases with the next extracted CR. The squared value of the canonical load, which is the correlation coefficient between the original variable and variate, informs about the contribution of the original variable to the CR or, in other words, about the amount of variance this variable explained by CR. The arithmetic mean of these squares computed for a given variate is called the extracted variance and informs about the percentage of variance on average explained by this variate in the dataset. The square of the canonical correlation multiplied by the extracted variance of a given set U or V is called redundancy and informs about the average variance of the corresponding set of original variables (e.g., DV) explained by a given variate taking into account the second set of original variables (e.g., IV).

How the canonical variable changes as a result of changes in the value of the original variable is indicated by the canonical weight coefficient. Canonical weights are interpreted similarly as the standardized beta coefficients in multiple regression, but they should be made in connection with the interpretation of the canonical load value. In other words, the original variables with a high value of weight are less significant the lower the value of the factor loading.

Results

Table 1 presents the selected descriptive statistics for the tests of WM and L2 grammar knowledge, length of study, self-assessment, and grades. The results show that the means for all the tests fluctuated around 50-70% of correct answers, with SDs being reasonably low, which excludes floor and ceiling effects. The participants' self-assessment was especially uniform, since most of them evaluated their level as quite high (4.0).

The analysis included the results of measurements for 11 variables. Only the cases with complete data were taken into account. The variables measured on ordinal scales, that is, final grades and self-assessment scores, were subjected to transformation into dummy variables (1, 0) to enable their introduction into the tested model. Since these data were collected with the use of a six-point rating scale, the number of classes of observations corresponding to the points of these scales was reduced according to the following scheme: "3" → "3", "3.5-4"

Table 1. *Descriptive Statistics*

Variable	M	*Median*	*SD*	*Min*	*Max*
Implicit (automatized) productive	5.77	*6.00*	2.80	0.00	12.00
Implicit (automatized) receptive	8.57	8.00	1.80	5.00	14.00
Explicit productive	7.03	7.50	2.74	0.00	14.00
Explicit receptive	10.09	10.50	1.92	4.50	15.00
Digit span	13.03	13.00	3.17	7.00	26.00
Nonwords	73.59	72.00	15.43	27.00	106.00
Reading span	24.53	25.00	5.92	8.00	40.00
Listening span	28.83	29.00	5.00	18.00	41.00
Length of study	11.93	12.00	2.46	5.00	17.00
Self-A	-	4.00	0.25	3.00	6.00
Grade	-	4.00	3.00	3.00	6.00

→ "4", "4.5-6" → "5". As a result, there were nine original IVs: the length of study (length), four variables regarding WM (listening span, reading span, digit span, nonwords), and four dummy variables created on the two original non-metric variables as grade (Grade) and self-assessment (Self-A). The abbreviation schema of the dummy variables is as follows: course grade (Grade_4, Grade_5); self-assessment (Self-A_4, Self-A_5). There were four original dependent variables (DV): implicit (automatized) productive, implicit (automatized) receptive, explicit productive, and explicit receptive.

The quality of the data was checked against the assumptions of canonical analysis. The ratio of the number of variables to the number of observations was sufficient (Laessig & Duckett, 1979), there were no outliers, and there was a lack of collinearity between independent variables. The value of Mahalanobis distance D^2 measure for the multivariate observation farthest from the center of the multivariate distribution was $D^2 = 24.77$, and proved not statistically significant ($p > .001$). As regards collinearity, the tolerance values were higher than .10 ($M = .57$, $Me = .58$, $SD = .13$, Min = .46, Max = .84) and the variance inflation factor (VIF) values were lower than the limit of 10 as established by the rule of thumb (Kutner et al., 2004, $M = 1.84$, $Me = 1.72$, $SD = 0.41$, Min = 1.63, Max = 2.63). The level of statistical significance was set at .05.

The links between IVs and DVs are presented in Table 2. L2 knowledge (response variables) had the strongest relationships with Grade_5 and Self-A_5, and the weakest with digit span as well as Grade_4. Although it is worth noting that Grade_4 is the variable with the highest negative correlation coefficient with automatized knowledge, this should be interpreted in the context of Grade_5, as both of these dummy variables represent the same original grade variable.

Out of the four canonical roots (four pairs of canonical variables), two were statistically significant. This means that canonical variables were correlated in the context of first and second canonical root (*CR1* and *CR2*). The correlations between the rest of the canonical variables were not statistically significant and,

Table 2. *Relationships Between Explanatory and Response Variables – Pearson's r Linear Correlation Coefficient*

Explanatory Variable	Implicit (automatized) productive	Implicit (automatized) receptive	Explicit productive	Explicit receptive
Digit span	.08	.05	.20	.14
Nonwords	.17	.12	.16	.15
Reading span	.15	.20	.15	.09
Listening span	.13	.19	.29	.16
Length of study	.08	.19	.20	.13
Grade_4	−.18	−.12	−.02	.06
Grade_5	.40	.27	.40	.20
Self-A_4	−.14	−.09	−.14	.05
Self-A_5	.38	.16	.38	.16

Note. Self-A = self-assessment.

therefore, the interpretation was limited to *CR1* and *CR2*. The value of canonical correlation (R) and the range of explained variance (R^2) by the canonical roots are presented in Table 3. The model including all canonical roots was characterized by a correlation of two weighted sums (i.e., of independent variables and dependent variables) equal to .59 and explained 34% of the variance of original variables by canonical variables. In the case of *CR1*, the canonical correlation between canonical variables *U* and *V* was .21 and the canonical R^2 was .19. The values of Wilk's λ, that is, the multivariate generalization of R^2 confirmed the result of moderately high correlation in the case of *CR1* and rather low correlation in the case of *CR2*.

In the structure of connections of both sets of variables, the role of grade, self-assessment, explicit productive knowledge, implicit (automatized) productive knowledge, and listening span could be observed. A review of the canonical weights and the canonical loadings (see Table 4) indicates that the largest contribution to the model was made by Grade_5, Self-A_5, as well as explicit productive and implicit (automatized) productive knowledge. Next in order were Grade_4, implicit (automatized) receptive knowledge, explicit receptive knowledge, as well as listening span. It should be emphasized that while canonical loadings inform about the contribution of the original variable to

Table 3. *The Test of Canonical Roots Significance*

Removed	Canonical *R*	Canonical R^2	X^2	*df*	*p*	Wilk's λ
0	.59	.34	98.27	36	.0000	.51
1	.38	.15	37.70	24	.0373	.77
2	.27	.08	15.05	14	.3747	.90
3	.16	.03	3.76	6	.7087	.97

Note. *R* = canonical correlation; R^2 = range of explained variance; X^2 = chi square; *df* = degrees of freedom.

the function which is *CR*, the canonical weights might be interpreted similarly to the beta coefficients in multiple regression. Accordingly, in the case of *CR1*, the changes of final grade to 5, self-assessment to 5, and listening span rise caused a marked increase of explicit productive knowledge as well as, though to a slightly lesser extent, (implicit) automatized productive and receptive knowledge. Loads and weights for *CR2* reinforce this interpretation, especially against implicit (automatized) productive knowledge: its level increased as the final grade changed from 4 to another and the listening span score decreased.

The factor load of listening span for CR1 was moderately high and positive, but it was rather low and negative in the case of CR2. Taking into account the value of the canonical correlation and the range of DV variance explained by the canonical variable for a given set of IVs (see Table 4), it should be concluded that listening span, together with the other IVs, positively affected the DVs. The understanding of the canonical factor loadings strengthens the interpretation based on the canonical weights review of the relationships between the set of IVs and the set of DVs, and the distinctive role of individual original variables in the structure of these relationships. At the same time, it indicates the need for further investigation of the role of listening span (WMC) in shaping the level of explicit and implicit (automatized) knowledge of L2 grammar.

Table 5 presents information about the variance of varieties explained by canonical roots. Regarding the significant canonical roots (CR1, CR2), the general amount of extracted variance by all CRs was 30% of the IV set (U) and 63% of

Table 4. *Canonical Weights and Loadings*

Original variable	Weights		Loadings	
	CR1	*CR2*	*CR1*	*CR2*
Variable *U*				
Digit span	.14	−.34	.31	−.27
Nonwords	−.02	.33	.32	.13
Reading span	.09	.50	.32	.12
Listening span	.16	−.70	.48	−.33
Length of study	.16	−.30	.36	−.25
Grade_4	.32	−.68	−.12	−.63
Grade_5	.73	−.24	.76	.30
Self-A_4	.20	.20	−.25	−.22
Self-A_5	.48	.64	.68	.31
Variable V				
Implicit (automatized) productive	.21	1.28	.76	.61
Implicit (automatized) receptive	.21	.06	.52	.08
Explicit productive	.72	−.88	.96	−.18
Explicit receptive	.07	−.29	.56	−.20

Note. CR = canonical root; *U* = independent variables; *V* = dependent variables; Self-A = self-assessment.

Table 5. *Extracted Variance and Redundancy*

Canonical root	Set U		Set F	
	Extracted variance	Redundancy	Extracted variance	Redundancy
1	.1982	.0680	.5200	.1785
2	.0996	.0145	.1140	.0166
3	.0673	.0051	.1967	.0148
4	.1288	.0033	.1693	.0044
Total	.4938	.0909	1.0000	.2143
$Sum_{CR1+CR2}$	.2978	.0825	.6340	.1951

Note. U = independent variables; *V* = dependent variables.

the DV set (V). The redundancy (the percent of variance in the first set of original variables based on the second set and its input to the CR) was, respectively, 8% of IV and 19% of DV. For a model with all canonical functions, the extracted variance and redundancy were obviously much higher, which indicated that it was appropriate to include the dependent and independent variables in the model.

Discussion

The analysis presented above corroborated the existence of a relationship between different aspects of WM, grammar attainment, length of study, and the mastery of the English passive. The canonical correlation for the entire model was rather high (cf. Plonsky & Oswald, 2014), which means that the original variables included in the model remained significantly related to each other. However, a closer analysis demonstrated that it was mainly grammar attainment as reflected in final grades, self-assessment, and then, to a lesser extent, the listening span results that determined the levels of both explicit productive and implicit (automatized) productive grammar knowledge. Other variables also explained both types of knowledge, although their role was less pronounced.

When it comes to RQ1 concerning the relationship between WM and explicit and implicit (automatized) grammar knowledge, the results showed that it was chiefly WMC, as measured by listening span, that affected both types of knowledge but primarily with respect to its productive dimension (the highest factor loadings in Table 4). At the same time, a much weaker, almost negligible, impact of PSTM was observed. In many studies, productive tasks have been shown to place heavy demands on the participants' WM, reflected in correlations between both WMC and PSTM and explicit productive knowledge (French & O’Brien, 2008; Santamaria & Sunderman, 2015; Williams & Lovatt, 2003). This effect is particularly noticeable in cognitively demanding tasks. In line with previous studies, our results indicate that it was the measure of explicit productive knowledge that proved to be the most demanding in this respect, with the effect that the students, especially with a greater listening span (WMC), were better able to gain access to rules in their long-term memory needed to provide the accurate forms of the verbs. The same applies to implicit (automatized) productive knowledge, which was shown to correlate with both PSTM and WMC (O'Brien et al., 2006; Serafini & Sanz, 2016). In the current study, the implicit (automatized) productive knowledge task was to describe a location with the help of a set of prompts in real-time, which required spontaneous language production. This test favored those with greater executive capacities, which was reflected in the correlation with WMC, although lower than that with explicit productive knowledge. At the same time, WMC and PSTM played a secondary role in determining the results of the tests of receptive explicit and implicit (automatized) knowledge. It could be hypothesized that the measure of receptive explicit knowledge (i.e., untimed grammaticality judgement) might not have placed heavy demands on WM since it did not require the use of rules to actually construct oral and written passive voice utterances. Moreover, the participants had time to access rules when deciding whether or not the sentences were grammatical and providing a justification in cases when they were not; therefore their WMC was not overloaded. Our result is different from the findings of Serafini and Sanz (2016), who found a correlation between WMC and receptive explicit knowledge. This divergence might be attributed to differences in cognitive load between tasks, with our tasks being less demanding in this respect. Indeed, this test was the easiest for the participants, as the median was the highest of all the measures of the knowledge of the passive. The lack of a stronger relationship between WM and performance on the test tapping receptive implicit (automatized) knowledge coheres with the results obtained by Grey et al. (2015), Tagarelli et al. (2015), and Suzuki and DeKeyser (2017), who found no significant effects of either PSTM or WMC in implicit receptive tasks. This test could have been the least demanding since not only did it not require the application of rules to construct passive voice forms, but it also did not demand the learners to provide explanations of relevant rules. Moreover, the substantial time pressure could have resulted in the learners making random choices when they could not promptly access explicit knowledge. Still, the canonical weights and factor loadings for the tests of WM as well as receptive implicit (automatized) knowledge and receptive explicit knowledge were notably different from zero (see Table 4 , CR1), which indicates that WM has an impact on both types of receptive grammar knowledge as well. These results are in most part in line with previous studies (Martin & Ellis, 2012; Li, et.al, 2019; Linck et al., 2014; Wen, 2019; Wen & Li, 2019) in that they confirm the role of WM in grammar knowledge development. Our results corroborate Sagarra's (2017) claim that it is not task explicitness but cognitive demands of the task that influence WM effects. Evidently, productive tasks consume more WM resources than cognitively less demanding receptive tasks such as grammaticality judgment tests with or without time constraints.

Somewhat surprising was the minor significance of PSTM. Some previous studies conducted on adult populations (Martin & Ellis, 2021; O'Brien et al., 2006; Serafini & Sanz, 2016; Williams & Lovatt, 2003) demonstrated a substantial impact of this type of memory on grammar knowledge. On the other hand, in other studies

(Grey et al., 2015; Suzuki & DeKeyser, 2017), a marginal impact of PSTM was observed. Therefore, our hypothesis regarding the role of this factor was speculative. However, even though its effect on the DVs was weaker than in the case of the other factors, it was still detected, as factor loadings for the nonwords and digit span task oscillated around .30 (see Table 4). Of these two, the digit span task scores, which increased the level of productive and receptive knowledge as they grew (see canonical weights for factor loadings respectively) seem significant.

Attention should also be directed to the stronger impact of the listening span task than that of the reading span task. It seems that although both tests are reliable measures of WMC, they are not equivalent. Perhaps this divergence can be explained in terms of a threshold for both the cognitive burden placed on the participants, in this case higher for the listening span task, which is a factor difficult to control, and their level of proficiency or/and a trade-off between these two factors. These results indicate that much of the divergence in the results of different studies can be attributed to test properties as well as the measurement procedure.

The factor of proficiency might have also contributed to the weaker correlations between WM and different types of grammar knowledge. It is likely that higher levels of proficiency are less subject to WMC limitations (Sagarra, 2017). In the current study, the participants were classified as intermediate-upper intermediate-advanced. However, it should be emphasized that their level of proficiency was uneven, with some skills being better developed than others. Apparently, our study confirms lower WM effects for higher proficiency levels.

With respect to RQ2, it seems that the role of overall target language grammar attainment was much more pronounced than the role of WM. In fact, all the dimensions of L2 grammar knowledge, that is, explicit and implicit (automatized) productive as well as explicit and implicit (automatized) receptive, were related to the final grades in a grammar course and self-assessment. On careful reflection, such outcomes are not overly surprising and point to the fact that years of formal study of the target language, which mostly involved intensive explicit instruction and deliberate efforts to consciously learn and apply grammar structures, may have trumped the effects of latent variables such as different aspects of WM. In other words, it is to some extent predictable that participants with greater overall mastery of grammar are likely to perform better on tests of explicit and implicit (automatized) knowledge with respect to a specific structure, in this case, the English passive. Thus, it would seem that instruction and feedback are crucial to the development of L2 grammar knowledge and that grades and self-assessment may serve as adequate measures of this knowledge. On the other hand, it cannot be ruled out that such performance was also a function of other ID variables that were not examined in the current study such as, for example, strategy use, learning styles, personality, or beliefs. The impact of such factors should surely be taken into account as well in future research exploring the link between WM and mastery of L2 grammar.

Limitations

To our knowledge, the current study was the first that examined in such a detailed way the role of WMC and PSTM, as well as grammar attainment, self-assessment, and length of study, focusing on the productive and receptive dimensions of explicit and implicit (automatized) knowledge of a specific target language structure. Confronting such a challenge inevitably entails some weaknesses, visible only after analyzing the results, which are in large part related to the design of the tasks intended to tap into different types of L2 knowledge. The test of implicit (automatized) productive knowledge might have actually targeted speeded explicit knowledge, which resulted in participants' partial reliance on memorized chunks. As far as the test of receptive implicit (automatized) knowledge is concerned, the amount of time needed to process the sentences in the PowerPoint presentation could have been better optimized. There could also be doubts about the reliability of the tests, specifically those addressing the receptive aspects of L2 knowledge which were lower for this sample than in the pilot studies. Moreover, the factor of proficiency was not sufficiently controlled for, as the sample was very diversified in this respect, from intermediate to advanced levels. Another controversy arises in connection with the WM tests. The current study clearly shows that two tests tapping into the same element can yield different results, which raises doubts about their equivalence. Moreover, there are other ID variables that affect the development of target language grammar knowledge, such as foreign language aptitude, overall proficiency, the use of language learning strategies and differences in the knowledge of the passive by the participants, that is, factors that were not taken into account.

Conclusions

Overall, the current results confirm the significant role of WMC as well as the less important role of PSTM in the development of L2 grammar knowledge (cf. French & O'Brien, 2008; Martin & Ellis, 2012; Pawlak & Biedroń, 2019; Pawlak & Biedroń, 2021; Williams & Lovatt, 2003). They also corroborate the association between WMC and cognitively demanding tasks requiring explicit knowledge such as the productive tasks of L2 knowledge employed in this study (Santamaria & Sunderman, 2015; Sagarra, 2017, Li et al., 2019). However, this effect was weaker for receptive tasks, which can be attributed to smaller cognitive burden in language reception. In the case of explicit receptive (untimed grammaticality judgement) and implicit (automatized) receptive (timed grammaticality judgement) tasks, participants could have relied more heavily on their intuitive judgment rather than engage their WM to a great extent. Moreover, WM is part of a wider set of factors that play a role in shaping grammar knowledge such as learning conditions, type of task, learner age and proficiency level, and other ID variables, but also the ways in which the knowledge of grammar is tested or the manner WM tests are conducted and scored.

The interpretation of the canonical analysis results is rather disappointing, as it shows the impact of WM on the development of grammar knowledge as smaller than it was expected. Still, the relationships between the set of independent variables and the set of dependent variables were observed and the distinctive role of individual original variables in the structure of these relationships was revealed. At the same time, the findings point to the need for further investigation of the role of WM in shaping the level of explicit and implicit (automatized) grammar knowledge.

Conflict of Interest Disclosure

The Authors report no conflicts of interest.

Funding

The study reported in this article represents a contribution to the research project no. 2015/17/B/HS2/01704 (2016-2019) funded by the National Science Centre, Poland.

Research Ethics Statement

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants involved in the study.

Authorship Details

Adriana Biedroń: Research concept and design, collection and/or assembly of data, data analysis and interpretation, writing the article, critical revision of the article, final approval of the article.

Sławomir Pasikowski: data analysis and interpretation, critical revision of the article.

Mirosław Pawlak: Research concept and design, collection and/or assembly of data, data analysis and interpretation, writing the article, critical revision of the article, final approval of the article.

References

Antoniou, M., Ettlinger, M., & Wong, P. C. M. (2016). Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning. *PLoS One 11*(7), e0158812. https://doi.org/10.1371/journal.pone.0158812

Baddeley, A.D., Gathercole, S, & Papagno, C, (1998). The phonological loop as a language acquisition device. *Psychological Review, 105*, 158–173. https://doi.org/10.1037/0033-295x.105.1.158

Biedroń, A. & Szczepaniak, A. (2012). Working-memory and short-term memory abilities in accomplished multilinguals. *Modern Language Journal, 96*(2), 290–306. https://doi.org/10.1111/j.1540-4781.2012.01332.x

Bielak, J., & Pawlak, M. (2013). *Applying cognitive grammar in the foreign language classroom: teaching english tense and aspect.* Springer.

Brzezinski, J., Gaul, M., Hornowska, E., Machowski, A., & Zakrzewska, M. (1996). *Skala inteligencji D. Wechslera dla dorosłych. Wersja zrewidowana. WAIS-R (Pl). Podręcznik* [Wechsler adult intelligence scale. A revised version. WAIS-R (Pl). Manual]. Pracownia Testów Psychologicznych PTP.

DeKeyser, R. (2009). Cognitive-psychological processes in second language learning. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 119–138). Wiley-Blackwell.

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition. An introduction* (pp. 94–112). Routledge.

DeKeyser, R. (2017). Knowledge and skill in SLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15–32). Routledge.

DeKeyser, R. M., & Juffs, A. (2005). Cognitive considerations in L2 learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 437–454). Lawrence Erlbaum.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford University Press.

Ellis, R. (2004). The definition and measurement of explicit knowledge. *Language Learning, 54*(2), 227–275. https://doi.org/10.1111/j.1467-9922.2004.00255.x

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*(2), 141–172. https://doi.org/10.1017/S0272263105050096

Ellis, R. (2006). Modeling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics, 27*(3), 431–463. https://doi.org/10.1093/applin/aml022

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Multilingual Matters.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (Eds.). (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Multilingual Matters.

Engel de Abreu, P. M., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology, 104*(4), 974–986. https://doi.org/10.1037/a0028390

Ettlinger, M., Bradlow, A. R., & Wong, P. C. M. (2014). Variability in the learning of complex morphophonology. *Applied Psycholinguistics, 35*(4), 807–831. https://doi.org/10.1017/S0142716412000586

French L. (2006). *Phonological working memory and second language acquisition: A developmental study of francophone children learning English in Quebec.* Edwin Mellen Press.

French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics, 29*(3), 463–487. https://doi.org/10.1017/S0142716408080211

Gagné, N., French, L. M., & Hummel, K. M. (2022). Investigating the contribution of L1 fluency, L2 initial fluency, working memory and phonological memory to L2 fluency development. *Language Teaching Research, 0*(0). https://doi.org/10.1177/13621688221076418

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition, 34*(3), 445–474. https://doi.org/10.1017/S0272263112000149

Grey, S., Cox, J. G., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *The Modern Language Journal, 99*(1), 137–157. https://doi.org/10.1111/modl.12190

Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition, 35*(3), 423–449. https://doi.org/10.1017/S0272263113000041

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory: L2 question development through recasts in a laboratory setting. *Studies in Second Language Acquisition, 37*(3), 549–581. https://doi.org/10.1017/S0272263114000618

Krashen, S. (1981). *Second language acquisition and second language learning.* Pergamon.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied linear regression models.* McGraw-Hill Irwin.

Laessig, R. D., Duckett, E. J. (1979). Canonical correlation analysis: potential for environmental health planning. *American Journal of Public Health, 69*(4), 353–360. https://doi.org/10.2105/AJPH.69.4.353

Larsen-Freeman, D. (2003). *Learning grammar: From grammar to grammaring.* Henle & Heinle.

Li, S. (2013). 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. (2019). Six decades of language aptitude research: A comprehensive and critical review. In Z. E. Wen, P. Skehan, A. Biedroń, S. Li, & R. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 78–97). Routledge.

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics, 40*(3), 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*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological STM and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*(3), 379–413. https://doi.org/10.1017/S0272263112000125

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27*(3), 377–402. https://doi.org/10.1017/S0142716406060322

Paradis, M. (2009). *Declarative and procedural determinants of second languages.* John Benjamins.

Pawlak, M. (2019). Tapping the distinction between explicit and implicit knowledge: Methodological issues. In B. Lewandowska-Tomaszczyk (Ed.), *Contacts & contrasts in educational contexts and translation* (pp. 45–60). Springer Nature.

Pawlak, M., & Biedroń, A. (2019) Verbal working memory as a predictor of explicit and implicit knowledge of English passive voice. *Journal of Second Language Studies, 2*(2), 258–280. https://doi.org/10.1075/jsls.19007.paw

Pawlak, M., & Biedroń, A. (2021). Working memory as a factor mediating explicit and implicit knowledge of English grammar. *Annual Review of Applied Linguistics, 41*, 118–125. https://doi.org/10.1017/S0267190521000052

Pienemann, M., & Lenzing, A. (2015). Processability theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (2nd ed., pp.

159–179). Routledge.

Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning, 62*(1), 93–132. https://doi.org/10.1111/j.1467-9922.2011.00690.x

Sagarra, N. (2017). Longitudinal effects of working memory on L2 grammar and reading abilities. *Second Language Research, 33*(3), 341–363. https://doi.org/10.1177/0267658317690577

Santamaria, K., & Sunderman, G. (2015). Working memory in processing instruction: The acquisition of L2 French clitics. In Z. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 205–223). Multilingual Matters.

Sanz, C., Lin, H.-J., Lado, B., Stafford, C. A., & Bowden, H.W. (2014). One size fits all? Learning conditions and working memory capacity in ab initio language development. *Applied Linguistics, 37*(5). 669–692. https://doi.org/10.1093/applin/amu058

Serafini, E. & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition, 38*(4), 607–46. https://doi.org/10.1017/S0272263115000327

Suzuki, Y., & DeKeyser, R. M. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning, 67*(4), 747–790. https://doi.org/10.1111/lang.12241

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics.* Pearson.

Tagarelli, K. M., Borges Mota, M., & Rebuschat, P. (2015). Working memory, learning conditions, and the acquisition of L2 syntax. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing: theory, research and commentary* (pp. 224–247). Multilingual Matters.

Ullman, M. T. (2015). The declarative/procedural model. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (2nd ed., pp. 135–158). Routledge.

Verhagen, J., Leseman, P., & Messer, M. (2015). Phonological memory and the acquisition of grammar in child L2 learners. *Language Learning, 65*(2), 417–448. https://doi.org/10.1111/lang.12101

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology, 49*(1), 51–79. https://doi.org/10.1080/713755607

Wen, E. Z. (2016). *Working memory and second language learning: Towards an integrated approach*. Multilingual Matters.

Wen, E. Z. (2019). Working memory as language aptitude: The phonological / executive model. In Z. E. Wen, P. Skehan, A. Biedroń, S. Li, & R. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 187–215). Routledge.

Wen, E. Z., Biedroń, A., & Skehan, P. (2016). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching 50*(1), 1–31. https://doi.org/10.1017/S0261444816000276

Wen, Z., & Li, S. (2019). Working memory in L2 learning and processing. In J. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 365–389). Cambridge University Press.

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning, 53*(1), 67–121. https://doi.org/10.1111/1467-9922.00211

Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition, 37*(3), 457–486. https://doi.org/10.1017/S0272263114000370

Zychowicz, K., Biedroń, A., & Pawlak, M. (2017). Polish Listening Span: A new tool for measuring verbal working memory. *Studies in Second Language Learning and Teaching, 7*, 601–618.

Zychowicz, K., Biedroń, A., & Pawlak, M. (2018). Polish Nonword Span (PNWSPAN): A new tool for measuring phonological loop capacity. *Glottodidactica, 45*(2). *309–327*. https://doi.org/ http://dx.doi.org/10.14746/gl.2018.45.2.18