

KATARZYNA ZYCHOWICZ

Akademia Pomorska w Słupsku
katarzyna.zychowicz@apsl.edu.pl
ORCID: 0000-0002-8420-403X

ADRIANA BIEDROŃ

Akademia Pomorska w Słupsku
adriana.biedron@apsl.edu.pl
ORCID: 0000-0002-4382-0223

MIROŚLAW PAWLAK

*Uniwersytet im. Adama Mickiewicza w Poznaniu / Państwowa Wyższa Szkoła Zawodowa
w Koninie*
pawlakmi@amu.edu.pl
ORCID: 0000-0001-7448-355X

Polish Nonword Span (PNWSPAN): A new tool for measuring phonological loop capacity¹

ABSTRACT. The phonological loop, which is a component of working memory, is considered to be one of the most significant factors affecting L1 and L2 learning. In order to measure this construct properly, a reliable instrument in the native language of the participants is needed. The purpose of this paper is to present the Polish Nonword Span PNWSPAN, which is a tool constructed to measure verbal working memory, in particular the phonological loop, in the case of adults. The article presents the theoretical framework of the study and the process of construction of the test, namely its structure, scoring and validation procedure.

KEYWORDS: working memory, phonological loop, nonword repetition.

¹ The study reported in this paper represents a contribution to the research project no. 2015/17/B/HS2/01704 (2016–2019) funded by the National Science Centre, Poland.

1. INTRODUCTION

Working memory (WM) is a concept adapted from cognitive psychology, defined as “the ability to mentally maintain information in an active and readily accessible state while concurrently and selectively processing new information” (Conway, Jarrold, Kane, Miyake & Towse 2007: 3). The phonological loop is the most frequently referred to aspect of WM; moreover, next to the central executive, it seems the most relevant to the theory of individual differences in SLA (Wen, Biedroń & Skehan 2016; Wen 2016). Evidence that WM storage and executive components affect foreign/second language (L2) learning and processing (Linck, Osthus, Koeth & Bunting 2014; Wen 2015; 2016) has accumulated over the last two decades. However, there are still many controversies surrounding this issue and many questions remain unanswered. One of the problems concerns the methods of measurement and scoring of WM components. Since a cognitive test should be constructed in the participants’ native language (Linck et al. 2014; Conway et al. 2007), we constructed two Polish tools for measuring WM capacity: a listening span, which is a measure of the central executive, and a nonword span, which is a measure of the phonological loop. The first of these, the Polish Listening Span (PLSPAN, Zychowicz, Biedroń & Pawlak 2016), is intended to measure the central executive. This article describes the process of the construction of the latter, the Polish Nonword Span (PNWSPAN), designed as a measure of the phonological loop for adult native speakers of the Polish language. At the beginning we outline the theoretical background of our study, that is, the construct of WM, placing emphasis on its two most important components, that is, the phonological loop and the central executive, as well as the ways in which these components can be measured. Subsequently, we describe the process of the construction of the tool, and its reliability and validity measures. Conclusions and suggestions for further research are provided at the end of the paper.

2. WORKING MEMORY

Deeply rooted in cognitive psychology, the concept of WM is defined as the ability to temporarily store, process and maintain a circumscribed amount of data while performing mentally demanding tasks. As elucidated by Baddeley (2003: 204), it is “a temporary storage system that underpins our capacity for thinking”. As such, it is essential for numerous complex cognitive activities, such as, for example, reasoning, mental calculations, as well as language learning, comprehension and production (Baddeley 2003; Juffs & Harrington

2011). Crucial in research in cognitive psychology and cognitive neuroscience (Conway, Macnamara & Engel de Abreu 2013), WM is the core of general intellectual functioning, and its measures are indicators of intellectual ability (Conway et al. 2013; Kane, Conway, Hambrick & Engle 2008).

Although the concept and definition of WM are still under debate, the most widespread and researched is the multicomponent model proposed by Baddeley and Hitch (1974), and Baddeley (2000). The model consists of the *central executive* (CE) component, which is a supervisory attention-limited control system, the *episodic buffer*, used for storing integrated information, and two slave storage systems, namely, the *phonological loop* (PL), responsible for processing verbal and acoustic information, and the *visuo-spatial sketchpad*, involved in processing visual and spatial information. Two of these, that is, the CE and the PL, seem to be the most crucial elements for language acquisition and have been referred to as verbal WM (Wen 2016). Thus, they have been extensively researched within the field of second language acquisition (SLA) as contributing to both the process and outcome of L2 learning (Biedroń & Szczepaniak 2012a; Biedroń & Pawlak 2016; DeKeyser & Juffs 2005; DeKeyser & Koeth 2011; Doughty 2013; Doughty et al. 2010; Juffs & Harrington 2011; Mackey, Philip, Egi, Fujii & Tatsumi 2002; Miyake & Friedman 1998; Papagno & Vallar 1995; Pawlak 2017; Robinson 2003; Sawyer & Ranta 2001; Skehan 2012; Suzuki & DeKeyser 2017; Wen & Skehan 2011; Wen, Mota & McNeill 2015; Wen 2016, Williams 2012).

Ample research in the field of cognitive psychology and SLA (for a review see Wen 2016) provides evidence that both verbal components of WM, namely the PL and the CE, consistently influence numerous features of SLA at different ages and at different proficiency levels. The PL seems to affect mainly language subsystems, while the CE can induce differences in language skills. Most research is based on the assumption that if WM is central to higher-level cognitive functions, then individual differences in WM capacity should result in significant differences in SLA. Table 1 briefly summarises some of the research investigating the role of WM in SLA.

As can be seen from Table 1, the relationship between SLA and WM is far from clear and the results of research are often ambiguous, grammar learning being perhaps the best example of this inconclusiveness and controversy. Several studies (e.g., Fortkamp 2003, Kormos & Sáfár 2008; Linck et al. 2014; Martin & N. Ellis 2012; O'Brien, Segalowitz, Collentine & Freed 2006; Suzuki & DeKeyser 2017; Williams & Lovatt 2003) concentrate on the relationship between grammar and WM. Fortkamp (2003) used a speaking span as a measure of the CE, correlated it with speech production, and found a positive link between CE and structural complexity, accuracy and fluency. Her conclusion was that key elements of CE, namely attention regu-

Table 1. Results and findings from WM-SLA studies. Adapted from Wen 2016

SLA domains and activities	PL	CE	Major SLA studies
L2 vocabulary acquisition and development	Instrumental in storing and acquiring novel phonological forms	Not yet clear	Bolibaugh and Foster (2013), Cheung (1996), Ellis and Sinclair (1996), Foster, Bolibaugh and Kotula (2014), French and O'Brien (2008), Service (1992), Speciale, Ellis and Bywater (2004)
Acquisition and development of L2 grammar	Facilitates the storage and chunking of morphosyntactic constructions	Not yet clear	Martin and Ellis (2012), Williams and Lovatt (2003)
L2 language comprehension (listening and reading)	Used to maintain a phonological record that can be consulted during offline language processing	Facilitates processing syntactic and semantic information	Alptekin and Erçetin (2011), Berquist (1997), Harrington and Sawyer (1992), Leeser (2007), Miyake and Friedman (1998)
Language production (speaking and writing)	Predicts narrative vocabulary at early stage; predicts grammatical accuracy at later stage	Is related to performance measures of L2 speech (e.g., accuracy)	Abu-Rabia (2003), Ahmadian (2012), Fortkamp (1999 and 2003), O'Brien Segalowitz, Collentine and Freed (2006), Payne and Whitney (2002)

lation and control, correspond with grammatical encoding in L2 speech production. Williams and Lovatt (2003) linked grammar rule learning and the PL in a semiartificial language; yet, their results failed to fully explain the variation in grammar acquisition leading them to the conclusion that both the PL and the CE are necessary to understand the cognitive basis for grammar learning. O'Brien et al. (2006) researched the role of the PL in lexical, grammatical and narrative skills of adults during speech production. What they established was the dependence of grammatical proficiency on the PL at later stages of L2 learning. The study by Kormos and Sáfár (2008) concentrated on L2 performance in an intensive English program, as measured by the results of Cambridge First Certificate in English exam, which they matched with the PL and CE capacities. Even though the study demonstrated a significant correspondence between measures of WM and L2 performance, it would be difficult to draw conclusions concerning grammar exclusively since no part of the FCE examination concentrates on grammar per se. Martin and N. Ellis (2012) analysed the influence of the PL and the CE on grammar and vocabulary learning in an artificial language. They documented significant independent effects of both components of WM on vocabulary and grammar learning. They also addressed the question whether the influence of WM on grammar is direct or rather it constitutes a by-

product of its influence on vocabulary, providing evidence of the direct effect of WM on grammar learning regardless of the mediating effect of vocabulary. Finally, Suzuki and DeKeyser (2017) investigated to what extent WM capacity predicted the acquisition of grammar under two learning conditions: distributed practice (7-day interval) and massed practice (1-day interval). It turned out that WM capacity was only related to performance after massed practice.

The relative scarcity of research on the contribution of WM to grammar learning and its inconclusiveness do not allow making claims about which component of WM is more important for the mastery of this subsystem. According to Linck et al. (2014), the CE is likely to be more strongly correlated with grammar production than the PL. However, research in the field of L1 acquisition has led some scholars, including Baddeley himself, to believe that it is the PL that is crucial for language learning in general, including the development of not only vocabulary but also grammar, be it in L1 or L2. As Baddeley, Gathercole and Papagno (1998: 166) point out, “the loop system mediates the acquisition of syntactic knowledge, as well as the learning of individual words”. They argue that even though the PL has been the most widely researched component of WM, its role has often been underestimated and ascribed to ‘dealing with phone numbers’ whereas, in fact, “the primary function of the phonological loop is the processing of novel speech input” (Baddeley et al. 1998: 166). Therefore, they refer to the PL as the language learning device.

3. THE MEASUREMENT OF WORKING MEMORY

No single measure can capture the capacity of WM for the simple reason that the construct is multicomponential. Additionally, several factors influence the choice of a specific task used for evaluation, some of them being implementation, complexity, content and language (Linck et al. 2014: 863). The *implementation factor* refers to the distinction between receptive and productive tests, with the emphasis being placed on productive rather than receptive performance, or, in other words, repetition rather than recognition tasks, due to the involvement of articulatory processes (Baddeley et al. 1998; Gathercole 2006; Wen 2016). The *language factor* concerns the language in which the tests are administered, that is, L1 or L2. Since proficiency level may impact scores on WM span tasks conducted in the L2, Wen (2012) postulates holding WM tests in participants’ L1, which will eliminate the risk of confounding the results, particularly in the field of SLA (see also Kormos & Sáfár 2008). Another factor, *the content of tests*, concerns the question whether

the content of a task is verbal, visual or spatial, among others. Botting, Psarou, Caplin and Nevin (2013) analysed the extent to which the level of verblity of a task correlates with linguistic input and concluded that, although the influence of content is not overwhelming, it is statistically significant. In line with that, Wen (2012) calls for the use of verbal tasks in researching the relationship between SLA and WM. The *complexity of the task factor* is related to the components of WM. The PL, as the storage element, is assessed with simple span tasks, which involve memorising lists of items in the right order, while the CE, as the processing component, is evaluated with complex span tasks, which require memorising material while processing verbal input. Typical verbal simple span tasks include letter span, word span and nonword span whereas typical verbal complex span tasks can take the form of reading span, listening span, speaking span and English opposites span (Linck et al. 2014: 865).

By far the most popular simple span measure is digit span, which constitutes one part of Wechsler Intelligence Scale (Wechsler 1997). During the test participants repeat strings of numbers of increasing length, from 3 to 10 elements. While the test is a valid and reliable measure of WM, its verblity is more and more often questioned (Linck et al. 2014) since the verbal input is easily transferable into visual input. This might indicate that digit span measures the capacity of not only the PL but also of the visuo-spatial sketchpad. Another simple span task is letter span, based on the same idea as digit span, the only difference being the repetition of strings of letters rather than digits. Similarly to digit span, there are doubts whether it can be regarded as a pure measure of verbal WM. The PL capacity can also be tapped by means of word span which is another example of a simple span measure. The participant's task is repeating increasingly longer lists of unrelated words. As popular as digit span, word span is used extensively as a measure of verbal memory. However, yet again, objections have been raised concerning its being a reliable measure of WM, as it is often claimed to tap into recall from long-term memory rather than the PL capacity (Gathercole 1995).

Apart from the aforementioned tasks, a measure that is intended to capture the complex nature of the construct of the PL is the nonword repetition span task (Dollaghan & Campbell 1998; Gathercole 2006; Gathercole, Willis, Baddeley & Emslie 1994; Gathercole, Hitch, Service & Martin 1997). Nonwords are phonologically probable combinations of sounds, which have the form of words but, since no meaning is attached to them, they cannot be referred to in this way. Nonword repetition span surpasses all the previously mentioned tests as the best-established and the most extensively researched method of measuring the PL in cognitive psychology (Gathercole 2006), most frequently used for cognitive, language and reading evaluation,

with particular emphasis on specific language impairment (SLI) diagnosis (Gray 2003). Besides, it has the most explicit verbal element and does not allow reliance on long-term memory. Therefore, Wen (2012: 5) argues that “the nonword repetition span task should be a better candidate for measuring the PL in SLA”.

Archibald and Gathercole (2007) enumerate the benefits of using nonword repetition spans over recall spans for the assessment of the PL in the case of language. First, as indicated above, in contrast to traditional span tasks which rely on prior knowledge, including vocabulary and visual information, nonwords represent a processing-based evaluation of the ability to react to new information. Second, nonwords appear to be unaffected by cultural bias and they can be easily adapted to multiple populations. Third, “although verbal short-term memory constrains both nonword repetition and serial recall performance, additional cues inherent in nonword repetition do lead to more accurate recall with greater retention of features of target phonemes” which “appears to be a signature of normal development” (Archibald & Gathercole 2007: 604). All of this makes nonword repetition spans ideal diagnostic tools in evaluating various language-related disorders, in particular SLI (Archibald & Gathercole 2006; Botting et al. 2013; Gray 2003, Lennon & Slesinski 2001). Consequently, such tasks have excellent construct validity.

4. THE STUDY

4.1. Aims

The aim of the study was to establish the reliability and validity of a tool for assessing the PL capacity in the case of Polish adults. As mentioned above, nonword tasks are primarily used for the evaluation of language and cognition development, and as such, they are typically employed with children, especially those who experience difficulties in their education. Such tests usually contain nonwords of increasing length, from one to five syllables, one per trial. However, these cannot reliably be used in studies with adult participants, as a strong ceiling effect is often observed. Yet, some tests, such as the Working Memory Test Battery for Children nonword list subtest (Pickering & Gathercole 2001), follow a different pattern where all nonwords have the same number of syllables, but they are arranged in sets of growing sizes, namely, from one to seven nonwords per set. One advantage of such input sequencing is the fact that no ceiling effect appears during testing all age groups, which led authors of Hi-Lab to include this particular

arrangement in their test battery (Doughty et al. 2010). The PNWSPAN test is based on the same principle as that followed by Pickering and Gathercole (2001) and Doughty et al. (2010), with Polish being the input language.

4.2. Participants

The participants of the validation study were expert judges and English Philology students. Two groups of judges were included: the first included five competent judges, that is, four linguists and a psychologist, and the other consisted of members of the target group, that is, ten English philology students enrolled in the third-year of a three-year BA program. The test was validated with 58 first- and second-year Polish university students of English Philology, 22 males and 36 females, who gave their consent to participate in the study. At the time of the administration of the PNWSPAN, they were all 19–23 years of age, with the mean of 21.6. They had been studying English as a foreign language both in and out of school for three to eleven years, mostly for six to eight years. Their proficiency in English could be described as intermediate, falling in between the B1 and B2 levels according to the *Common European framework of reference*. As part of their BA program, participants attended a number of content classes in English, including those focusing on strategic training, varieties of English, introduction to linguistics and introduction to literary studies, as well as practical English classes, including grammar, pronunciation and the four language skills.

4.2. The test

The test comprises sequences of Polish nonwords, each being a 2-syllable, phonologically likely string of five Polish sounds in the same order, that is cvcvc (e.g., *nomin, gares, mizek*), which is the most popular pattern for 2-syllable words in Polish. All the nonwords are high in phonotactic probability and all have been checked against Polish, English and German languages corpora in order to eliminate any actual words in those three languages. Their wordlikeness was assessed by a group of competent judges (see below). The nonwords are sequenced in sets of 2, 3, 4, 5 and 6, three trials per stage, recorded and played to participants in the order from the shortest level of two items to the longest one consisting of six nonwords, producing a total of 60 nonwords. The participants' task is to repeat the nonwords in the correct order.

4.3. Administration

Similarly to most cognitive tests, the test is taken individually, which ensures a focus on the task. The administration procedure takes about four to five minutes. Before beginning the test, participants are informed of its content and the task they are supposed to perform. A simplified definition of a nonword is presented, followed by the information on set arrangement, and finally immediate serial recall is emphasised. These instructions are accompanied by three trial sets which aim to familiarise participants with the task and to ensure that they understand it thoroughly. The trial sets are presented below:

Wysłuchaj uważnie, a gdy wskażę na Ciebie, powtórz (Listen carefully and, when I point at you, repeat):

soden, ruloj

Teraz kolejna próba. Uważaj (Another trial, concentrate):

Kubor, wonet, mytaf.

I jeszcze jedna próba (And one more):

Gudal, tomis, derap, bawuk.

After the presentation of the trial sets, the actual test starts. The sequences are pre-recorded with the use of Audacity software using a male Polish lector's voice with the consistent speed of one nonword per second, as suggested by Wechsler (1997) and Engel de Abreu and Gathercole (2012). Obviously, to ensure steady rhythm and optimal pronunciation, the lector is required to practise all the nonwords before the actual recording. Each set is followed by a pause, from five to twenty seconds, depending on the length of the set, which is the time for the participant to repeat the stimuli. Each participant is recorded for later analysis since no online scoring is performed. The entire test lasts three minutes.

4.4. Scoring

Most early cognitive tests apply absolute span score procedure, that is, each set is scored as either correct or incorrect. After two or three incorrect trials per level, the test is discontinued, the span score being the last item size (e.g., 4 or 5) recalled. However, the absolute span procedure has numerous disadvantages (Conway et al. 2005; Linck et al. 2014). First, the sensitivity of the measure is very low, as the value is usually between two and six, resulting in poor discrimination among the results. Besides, span reliability may be threatened by the variety in the level of item difficulty, which is a factor difficult to control for. Second, a great deal of data is lost as a conse-

quence of the discontinuation of the test after a failure to repeat a certain list. Third, the item size seems to be insufficient for in-depth analyses because all information on other trials is ignored. To conclude, absolute span measures should be excluded from research on individual differences in favour of partial scoring, where individual elements within each set are assigned points. Thanks to the use of this procedure, floor and ceiling effects are avoided and the results are diversified, which, in turn, ensures a higher level of sensitivity (Conway et al. 2005).

In effect, the result of the PNWSPAN is a partial score, with each element being assigned points in the range of 0–3, depending on the quality of its recall. If an item is recalled correctly, it receives three points. If it is not, the type of error is taken into consideration with a basic distinction being made between order and item errors. The former are most often encountered in serial order research with lists of highly familiar stimuli, such as digits or letters. The latter appear more often in lists of lower frequency words or nonwords (Jeffries, Frankish & Ralph 2006). Also, nonword recall often suffers from item fragmentation, where errors at the level of the phoneme rather than the whole item appear, mostly involving phoneme substitutions that preserve syllabic structure (Archibald & Gathercole 2007; Gathercole, Service, Adams & Martin 1999). In light of this, it might be suggested that while calculating the results, order errors should not be given the same weight as item errors, and many nonwords are partly correct, although one or two phonemes may be altered. Thus, the scoring for the PNWSPAN is as follows:

- 3 points – correct word and correct order
- 2 points – correct word and incorrect order
- 1 point – partly correct word and correct order
- 0 points – lack of answer or partly correct/incorrect word and incorrect order.

Recorded tests are listened to by the researcher and analysed on the phoneme level. To assure score objectivity in this study, 10% of the recordings were assessed independently by two examiners and inter-rater reliability of at least 90% was maintained across all samples with the mean score of .93.

4.5. Analysis

The reliability and validity of the test were verified in two ways, with the stimuli first being evaluated by the judges and then the validation study being undertaken. In order to determine the reliability of the PNWSPAN, the test-retest procedure was implemented, in which case Pearson's r was

tabulated, and Cronbach alpha was calculated in order to establish the internal reliability consistency of the tool.

An attempt was made to determine construct, content and face validity of the tool. As is the case with other tests, construct validity can be tapped by determining convergent and discriminant validity. Nonword span tasks correlate mildly with more traditional simple span tasks and, at the same time, they are moderately related to complex span tasks. Therefore, in order to evaluate the convergent validity of the PNWSPAN and in this way shed light on its construct validity, its results were compared with the scores on: (1) the digit span taken from Wechsler Adult Intelligence Scale, which constitutes a simple span task, (2) the Polish Listening Span (PLSPAN), developed by Zychowicz, Biedroń and Pawlak (2017), and the Polish Reading Span (PRSPAN), developed by Biedroń and Szczepaniak (2012a; 2012b), both of which are examples of complex span tasks. This involved calculating Pearson's correlations between the PNWSPAN and these three tests.

In the case of content validity, the five competent judges (four linguists and one psychologist) were familiarised with the concept of WM and its measurement and then requested to assess each element of the task on a 5-point Likert-scale, where 1 indicated *absolutely not* and 5 *absolutely yes*. In order to obtain the ratings of wordlikeness the judges were asked the following questions: (1) Does this word exist in Polish?, (2) Is it likely to pass for a Polish word?, and (3) Does it sound like a foreign word? After rating all the items in a set, the judges were asked two more questions: (1) Which words in this set sound similar? (Indicate their numbers), and (2) How similar do you feel they are? (Comment). All elements with mean values above 1.5 for Questions 1 and 3, and below 4.5 for Question 2 were eliminated and replaced with new ones, which were also assessed. Every indication of stimuli similarity was accounted for and sets were mixed until no similarities were observed. The level of agreement was established by calculating the Kendall's coefficient of concordance. As regards face validity, 10 third-year students were asked to listen to each set and answer two questions on the same 5-point Likert scale as the one used by the competent judges: (1) Do the breaks between words allow you to immediately decide where one word ends and another begins?, and (2) Is this set possible to pronounce as a whole? Subsequently, they listened to the separate items in the test and answered the following two questions: (1) Can you hear the word well?, and (2) Do you think the word is pronounceable? Also in this case, the level of agreement was determined by tabulating the Kendall's coefficient of concordance. In addition, the assessment of test items was accompanied by a focus session, which gave the students an opportunity to express their opinions concerning the test.

Finally, the sensitivity of the PNWSPAN was also evaluated by analysing the scores obtained by the participants and determining the extent to which floor and ceiling effect could be observed in the data.

4.6. Results

4.6.1. Reliability

As mentioned above, the test-retest method was applied to establish the reliability of the PNWSPAN, with a period of two weeks separating the first and second administration of the test. The results obtained on the two occasions correlated at $r = 0.64$ ($p = 0.05$), which testifies to high reliability of the instrument. Although the discriminating power of several positions within the test was weak, the internal consistency reliability of the tool was satisfactory, as evident in the Cronbach alpha value of 0.68. Those results are consistent with those of previous research on nonword repetition span tasks where these values ranged between 0.49 (Archibald & Gathercole 2007) and 0.78 (Jeffries et al. 2006), which can be taken as evidence that the PNWSPAN is a reliable test.

4.5.2. Validity

The results of the correlational analyses between the PNWSPAN and the other measures were as follows: for the PNWSPAN and the digit span, Pearson's coefficient r equaled 0.32 ($p = 0,019$), in the case of the PNWSPAN and the PLSPAN, Pearson's coefficient r was 0.46 ($p = 0,000$), and for the PNWSPAN and the PRSPAN, Pearson's coefficient r was 0.31 ($p = 0,011$), all of which can be interpreted as low moderate correlations. As expected, the PNWSPAN and the PLSPAN, the two verbal memory tests using aural modality, correlated somewhat better than the others. Moreover, the PNWSPAN and digit span correlated mildly, which is consistent with the research conducted by Gathercole et al. (1999), where the correlation decreased with age, from 0.57 for 5-year-olds to 0.32 for 13-year-olds. Also, the lowest correlation held between the PNWSPAN and the PRSPAN, which was to be expected as these two tests share no common characteristics, except for the fact that they measure WM. On the whole, such results allow us to conclude that although all the tests measure one concept, that is WM, each taps a different aspect of the construct.

When it comes to the assessment made by the five competent judges, Kendall's coefficient of concordance for all the items exceeded 0.9, with the mean value of 0.92. This allows the conclusion that the PNWSPAN is characterized by high content validity. A similar conclusion can be reached in the case of face validity since Kendall's coefficient of concordance for the responses of the ten students amounted to 0.94. As transpired from the focus session, the students considered the test to be very well audible, well timed and not too long, which was the desired outcome. However, the PNWSPAN evoked a wide array of strong emotions. These ranged from amusement (e.g., "Yeah, good memory exercise, so different from anything I've ever done.") to irritation (e.g., "I couldn't focus on the words because I didn't know what they meant, and that was very irritating."), but also from boredom (e.g., "It was like learning linguistics. You can't understand a word but still try to memorise it all.") to disorientation (e.g., "I searched my memory in all languages and still couldn't make sense of the words.", "I have never experienced anything like that. That's probably how I'd feel in the Amazon trying to communicate with some tribes"). Perceptions of the level of difficulty varied widely as well, representing such extremes as "easy, just a few words" and "deadly, I'd never manage more than 3-4 items". The focus session turned into a lengthy discussion concerning the role of memory and its relationship with general cognitive functioning, culminating in a cliché pronouncement: "If you can't understand something you'll never remember it". However, one comment, that is, "I tried to find a helpful strategy to remember the input, but failed", assured the researchers of the validity of the test, as it indicated that PL capacity was the only factor impacting the results.

4.5.3. Sensitivity

The maximum score on the test is 180 points. The mean result during the study was 92.33, the minimum score being 64 and the maximum 126. Such results indicate that no floor or ceiling effect could be observed. They also demonstrate that the PNWSPAN is an accurate and sensitive measure of the PL component of WM.

5. DISCUSSION

The PNWSPAN possesses all the properties of a good research instrument, namely high reliability, sound validity, dependable accuracy and excellent sensitivity. The only unsatisfactory result obtained during the pro-

cess of validation is the low discriminating power of several positions within the test. A possible explanation for this is the very strong primacy and recency effect observed during the analysis, which is consistent with earlier research on serial order tasks. As suggested by Archibald and Gathercole (2007: 588), "Immediate repetition of items for ordered recall forms a classic 'serial position curve' in which recall starts very accurately, decreases throughout the list, and then improves toward the end of the list". In light of the above, the low discriminating power of the tool is typical of serial recall, and as such, should not be treated as a drawback, but rather as an inherent characteristic of this type of measurement. Another problem often appearing in serial recall tasks is the phonological similarity effect (Archibald & Gathercole 2007; Baddeley et al. 1998; Engel de Abreu & Gathercole 2012; Gathercole et al. 1999; Lennon & Slesinski 2001). The phenomenon was taken into consideration in the process of test construction and, since similar words do not appear in close proximity, the results are not affected by it.

6. CONCLUSIONS

The purpose of the study reported in this paper was to validate an instrument which could be used with the Polish population to examine the PL which is regarded the most relevant component of WM in studies of SLA. In accordance with suggestions offered in the literature, we constructed the PNWSPAN, which is a simple span test based on verbal input intended to measure the PL in the case of adults and young adults. The procedures applied to assess the reliability and validity of the test provided evidence that the instrument constitutes a valid and reliable measure of PL, which is a crucial component of WM. Nevertheless, the test suffers from a number of limitations which are typical of cognitive tests of WM. One of these is low discriminating power of some positions, resulting from strong primacy and recency effects. Another problem is the level of difficulty of the words, which should be the same for all the positions in all the tests, a goal which is very difficult to attain as some words are more easily retrievable than others.

What should also be emphasized at this juncture is that while seeking valid, reliable and sensitive measures of different components of WM is commendable, it should be kept in mind that it constitutes just one of a wide array of individual difference factors that mediate the effects of instruction as well as different dimensions of L2 knowledge that may be the outcome of instructed and uninstructed learning. For this reason, there is a need to examine the role of different facets of WM, in particular the PL and CE, in con-

junction with other moderating variables, such as motivation, willingness to communicate, beliefs or strategies, as only then will it be possible to obtain a more complete picture of influences on the development of explicit and implicit (highly automatized) L2 knowledge (DeKeyser 2010; Ellis 2009). This means that while the study of WM is of paramount importance, it should run in parallel to investigations of other aspects of individual variation, be they cognitive, affective and social, or constitute amalgams of these.

REFERENCES

- Abu-Rabia, S. (2003). The influence of working memory on reading and creative writing processes in a second language. *Educational Psychology*, 23, 209–222. DOI: 10.1080/01443410303227.
- Ahmadian, M. J. (2012). The relationship between working memory and oral production under task-based careful online planning condition. *TESOL Quarterly*, 46 (1), 165–175. DOI: 10.1002/tesq.8.
- Alptekin, C. / Erçetin, G. (2011). The effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *TESOL Quarterly*, 45, 235–266. DOI: 10.5054/tq.2011.247705.
- Archibald, L. M. / Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, 41 (6), 675–693. DOI: 10.1080/13682820500442602.
- Archibald, L. M. / Gathercole, S. E. (2007). Nonword repetition and serial recall: Equivalent measures of verbal short-term memory? *Applied Psycholinguistics*, 28 (4), 587–606. DOI: 10.1017/S0142716407070324.
- Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4, 417–123. DOI: 10.1016/S1364-6613(00)01538-2.
- Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36, 189–208. DOI: 10.1016/S0021-9924(03)00019-4.
- Baddeley, A. D. / Gathercole, S. / Papagno, C. (1998). The phonological loop as a language acquisition device. *Psychological Review*, 105, 158–173. DOI: 10.1037/0033-295X.105.1.158.
- Baddeley, A. D. / Hitch, G. J. (1974). Working memory. In: G. Bower (ed.), *The psychology of learning and motivation* (vol. 8, pp. 47–90). New York, NY: Academic Press.
- Berquist, B. (1997). Individual differences in working memory span and L2 proficiency: Capacity or processing efficiency? In: A. Sorace / C. Heccock / R. Shillcock (eds.), *Proceedings of the GALA' 1997 Conference on language acquisition* (pp. 468–473). Edinburgh: The University of Edinburgh.
- Biedroń, A. / Pawlak, M. (2016). The interface between research on individual difference variables and teaching practice: The case of cognitive factors and personality. *Studies in Second Language Learning and Teaching (SLLT)*, 6 (3), 395–422. DOI: 10.14746/ssllt.2016.6.3.3 2016.
- Biedroń, A. / Szczepaniak, A. (2012a). Polish reading span test – an instrument for measuring verbal working memory capacity. In: J. Badio / J. Kosecki (eds.), *Cognitive processes in language. Lodz Studies in Language* (vol. 25, pp. 29–37). Frankfurt am Main, Berlin: Peter Lang.
- Biedroń, A. / Szczepaniak, A. (2012b). Working-memory and short-term memory abilities in accomplished multilinguals. *Modern Language Journal*, 96, 290–306. DOI: 10.1111/j.1540-4781.2012.01332.x.

- Bolibaugh, C. / Foster, P. (2013). Memory-based aptitude for nativelike selection: The role of phonological short-term memory. In: G. Granena / M. Long (eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 203–228). Amsterdam: John Benjamins.
- Botting, N. / Psarou, P. / Caplin, T. / Nevin, L. (2013). Short-term memory skills in children with specific language impairment. The effect of verbal and nonverbal task content. *Topics in Language Disorders*, 33 (4), 313–327. DOI: 10.1097/01.TLD.0000437940.01237.51.
- Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology*, 32 (5), 867–873. DOI: 10.1037/0012-1649.32.5.867.
- Conway, A. R. A. / Jarrold, Ch. / Kane, M. J. / Miyake, A. / Towse, J. N. (2007). Variation in working memory. An introduction. In: A. R. A. Conway / Ch. Jarrold / M. J. Kane / A. Miyake / J. N. Towse (eds.), *Variation in working memory* (pp. 3–17). Oxford: Oxford University Press.
- Conway, A. R. A. / Kane, M. J. / Bunting, M. F. / Hambrick, D. Z. / Wilhelm, O. / Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786. DOI: 10.3758/BF03196772.
- Conway, A. R. A. / Macnamara, B. / Engel de Abreu, P. (2013). Working memory and intelligence: An overview. In: T. P. Alloway / R. G. Alloway (eds.), *Working memory: The new intelligence* (pp. 13–36). New York, NY: Psychology Press.
- DeKeyser, R. M. (2010). Cognitive-psychological processes in second language learning. In: M. H. Long / C. J. Doughty (eds.), *The handbook of language teaching* (pp. 119–138). Oxford: Wiley-Blackwell.
- 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). Mahwah, NJ: Lawrence Erlbaum.
- DeKeyser, R. M. / Koeth, J. (2011). Cognitive aptitudes for second language learning. In: E. Hinkel (ed.), *Handbook of research in second language teaching and learning* (pp. 395–407). New York, NY: Routledge.
- Dollaghan, C. / Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41, 1136–1146. DOI: 10.1044/jslhr.4105.1136.
- Doughty, C. J. (2013). Optimizing post-critical-period language learning. In: G. Granena / M. H. Long (eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 153–175). Amsterdam: John Benjamins.
- Doughty, C. J. / Campbell, S. G. / Mislevy, M. A. / Bunting, M. F. / Bowles, A. R. / Koeth, J. T. (2010). Predicting near-native ability: The factor structure and reliability of Hi-LAB. In: M. T. Prior / Y. Watanabe / S.-K. Lee (eds.), *Selected proceedings of the 2008 Second Language Research Forum* (pp. 10–31). Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2382.
- 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). Bristol: Multilingual Matters.
- Ellis, N. C. / Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology*, 49A (1), 234–250.
- Engel de Abreu, P. M. J. / Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology*, 104 (4), 974–986. DOI: 10.1037/a0028390.
- Fortkamp, M. B. M. (1999). Working memory capacity and aspects of L2 speech production. *Communication and Cognition*, 32, 259–296. DOI: 10.17851/2237-2083.16.2.129-144.

- Fortkamp, M. B. M. (2003). Working memory capacity and fluency, accuracy, complexity and lexical density in L2 speech production. *Fragmentos*, 24, 69–104. DOI: 10.5007/fragmentos.v24i0.7659.
- Foster, P. / Bolibaug, C. / Kotula, A. (2014). Knowledge of native like selections in an L2: The influence of exposure, memory, age of onset and motivation in foreign language and immersion settings. *Studies in Second Language Acquisition*, 36 (1), 101–132. DOI: 10.1017/S0272263113000624.
- French, L. M. / O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, 29, 463–487. DOI: 10.1017/S0142716408080211.
- Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, 23 (1), 83–94. DOI: 10.3758/BF03210559.
- Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27 (4), 513–543. DOI: 10.1017/S0142716406060383.
- Gathercole, S. E. / Hitch, G. J. / Service, E. / Martin, A. J. (1997). Short-term memory and new word learning in children. *Developmental Psychology*, 33 (6), 966–979. DOI: 10.1037/0012-1649.33.6.966.
- Gathercole, S. E. / Service, E. / Hitch, G. J. / Adams, A.-M. / Martin, A. J. (1999). Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Applied Cognitive Psychology*, 13 (1), 65–77. DOI: 10.1002/(SICI)1099-0720(199902)13:1<65::AID-ACP548>3.0.CO;2-O.
- Gathercole, S. E. / Willis, C. S., Baddeley, A. D. / Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, 2 (2), 103–127. DOI: 10.1080/09658219408258940;
- Gray, S. (2003). Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, 36 (3), 129–151. DOI: 10.1016/S0021-9924(03)00003-0.
- Harrington, M. / Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, 14, 25–38. DOI: 10.1017/S0272263100010457.
- Jefferies, E. / Frankish, C. R. / Ralph, M. A. L. (2006). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language*, 54 (1), 81–98. DOI: 10.1016/j.jml.2005.08.001.
- Juffs, A. / Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, 44, 137–166. DOI: 10.1017/S0261444810000509.
- Kane, M. J. / Conway, A. R. A. / Hambrick, D. Z. / Engle, R. W. (2008). Variation in working memory capacity as variation in executive attention and control. In: A. R. A. Conway / Ch. Jarrold / M. J. Kane / A. Miyake / J. N. Towse (eds.), *Variation in working memory* (pp. 21–49). Oxford: Oxford University Press.
- Kormos, J. / Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, 11 (2), 261–271. DOI: 10.1017/S1366728908003416.
- Leeser, M. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning*, 57 (2), 229–270. DOI: 10.1111/j.1467-9922.2007.00408.
- Lennon, J. / Slesinski, C. (2001). Comprehensive test of phonological processing (CTOPP): Cognitive-linguistic assessment of severe reading problems. *Communique*, 29, 38–40.
- 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. DOI: 10.3758/s13423-013-0565-2.

- Mackey, A. / Philip, J. / Egi, T. / Fujii, A. / Tatsumi, T. (2002). Individual differences in working memory, noticing interactional feedback and L2 development. In: P. Robinson (ed.), *Individual differences and instructed language learning* (pp. 181–209). Philadelphia, PA: John Benjamins.
- 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. DOI: 10.1017/S0272263112000125.
- Miyake, A. / Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In: A. Healy / L. Bourne (eds.), *Foreign language learning* (pp. 339–364). Mahwah, NJ: Lawrence Erlbaum.
- 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, 377–402. DOI: 10.1017/S0142716406060322.
- O'Brien, I. / Segalowitz, N. / Collentine, J. / Freed, B. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, 29, 557–582. DOI: 10+10170S027226310707043X.
- Papagno, C. / Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology*, 38A, 98–107. DOI: 10.1080/14640749508401378.
- Pawlak, M. (2017). Overview of learner individual differences and their mediating effects on the process and outcome of interaction. In: L. Gurzynski-Weiss (ed.), *Expanding individual difference research in the interaction approach. Investigating learners, instructors, and other interlocutors*. Amsterdam, Philadelphia: John Benjamins.
- Payne, J. S. / Whitney, P. J. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal*, 20, 7–32. www.learntechlib.org/p/95531/.
- Pickering, S. J. / Gathercole, S. E. (2001). *The working memory test battery for children*. Hove: The Psychological Corporation.
- Robinson, P. (2003). Attention and memory during SLA. In: C. J. Doughty / M. H. Long (eds.), *The handbook of second language acquisition* (pp. 631–679). Oxford: Blackwell Publishing.
- Sawyer, M. / Ranta, L. (2001). Aptitude, individual differences, and instructional design. In: P. Robinson (ed.), *Cognition and second language instruction* (pp. 319–354). Cambridge: Cambridge University Press.
- Service, E. (1992). Phonology, working memory and foreign-language learning. *Quarterly Journal of Experimental Psychology*, 45 (1), 21–50. DOI: 10.1080/14640749208401314.
- Skehan, P. (2012). Language aptitude. In: S. Gass / A. Mackey (eds.), *Routledge handbook of second language acquisition* (pp. 381–395). New York, NY: Routledge.
- Speciale, G. / Ellis, N. C. / Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, 25 (2), 293–321. DOI: 10.1017/S0142716404001146.
- Suzuki, Y. / DeKeyser, R. (2017). Exploratory research on second language practice distribution: An Aptitude × Treatment Interaction. *Applied Psycholinguistics*, 38 (1), 27–56. DOI: 10.1017/S0142716416000084.
- Wechsler, D. (1997). *Manual for the Wechsler adult intelligence scale*. Third edition (WAIS III). San Antonio, Tx: The Psychological Corporation.
- Wen, E. Z. (2012). Working memory and second language learning. *International Journal of Applied Linguistics*, 22, 1–22. DOI: 10.1111/j.1473-4192.2011.00290.
- Wen, E. Z. (2015). Working memory in second language acquisition and processing: The phonological/executive model. In: E. Z. Wen / M. B. Mota / A. McNeill (eds.), *Working memory in second language acquisition and processing* (pp. 41–62). Bristol: Multilingual Matters.

- Wen, E. Z. (2016). *Working memory and second language learning: Towards an integrated approach*. Bristol: Multilingual Matters.
- Wen, E. Z. / Biedroń, A. / Skehan, P. (2016). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, 50 (1), 1–31. DOI: 10.1017/S0261444816000276.
- Wen, E. Z. / Mota, M. B. / McNeill, A. (eds.) (2015). *Working memory in second language acquisition and processing*. Bristol: Multilingual Matters.
- Wen, E. Z. / Skehan, P. (2011). A new perspective on foreign language aptitude: Building and supporting a case for “working memory as language aptitude”. *Ilha Do Desterro: A Journal of English Language, Literatures and Cultural Studies*, 60, 15–44. DOI: 10.5007/2175-8026.2011n60p015.
- Williams, J. N. (2012). Working memory and SLA. In: S. Gass / A. Mackey (eds.), *Handbook of second language acquisition* (pp. 427–441). Oxford: Routledge/Taylor & Francis.
- Williams, J. N. / Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning*, 53, 67–121. DOI: 10.1111/1467-9922.00211.
- 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 SSLT*, 7 (4), 601–618. DOI: 10.14746/sslst.2017.7.4.3 2017.

Received: 1.08.2018; **revised:** 10.09.2018

