UNIVERSITY OF JYVÄSKYLÄ

NONWORD REPETITION AND L2 KNOWLEDGE

A Pro Gradu Thesis in English

by

Helena Miettinen

Department of Languages
2003

HUMANISTINEN TIEDEKUNTA
KIELTEN LAITOS

Helena Miettinen
NONWORD REPETITION AND L2 KNOWLEDGE

Pro Gradu –tutkielma

Englannin kieli
Maaliskuu 2003 102 sivua + 12 liitettä

Tutkielma on osa Jyväskylän yliopiston Soveltavan kielentutkimuksen keskuksen hanketta Tilanteinen kielellinen tietoisuus ja vieraan kielen oppiminen. Projektissa on vuodesta 1998 lähtien tutkittu lasten kielellistä tietoisuutta vygotskilaisessa sosiokulttuurisessa ja bahtinilaisessa dialogisessa viitekehyksessä. Tutkimusprojekti seuraa ryhmää lapsia läpi perusopetuksen luokkien 1—6. Tutkimushenkilöt (N=16) ovat opiskelleet englantia kolmannelta luokalta lähtien ja he olivat tutkielmaa tehtäessä viidennellä luokalla. Tutkielmassa käytetään osittain hankkeessa jo saatavilla olevaa aineistoa.

Tutkielman tarkoituksena on selvittää lasten epäsanantoiston ja vieraan kielen oppimisen yhteyttä, fonologisen työmuistin kielispesifisyyttä ja kehittymistä. Epäsanat näyttävät ja kuulostavat yleensä jonkin tietyn kielen sanoilta, mutta ne on keksitty jotain testiä tai tehtävää varten. Tyypillisessä epäsanatestissä tutkimushenkilö toistaa kuulemansa epäsanat välittömästi ja mahdollisimman tarkasti ja saa pisteitä oikeiden suoritusten perusteella. Testin taustalla on työmuistimalli, jonka yhtä osaa, fonologista työmuistia, epäsanatestin on tarkoitus operationaalistaa. Oletuksena on, että vaikeudet epäsanojen toistamisessa ovat johdettavissa työmuistin ongelmiin tai kehittymättömyyteen. Kun epäsanatestin tuloksia verrataan kielitaitoon, voidaan tehdä päätelmiä fonologisen työmuistin ja kielen oppimisen yhteyksistä. Useissa tutkimuksissa on saatu näyttöä tällaisen yhteyden olemassaolosta.

Tutkielman aineisto kerättiin kahden suomenkielisen ja yhden englanninkielisen epäsanatestin avulla, sekä erinäisillä kielitaitotehtävillä. Tutkimushenkilöt osallistuivat suomenkieliseen epäsanatestiin jo toisella luokalla, mutta varsinaiset tämän tutkielman epäsanatestit ja kielitaidon arviointiin tarvittavat tehtävät tehtiin viidennellä luokalla, loppuvuodesta 2002. Testejä verrattiin toisiinsa korrelaatioiden (Spearman) avulla, ja tulokset olivat odotetunlaisia. Englanninkielinen epäsanatesti oli yhteydessä englannin kielen taitoon (? = 0,765**) muttei suomenkielisiin epäsanatesteihin (? = 0,416 ja ? = 0,248). Suomenkieliset epäsanatestit puolestaan korreloivat keskenään tilastollisesti merkitsevästi (? = 0,512*). Tulokset viittaavat siihen, että fonologinen työmuisti on yhteydessä vieraan kielen taitoon. Lisäksi fonologinen työmuisti näyttäisi olevan kielispesifi, eli se toimii eri tavalla käytettäessä eri kieliä.

Asiasanat: phonological working memory, nonword repetition, L2 knowledge

CONTENTS

1 INTRODUCTION

Research indicates that memory and language are closely connected. Memory is a widely studied concept and there are several theories about how it works. One of these theories includes a construct called phonological working memory, which is believed to be an essential part of language learning. It allegedly functions as a gateway for linguistic input to get to long-term memory. Researchers in the field of cognitive psychology have developed and used nonword repetition tests as one way to operationalize their subjects' phonological working memory. The success in nonword repetition tests has then been used to study the connection between phonological working memory and different aspects of language learning. This relationship has been investigated from several angles: Subjects have included children, adults, normal subjects, subjects with memory or language deficiencies, and those learning their first language (L1) or second/foreign language (L2), while the focus of research has ranged from the general development of phonological working memory to its language-specific nature. Studies with L2 learners are a small minority among this research.

The present study is part of a larger research project titled Situated metalinguistic awareness and foreign language learning (http://www.solki.jyu.fi/meta). It was launched in 1998 at the Centre for Applied Language Studies of the University of Jyväskylä and it is headed by Academy Research Fellow Riikka Alanen. The project focuses on the development of language awareness of a group of children through grades 1–6. So far the project has produced several papers in international conferences, publications, and master's theses (e.g. Alanen 2000, Aro 2001, Dufva, Alanen, and Mäntylä 2001). Future publications include book chapters by Alanen (in press) and Dufva (in press). When the participants first began school in 1998, they were six

or seven years old. Originally, there were 20 of them in the research group, but at the moment there are 16 left. At present, the participants are in the fifth grade and 11 to 12 years old. They are Finnish and they speak Finnish as their L1. They began studying English as an L2 in the third grade in the year 2000.

The participants' metalinguistic development is studied in a Vygotskian socio-cultural and Bakhtinian dialogical framework. Both qualitative and quantitative methods are used in analyzing the data, which consist of, for example, interviews, classroom observation, various language tasks and quantitative tasks, one of which is nonword repetition. The aim, however, is not in generalizations, but in studying the children's metalinguistic development as individuals. The emphasis is on the situated nature of the participants' metalinguistic awareness, what kind of an effect it has on L2 learning, and how these children see their L1 and the L2 they are learning. The present study uses partly data that are already available in the project.

A Finnish nonword repetition test was carried out as part of this longitudinal research project in May 2000, partly to test the participants' phonological working memory, but also to see whether a Finnish nonword repetition test could yield interpretable results. The experience from previous research seemed to be that Finnish nonwords could not be used because they are too easy for the subjects.

The present study is a continuation of that first Finnish nonword repetition test. Since it seemed to be possible to test phonological working memory with Finnish nonwords when the participants were in the second grade, it made sense to pursue this type of testing. The longitudinal project offered a chance to test the same participants again in the fifth grade with Finnish nonwords as well as English ones and make comparisons between the scores.

Since several previous studies had failed to get results with Finnish nonwords, the main concern in the fifth grade still was to

construct a good Finnish nonword repetition test. Even though the second grade one seemed to work very well, the participants were now considerably older and age was known to be an important factor in success in nonword repetition. The Finnish repetition test was thus a major priority. The aim was then to compare the two Finnish nonword repetition tests to try to discover if the same participants were performing well or poorly at both times, or if the development of phonological working memory had been different within the group of participants.

What most often happens in studies using nonword repetition tests is comparison between the repetition scores and language knowledge. The purpose then is to see if phonological working memory and language really are connected. There is plenty of evidence to indicate a clear connection between phonological working memory and L1, but there is not yet very much research on the connection between this memory construct and knowledge of L2. From memory research point of view L2 studies are important for defining phonological working memory more clearly, but they also contribute to the understanding of L2 learning processes. If the connection between nonword repetition and L2 knowledge is not as clear as in the case of L1, some other mechanisms must be available for L2 learners to utilize. Furthermore, comparing different language nonword repetition results to knowledge of those same languages best reveals the differences between the connections, i.e. if phonological working memory shows signs of language-specificity by functioning differently in contexts where knowledge of languages other than L1 are assessed.

One objective of the present study was to add to the L2 studies of phonological working memory. The participants have been learning English as an L2 since third grade, so it made sense to test their success in repeating English-sounding nonwords as well. It was also essential to construct a measure of knowledge of English that was more

comprehensive than in similar earlier studies, since most previous research seems to have focused on vocabulary knowledge. A broader measure of English was achieved by constructing some measures specifically for the purposes of the present study, but also by utilizing other tasks of the larger research project and those done as part of regular English classes. Once there was a measure of knowledge of English, it could be compared to English nonword repetition success, but also to Finnish nonword repetition to test the claims made by some researchers about the language-specificity of phonological working memory.

To summarize, the present study attempted to answer three general research questions: Is children's knowledge of an L2 connected to their success in repeating nonwords sounding like that L2? Is phonological working memory language-specific? Are the same children good at repeating Finnish nonwords in the fifth grade as in the second grade?

The issues mentioned above are studied quantitatively, since all these tasks and tests can be reduced to number scores. The connection between English nonword repetition and knowledge of English, the language-specificity of phonological working memory, and the nature of phonological working memory development are all studied using statistical analysis by calculating correlations between the scores, to be exact. Since the number of the participants is rather small, more sophisticated methods of analysis were thought not to yield reliable results and thus were not used.

To begin with, chapter 2 introduces previous research carried out into the relationship between phonological working memory and language, L1 as well as L2. It also briefly describes the memory model, which the concept of phonological working memory and the nonword repetition test are based on. The design of the present study and the research questions and hypotheses are described in more detail in

chapter 3. Also the participants, the tests and measures and processing of the data are described in the same chapter. The major results from the correlation test are reported in chapter 4, and discussed and compared to previous research in chapter 5. Concluding remarks are made in chapter 6.

2 PREVIOUS RESEARCH ON PHONOLOGICAL WORKING MEMORY AND LANGUAGE

For quite some time researchers have been trying to explain how human memory operates and how it is structured. Several theories and models of memory exist and in a way they are in competition with each other. The focus in this chapter is on one alleged aspect of memory, phonological working memory, which is an element of a working memory model. The nonword repetition tests are based on this working memory model, so its structure and functioning are described first. Then the idea behind nonwords and the structure of a basic nonword repetition test are explained in brief, and some suggestions are made for why nonword repetition tests are needed. This theoretical basis is followed by a review of two types of studies, both of which have used nonword repetition tasks to study the relationship between phonological working memory and language learning. Studies focusing on the different elements of L1 (vocabulary, reading etc.) are reviewed first and those focusing on L2 learning after that.

2.1 Phonological Working Memory

Since it is (apparently) impossible to physically locate memory in the brain and actually see it in action, we have to rely on models, or hypotheses of what memory is and how it processes information. One such model is the working memory model of Baddeley and Hitch. It was first published in 1974 but has been revised since (for further details, see Baddeley 1986). The essence of working memory is that while people are involved in a cognitive task such as learning, for

example, the information is held and processed in working memory (Baddeley 1986:33-34).

The working memory model comprises three components: the central executive, the visuo-spatial sketchpad, and the phonological loop, as illustrated in Figure 1 (Baddeley and Hitch 1974, Baddeley 1986, Baddeley 1992, Gathercole and Baddeley 1993a:4). According to Gathercole and Baddeley (1993a:4-5), the central executive is the most important part and the other two are "slave systems" to it. The central executive has many functions, but limited capacity to do everything. It regulates how information flows in the working memory system, gets information from other parts of memory, e.g. from long-term memory, and it processes and also stores information. The more there is for the central executive to handle simultaneously, the less efficient it is in its functioning.

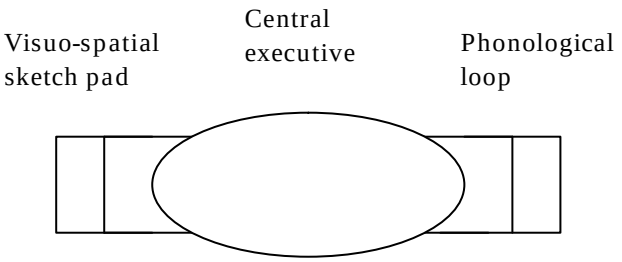


Figure 1. The working memory model as illustrated in Gathercole and Baddeley (1993a:4).

The two slave systems of the central executive deal with two different types of information (Gathercole and Baddeley 1993a:4). The visuo-spatial sketchpad processes and maintains visual and spatial information whereas the phonological loop deals with verbal information, the last-mentioned being the part also referred to as phonological working memory (Lehto 1996:15, Gathercole and Baddeley 1993a). It was to investigate the nature of phonological

working memory that the nonword repetition test (discussed in section 2.2) was developed.

The phonological loop, in turn, consists of two components: phonological store and articulatory rehearsal process. The structure of the phonological loop is illustrated in Figure 2. It also shows how speech input has direct access to the phonological store but nonspeech input only gains access to it through the articulatory rehearsal process. If a stored phonological code is not kept fresh through the rehearsal process, it gradually disappears from the phonological store (Baddeley 1986:84, Gathercole and Baddeley 1993a:8). There are several phenomena associated with and supporting the hypothetical phonological loop component of working memory, for example the acoustic similarity effect, the irrelevant speech effect, the word-length effect, and articulatory suppression (Baddeley 1992:558).

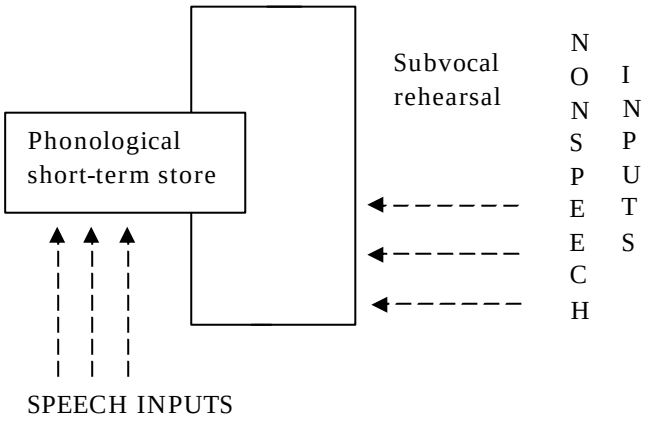


Figure 2. The phonological loop model as illustrated in Gathercole and Baddeley (1993a:8).

According to Gathercole and Baddeley (1993a:25–26), working memory is something that already exists in newborns and develops as people get older. This is based on research indicating that the amount of verbal material that people can hold in their memory for short periods of time increases tremendously from early childhood to adulthood. The likely

way that this development happens is that the operating efficiency of each of the three components of working memory will grow and people learn to use different strategies more efficiently to maximize the functioning of working memory.

As to phonological working memory development specifically, according to Gathercole, Willis, Emslie, and Baddeley (1991:365), the early school years are a time of rapid development in phonological working memory and it reaches adult level at about the age of 12. Gathercole and Baddeley (1993a:26–31) contribute this development to changes in the rate of subvocal rehearsal. While studies indicate that 4-year-olds have a fully functioning phonological loop (that is, phonological working memory), the rehearsal component of it does not get very efficient until children are older. In fact, the rate of subvocal rehearsal will increase as speaking rate increases. These increases enable more information to be held in the phonological loop without the information being deteriorated.

As Gathercole and Baddeley (1993a:70–73) put it, a properly functioning phonological working memory is a key to the long-term learning of languages. When learning new vocabulary, for example, the phonological information is first and temporarily held in phonological working memory and then transferred “into some more permanent knowledge structure in the lexical-semantic memory system” (Gathercole and Baddeley (1993a:71). If the temporary trace in the phonological working memory is not distinct and durable enough, it is unlikely or at least difficult to form a more permanent trace in the long-term memory. If there are problems with the phonological working memory, the phonological material can either be encoded wrongly or it can be lost quickly (Gathercole and Baddeley (1993a:71).

2.2 The Nonword Repetition Test

Nonword repetition has been used for at least two decades by psychologists to operationalize phonological working memory. Countless variations of the test can be found in literature, but the Children's test of nonword repetition (CNRep) by Gathercole, Willis, Baddeley, and Emslie (1994) is a widely used and translated one. The following description of the basic form of a nonword repetition task is largely based on an article by Gathercole et al. (1994) describing CNRep.

First of all, nonwords are phonological items that conform to the phonotactics of some language, but are not real words with any meaning. Which language they are modeling depends on the purposes for which they are used. Usually it is the subjects' L1 or the L2 they are learning.

In a typical nonword repetition task the subject hears taped or live nonwords and is asked to attempt to repeat each item immediately after hearing it. The number and length of the nonwords depends on the research agenda, but in CNRep the children repeat a total of 40 nonwords, ten of each length between one and four syllables (two to five syllables in later versions of the test, e.g. Masoura and Gathercole 1999). Usually there are a few practice items before the actual test.

The repetition attempts are taped and then scored, usually by giving participants either one point per nonword or a point per syllable. The score represents the subjects' phonological working memory capacity and it is usually used to calculate correlations with scores on aspects of language knowledge such as vocabulary. Gathercole and Baddeley (1993a:48–49) give two reasons for using a nonword repetition test to measure phonological working memory. Firstly, it is possibly more sensitive than e.g. digit span, since subjects have no long-term lexical representations that match nonwords. Therefore, long-term

memory knowledge does not influence the results that much, even though its effect cannot be completely eliminated even with nonwords. Secondly, the nonword repetition test is thought to be simple, since hearing and repeating new and strange words is something natural and common to children and everyone else learning a language. What adds to its simplicity is that it is not sensitive to the use of "higher-level strategic processes" (Gathercole and Baddeley 1993a:49) like cumulative rehearsal, which digit span demands.

A number of studies using some form of the nonword repetition test are reviewed in the following two sections. They serve to illustrate the great number of ways in which nonword repetition has been used to evaluate phonological working memory and the great variety of objectives embodied in such research. It should be mentioned that there were some problems with terminology. It was not always clear in the studies reviewed below what it was exactly that was being measured with the nonword repetition task. Even within a single research article there could be several different terms used to refer to the memory system supposedly under study. It appeared that the researchers did associate the task with a certain memory model (namely that by Baddeley and Hitch 1974), but varied the terms in their writing freely. The terms used in a particular study are also used in the present study when referring to that particular study. When speaking in general, the intent has been only to use the term phonological working memory.

2.3 Nonword Repetition and L1 Learning

Since the present study deals with nonword repetition in the context of L2 learning, the emphasis in reviewing previous research is on the studies of the connection between nonword repetition and L2 learning. However, a clear majority of the studies using nonword repetition are

interested in the different aspects of L1 learning and so deserve a brief introduction here. Only the most widely studied areas of such research are covered and only the main results are reported. The research methods and especially the nonword tasks used are not described extensively until section 2.4, because the emphasis in the present study is on L2 learning and the relevant methodological issues will be dealt with in the section focusing on L2 learning studies.

Studying L1 development often means studying young children: speech development, vocabulary, learning to read, and various aspects of language impairment. Of these the extent and development of L1 vocabulary has drawn the most attention.

2.3.1 Vocabulary Development

Researchers disagree to some extent on the relationship between vocabulary knowledge and phonological working memory. There are studies indicating a strong connection and also those denying it. Some studies take the link between nonword repetition and vocabulary knowledge as a fact and study the issue further. The possible relationship remains a matter of debate.

Perhaps one of the first studies to systematically look for and find a significant correlation between vocabulary development and phonological working memory was conducted by Gathercole and Baddeley (1989). Examples of more recent studies coming to the same conclusion are Avons, Wragg, Cupples, and Lovegrove (1998) and Gathercole, Service, Hitch, Adams, and Martin (1999). There are also a few studies that especially looked whether nonword repetition test could be used to predict later vocabulary knowledge or vice versa, and also studied the causal links between the tests and vocabulary

knowledge (Gathercole, Willis, Emslie, and Baddeley 1992 and Michas and Henry 1994).

Among those who have studied and found a connection between nonword repetition and vocabulary, most have examined the existing L1 vocabulary of their subjects. Some, however, have been more creative. Michas and Henry (1994), for example, added to their study a task where subjects were taught three new words and three definitions for each of them. Gathercole, Hitch, Service, and Martin (1997) included in their study four vocabulary learning tasks, three of which tested the learning of nonwords.

Several researchers have attempted to clarify what it is that nonword repetition really measures, since such a strong relationship seems to exist between vocabulary and the ability to repeat nonwords. According to Gathercole, Willis, Emslie, and Baddeley (1991) the two most apparent reasons for this connection are phonological working memory capacity and existing vocabulary knowledge. That is to say that repetition scores correlate with vocabulary because both of them benefit from good working memory capacity or because good existing vocabulary also aids nonword repetition.

There are several further factors that have been studied – with conflicting results – as possible sources for the connection between nonword repetition and vocabulary knowledge. Snowling, Chiat, and Hulme (1991) suggested other more complicated phonological processes besides memory, such as phonological segmentation, assembly of articulatory instruction, perceptual problems, and difficulties with the storage of phonological information. They also drew attention to the importance of the prosodic structure of nonwords, and the aiding influence of affixes. Gathercole, Willis, and Baddeley (1991) dismissed the influence of these nonmemory processes based on their earlier work. Wordlikeness of nonwords was found to have a positive influence on the relationship (Gathercole 1995), lexical

morphemes integrated into nonwords can have either a positive or a negative effect (Dollaghan, Biber, and Campbell 1995), and speech output skills do not explain the link between nonword repetition and vocabulary skills (Gathercole, Service, Hitch, Adams, and Martin 1999).

Not all believe that the connection between nonword repetition and vocabulary knowledge has to do with phonological working memory. Metsala (1999) explained the relationship with the "underlying structure of lexical items" (Metsala 1999:11) and claimed that phonological awareness mediated the relationship. Bowey (1996), too, denied a direct causal role of phonological memory on vocabulary acquisition. She claimed that there was a latent phonological processing factor working behind phonological memory and phonological sensitivity alike.

In summary, there is much evidence to indicate that there, indeed, is a relationship between vocabulary development and phonological working memory as measured by the nonword repetition test although some researchers disagree. Some doubt whether nonword repetition is a very good test of phonological working memory, while others question the role of phonological working memory in vocabulary development altogether, regardless of how it is measured. If we accept that nonword repetition and vocabulary development are connected, there are still many possibly intervening elements to consider. Based on the research reviewed above, it is not exactly clear what lies behind the apparent link between phonological memory and vocabulary development and conclusions based on nonword repetition tests need to made with care.

2.3.2 Reading

The nonword repetition test has also been used in studies investigating the development of L1 reading. Based on these few studies, there appear to be several ways in which L1 reading may be affected by (impaired) phonological working memory.

According to Brady, Poggie, and Rapala (1989), poor readers do not suffer from poor vocabulary but have problems in forming accurate phonological representations. This causes encoding difficulties, which in turn lead to memory difficulties and poorer reading. Gathercole and Baddeley (1993b) measured their subjects' phonological working memory with nonword repetition before they were able to read and found it to predict reading skill in childhood. Gathercole (1995) reported similar findings from comparing her subjects' nonword repetition at age four to their reading at age five. The phonological working memory skills were important in early reading development, but they no longer had such an influence after one year of reading.

Also Muter (1998) found nonword repetition (at ages five and six) to be a long-term predictor of reading (at age nine). Nonword repetition also discriminated between good and poor readers. Van Bon and van der Pijl (1997) were the only ones to doubt that the differences in repetition scores – which they also found in their study of good and poor readers – were due to differences in phonological working memory. They were inclined to explain the results with the subjects' differing phonological segmentation abilities instead.

2.3.3 Language Impairment

Another area in which nonword repetition tests have been used quite extensively - and the last area discussed here - are studies concerning

the connection between language or general learning impairment and phonological working memory capacity. First of all, nonword repetition tests have been found to discriminate between learning-disabled and nondisabled children (Taylor, Lean, and Schwartz 1989). As to difficulties with language, twin-studies of children with persistent language impairment have indicated that problems in repeating nonwords – and therefore also in phonological working memory – are heritable (Bishop, North, and Donlan 1996). Also children with specific language impairment, whether in intervention or not, have been found to repeat nonwords less successfully that normal children (Weismer, Tomblin, Zhang, Buckwalter, Chynoweth, and Jones 2000). The same study also indicated that nonword repetition can be used as a tool to rule in or out language disorder, but it is not adequate alone. A study of dyslexics, on the other hand, revealed that even though the subjects had shown impaired performance in nonword repetition when they began school, their problems with learning pseudonames at age nine could not be explained with poor phonological working memory (Mayringer and Wimmer 2000).

The studies reviewed in sections 2.3.1–2.3.3 have been quite homogenous in many ways. Since the focus has been on L1 development, it is only natural that the subjects have been quite young. Most of them have also been native speakers of English. In the next section the focus shifts from L1 to L2 learning and along with it the subjects and languages become quite varied.

2.4 Nonword Repetition and L2 Learning

Studies focusing on the relationship between nonword repetition tasks and L2 learning are relatively scarce. Furthermore, those existing are

somewhat difficult to compare due to different methodology and various other aspects such as language, subjects, and the actual nonword task used. Even though the main focus in the present study is on nonword repetition test as a means to assess learners' working memory, studies using methods such as nonword span or some of the more traditional short-term memory span measures are also reviewed below. After a study-by-study review the findings as well as the methods used in the field of nonword repetition and L2 learning are compared and summarized.

2.4.1 Children Learning English as an L2

Service 1989. What in the first place inspired the inclusion of the nonword repetition task in the larger research project that the present study is part of was a dissertation by Service (1989, 1992). She conducted a longitudinal study to examine the effect of the phonological working memory system on L2 learning. More particularly, one section of the dissertation deals with the ability of a pseudoword repetition task to predict the success of learning English (Service 1989:43-85).

Service's study was part of an educational experiment where three primary school classes began their English studies a year later than usual (referred to as late starters). The extra time was used for Finnish studies and the students caught up with others (the normal starters) in English by having correspondingly fewer Finnish classes and more English later on. Thus the subjects were Finnish school children, 22 of whom started to learn English in the first year of testing, while 22 started one year later. The subjects were nine years old, when the study began. They were tested four times at one-year intervals. Each time they repeated a list of 10 Finnish nonwords as well as a 10-item list

of English nonwords. Half of the nonwords in each language had two syllables and half had four syllables. The stimuli conformed to Finnish or English phonotactics. The Finnish nonwords were old Finnish words no longer in use and the English nonwords were constructed by interchanging certain syllables in a group of English words.

The subjects' performance was scored syllable by syllable so that a correctly repeated syllable had all the phonemes right. The Finnish nonwords were excluded from the study, since performance on them was nearly perfect. English nonwords, in contrast, did yield significant main effects of teaching group, syllable length, and year of testing. In other words, normal starters were better than late starters, shorter nonwords were easier to repeat than longer ones, and the subjects got better repetition results year by year.

Service's study compared the subjects' English nonword repetition scores to their English grades two and a half years after the first testing and found that they correlated significantly. From this Service concluded that nonword repetition scores predicted how well Finnish children learned English in the first few years of formal English education. The possibility that general academic ability might explain the correlations was ruled out by statistically examining the effect of the math grade on the established relationship.

Service also calculated correlations for repetition and the English subskills of listening comprehension, reading comprehension, and production. The measures were taken from two tests the subjects had taken during the years of testing. Answers to the comprehension tasks were given in Finnish and the production tasks required written translations, answering questions, and writing a composition. All the correlations between nonword repetition and English subskills turned out equally significant. In other words, nonword repetition was not related to any particular subskills, but to English skills in general. However, as Service points out, there may not exist such a clear

division between receptive and productive subskills for such young learners to begin with.

Service and Kohonen 1995. Service and Kohonen used partly the same data as Service (1989) in investigating the relationship between nonword repetition and L2 vocabulary learning. They focused on L2 vocabulary learning as the possible mediator between repetition scores and L2 knowledge, since they assumed that success in nonword repetition is really associated to vocabulary learning and only through that to other aspects of L2 knowledge.

Of the original 44 subjects in Service's (1989) study, 42 were available for further testing. The English nonword repetition tests were the ones reviewed above (Service 1989), but they were now compared to different measures of school achievement. These new measures were taken one year later than in the previous study by Service (1989), at the end of the spring term of the fourth year of the larger educational experiment. The means of the school report grades in nine theoretical subjects formed the measure of general academic achievement. Knowledge of English was measured in four ways: English grade, a communicative test, a traditional test, and a vocabulary test.

The communicative test measured the subjects' communicative skills in English with four types of tasks: reading comprehension, reproduction of structures (i.e. writing down what was dictated), listening comprehension, and written production (for details of the tasks, see Service and Kohonen 1995:161). The traditional test measured mostly the same skills, but in more traditional ways. The traditional tasks were listening comprehension, reading comprehension, knowledge of phrases, recognition of adjective forms, knowledge of words, correct word forms, and past tense forms (details in Service and Kohonen 1995:161-162). The vocabulary test required 40 written translations from English to Finnish and vice versa.

Service and Kohonen conducted stepwise multiple regression analyses on the data and the resulting partial correlations showed that there was a relationship between the nonword repetition tasks and the different measures of knowledge of English. This association was not mediated by general academic achievement. Another set of regression analyses was carried out to test the hypothesis that vocabulary was an intervening factor between repetition and other aspects of L2 knowledge. Indeed, when the effect of English vocabulary knowledge was eliminated, there no longer was a relationship between nonword repetition and the traditional and communicative tests of English. This led Service and Kohonen (1995:168-169) to conclude that "the relationship between phonological memory and language learning may depend on vocabulary knowledge".

Dufva and Voeten 1999. Dufva and Voeten conducted a longitudinal study aiming to discover whether L1 literacy and phonological memory predicted L2 learning. The subjects were 160 Finnish school children, who spoke Finnish as their L1 and had begun studying English as an L2 in the third grade. The study continued from the spring of the first grade to the end of the third; the subjects were seven years old, on average, when first tested.

The subjects' phonological memory was tested at the end of the second grade with a nonword repetition task modeled after Service (1989). The subjects heard two taped lists of ten nonwords. They sounded like English and were made from real English words. Half of the nonwords had two and half had four syllables. The subjects repeated the nonwords and the attempts were taped and scored syllable by syllable. Each correctly repeated syllable gave a point, so the maximum score for a list was 30.

Knowledge of English was measured with three tasks at the end of the third grade. Vocabulary knowledge was tested by having the

children produce short lists of English words and their Finnish translations on certain subjects such as colors. In a communicative task the children filled in gaps in an English text with the help of Finnish cues. Thus the task involved reading comprehension and written production. The third task was listening comprehension. The children were read a story twice and they then answered questions based on the story. The questions were in Finnish and they were also answered in Finnish and in writing.

The data were analyzed statistically (LISREL modeling) and it indicated that phonological memory – as well as some L1 literacy measures – predicted the children's performance in English after one year of learning. Dufva and Voeten (1999:345) suspected that there might exist a phonological skill that both word recognition (one of the literacy measures) and phonological memory stem from. Furthermore, they suggested that in previous studies the effect of phonological memory might have appeared to be too significant, since the effect of word recognition had not been taken into account.

Cheung 1996. Cheung studied the relationship between phonological memory and learning of L2 vocabulary in an experimental situation in order to find out whether phonological memory capacity could predict learning rate. The subjects were 84 native speakers of Cantonese Chinese who were defined as bilinguals, since they had been learning English since age three or four. At the time of the study they were in the seventh grade and on average 12 years old.

Different measures of the subjects' knowledge of English were taken, but the main focus was on the vocabulary learning task. The subjects were presented with three English words that were new to them and they had to learn the correct pronunciation and the Cantonese translation of the words. The score on this experiment was the number

of trials it took the subjects to learn the new items. What counted as learning were three consecutive errorless trials.

Phonological memory was measured with a nonword span task, a version of nonword repetition modified to suit older subjects. All English nonwords had two syllables and the subjects heard and repeated them in sequences starting with one and going up to seven nonwords. There were two trials at each sequence length. The task was continued until the subject made errors in both trials of the same sequence length.

Cheung carried out forced-entry and stepwise multiple regression analyses with the variables described above together with English and general measures. The results pointed to nonword span as the only measure that uniquely predicted vocabulary learning speed, i.e. the number of vocabulary learning trials. However, it appeared that there was some interaction between how much of an effect phonological memory has on vocabulary learning and former knowledge of vocabulary, since the results given above were only significant for students whose knowledge of English vocabulary was below the group median. In other words, phonological memory did not predict vocabulary learning for those with vocabulary knowledge above the group median.

2.4.2 Language-specificity of Phonological Working Memory

Thorn and Gathercole 1999. A few studies in the field of phonological working memory and L2 learning have concentrated on the possibility that phonological working memory might in fact be a language-specific phenomenon. Thorn and Gathercole conducted two experiments. In the first one they compared three groups of children: native English and French bilinguals, native English speakers learning French as an L2, and

English monolinguals. The native bilinguals had started learning French before the age of three, whereas the L2 learners had started learning French at school when they were three years of age or older. The L2 learners were selected so that their knowledge of French vocabulary was comparable to that of the native bilinguals, which inevitably made them somewhat older.

The aim of the study was to contrast the phonological memory scores of the different groups with their French and English vocabulary knowledge in order to determine the role lexical and sublexical knowledge play in phonological short-term memory capacity. If phonological memory is language-specific and thus reflects the vocabulary knowledge of a specific language, three things should happen: Native bilinguals should do better than native English monolinguals in the French phonological working memory tasks, but the performance of the L2 learners should be equal to that of bilinguals. In the English phonological memory tasks, the bilinguals and English monolinguals should do equally well, but the L2 learners should outperform both simply because they are older and thus hold a better command of English vocabulary.

The children's receptive vocabulary knowledge was measured with the short form of the British Picture Vocabulary Scale in which the child had to point to a picture that corresponded to the word spoken by the experimenter. The two measures used to assess phonological short-term memory were digit span and nonword repetition. Both were done in English and French. The nonword repetition task was a hybrid of the Children's Test of Nonword Repetition (see e.g. Gathercole, Willis, Baddeley, and Emslie 1994) and its French version. The subjects repeated two lists of 40 nonwords, half of which sounded like English, half like French. The length of the stimuli varied from two to five syllables and there were equally many nonwords of each syllable length.

Statistical analyses were carried out to investigate the relationship between the subject groups' performance on the vocabulary and phonological short–term memory tasks. The results on the memory tasks were as expected and reflected the children's vocabulary knowledge in the two languages. The L2 learners and the English monolingual children repeated English nonwords more successfully than French nonwords, whereas the native bilinguals scored equally well on both English and French nonwords. Repeating French nonwords was as easy for the L2 learners as for the native bilinguals. These results were interpreted as a strong indication that phonological short-term memory functions language-specifically.

In the second experiment, Thorn and Gathercole studied native and non-native English-French bilinguals (N=25). The non-native bilinguals had been learning the L2 since, on average, age two. Half of them had acquired English first, half French. In this experiment vocabulary knowledge was examined more broadly with both receptive and expressive tests. In the receptive test the experimenter spoke a word and the children had to choose the corresponding picture from those presented to them. The expressive test required the children to name objects in pictures. Both tests were done in English and French. The second experiment also yielded results supporting the claim that language-specific knowledge affects phonological short-term memory capacity: The native bilinguals repeated both French and English nonwords equally well, but the non-native bilinguals were better at repeating nonwords that sounded like their L1.

Masoura and Gathercole 1999. Masoura and Gathercole studied Greek school children who had been learning English for one to five years. They were interested in L2 learning in the formal context. They wanted to look at the relationship between L2 and L1 vocabulary knowledge and whether there was a connection between phonological short-term

memory and L2 vocabulary when the language was being learned in school.

The subjects were 45 Greek school children who were eight to eleven years old. Each child participated in a productive and receptive Greek vocabulary test and two translation tasks to measure their English vocabulary. Phonological memory was measured in both languages with the Children's Test of Nonword Repetition and its Greek version. There were ten more nonwords in the Greek test so that the stimuli extended from two to six syllables, whereas the longest English nonwords had five syllables.

Masoura and Gathercole calculated partial correlations between the test scores and concluded that the link between phonological short-term memory and vocabulary was significant. Both L1 and L2 vocabulary measures correlated with Greek and English nonword repetition scores alike. There was also a significant relationship between L1 and L2 vocabulary and it remained so after a composite repetition score, i.e. Greek and English nonwords combined, was partialled out. From this Masoura and Gathercole concluded that there was a connection between vocabulary knowledge of the two languages that was independent of the effect of phonological short–term memory.

Furthermore, nonword repetition remained significantly linked with L2 vocabulary after partialling out L1 vocabulary, but when L2 vocabulary was, in turn, partialled out, L1 vocabulary was no longer associated with nonword repetition. This was interpreted by Masoura and Gathercole as a sign of possible language specificity. All these connections were independent of general factors such as age, nonverbal ability and length of English studies.

It is worth mentioning that Masoura and Gathercole's view of language-specificity is a bit different from that in Thorn and Gathercole (1999) above. Vocabulary knowledge in both L1 and L2 correlated significantly with repetition scores in both languages. However, the

situation changed when a composite repetition score was made of both L1 and L2 nonwords. While general factors were also controlled, the composite repetition score, L1 vocabulary, and L2 vocabulary were compared to each other in pairs so that the effect of the third variable was statistically excluded. In this situation, only L2 vocabulary remained significantly associated with repetition capacity. This together with the result that L2 nonwords were repeated less successfully was an indication of language-specificity to Masoura and Gathercole. Thus, here language-specificity does not mean that nonword repetition in a particular language is connected to knowledge of that same language, but that a more general nonword repetition capacity is connected to L2 but not L1 knowledge.

2.4.3 Adult Learners of L2

Baddeley 1993. Several of the studies on the relationship between nonword repetition and L2 learning focus on adults as language learners. Baddeley conducted a case study of a 23-year-old male (S.R.) with a suspected phonological short-term memory deficit. The aim of the study was to clarify whether or not S.R. actually had such a deficit, how it affected his long-term learning, and what the practical consequences were. The subject and six peer controls took part in eight tasks of verbal short-term memory, a task of visual short-term memory, two tasks of visual and verbal long-term memory, L1 and speech production, and an L2 vocabulary learning task. One of the verbal short-term memory tests was a phonological working memory measure, an adult version of the Children's Test of Nonword Repetition (Gathercole and Baddeley 1989). In this task the participants repeated nonwords of increasing length, starting with one syllable, in trials of

three, until they failed to repeat correctly any of the three nonwords of that particular length.

From Baddeley's measurements it was apparent that S.R. indeed had a much poorer verbal memory span than his peers, but that his deficit surfaced with verbal material only. This defective phonological short-term memory was mirrored in poor long-term phonological memory and learning as measured by verbal recognition and name recall. Whereas S.R's limited short-term memory had not affected L1 development, he had never been able to learn an L2.

To further examine the implications of S.R's deficit on L2 acquisition, Baddeley conducted a learning experiment. Through several learning and test trials S.R. and the controls tried to learn eight pairs of English words and eight English-Finnish pairs. With the help of semantic coding, S.R. did well on the L1 items, but he performed poorly on the English-Finnish pairs. The phonologically new items would have required rote rehearsal, but S.R. was apparently incapable of doing that.

The results of the case study indicated that impaired phonological short-term memory was connected to defective long-term phonological learning. From the point of view of the working memory model in question, this means that a proper functioning of the phonological loop is mainly important for its effect on long-term phonological learning. In practice, poor phonological working memory hinders L2 learning as it did in S.R's case.

Papagno and Vallar 1995. Also Papagno and Vallar studied adults, but they were interested to see if polyglots differed from non-polyglots in terms of phonological short-term memory skills and new vocabulary learning. The polyglots spoke Italian as their L1, but they were fluent in at least three languages. The non-polyglots had studied only one L2 in school. Otherwise the groups were comparable and no differences in other skills and qualities were expected.

The subjects were tested in several ways. They took part in tests of general intelligence, L1 vocabulary knowledge, visuo-spatial span, and visuo-spatial learning. Of most interest here are the immediate verbal memory tasks and paired-associate learning of real and nonwords.

Verbal memory (i.e. phonological working memory) was measured with auditory digit span and a nonword repetition task. Again, a more complicated version of the nonword repetition test was used evidently because the subjects were adults, even though Papagno and Vallar did not give any details of how the test was constructed or reasons for using this particular test. It remained unclear whether or not the nonwords conformed to the phonotactics of any particular language, if they sounded like Italian, for example, but they were described as "meaningless" and "pronounceable" (Papagno and Vallar 1995:100). The items were between two and nine syllables long and they were presented and repeated one by one starting with the shortest ones. There were three nonwords of each syllable length and the task continued until the subject made three consecutive errors. It is not mentioned what counted as an error, but the score was the number of syllables in the longest nonword that was repeated correctly.

The subjects' ability to learn new words was measured with an experimental task where the subjects tried to learn eight pairs of Italian (L1) words and eight word-nonword pairs. The nonwords were pronounceable Russian words. Half of the Italian items in both word-word pairs and in word-nonword pairs had two syllables and the other half had three. Also the Russian nonwords had two or three syllables. After hearing the pairs once the subjects were presented with the first item of a pair and required to say the second half. This was continued until the subject got all eight items right in two consecutive trials. The maximum number of trials was set at ten. The score was the number of

items that the subjects remembered correctly, so it was between zero and eight for each trial.

The data were analyzed, for example, by using unpaired t-tests and principal component analysis. The results showed that the polyglots performed significantly better than the non-polyglots on nonword repetition and auditory digit span. They also learned nonwords faster than the non-polyglots, but there was no difference between the groups on learning pairs of Italian words. These results appeared to be independent of general intelligence and L1 vocabulary knowledge, and Papagno and Vallar considered them to be evidence of a close association between phonological memory capacity and L2 learning.

2.4.4 Other Measures of Phonological Working Memory

As mentioned at the beginning of section 2.4, there are also studies that have yielded interesting findings about the relationship between phonological working memory and L2 learning without using any type of nonword repetition task. Two such studies are briefly summarized here.

Papagno, Valentine, and Baddeley 1991. Papagno, Valentine, and Baddeley studied the role of short-term phonological store in adult L2 vocabulary learning. The method used to get to the phonological loop was articulatory suppression, which in this case meant repeating the sound "bla" while doing vocabulary learning tasks. Articulatory suppression was assumed to disrupt the operation of the phonological loop by preventing subjects from rehearsing words subvocally or by interfering with the phonological recoding when words were presented visually.

In a number of experiments, Italian and English subjects had to learn pairs of words, half of which consisted of two L1 words and the other half were L1 - L2 pairs. The pairs were presented to the subjects either visually or auditorily in lists of eight pairs. The length of the words varied between languages. The response was written and there was a maximum of five test trials following the practice and learning trials. The score for each test trial was the total of pairs that were remembered correctly.

While first learning the pairs, the subjects did a secondary task, either an articulatory suppression task of repeating a single syllable all through the list or tapping with a finger. The results showed that articulatory suppression was more harmful to learning L2 than L1 vocabulary, which indicates that short-term phonological coding is indeed important for the acquisition of L2 vocabulary.

Atkins and Baddeley 1998. Atkins and Baddeley studied the relationship between verbal memory span and L2 vocabulary learning. Their assumption was that the longer the memory span the more successful the learners would be at learning L2 vocabulary. The subjects were 30 adults between the ages of 19 and 40.

Both visual and verbal short-term memory were tested. The only measure of the visual aspect was a test of visuo-spatial recall, in which the subjects had to reproduce patterns shown to them. Verbal span was measured with six different tasks: digit span, phonologically similar letter span, and phonologically dissimilar letter span all with both auditory and visual stimuli.

In the vocabulary learning task there were four lists of 14 pairs of words, an English word and its Finnish translation. The words were learned either as pairs of single words or as pairs of two-word sentences. Half of the pairs were learned during the first session and the other half one week later. Learning and testing were subject paced so

that the subjects were working alone with a computer and they could focus on those items difficult for them. What was analyzed from the learning task was vocabulary learning speed. It was measured with the average number of errors the subjects made, while learning the pairs or sentences. Scoring was done in two ways. The strict score did not allow any errors or omissions, while the lenient version did not count spelling mistakes as errors.

A principal components analysis was conducted on the short-term memory tasks and two factors were found, a verbal and a visual factor. From these, only the verbal factor correlated with the speed of learning vocabulary items. It did not matter whether the words were learned in pairs or sentences, whether the presentation was auditory or visual, or whether the scoring was lenient or strict; all learning speed scores correlated with verbal span. In contrasts, visuo-spatial span did not correlate with any aspect of vocabulary learning speed. It was concluded that verbal span significantly predicted the speed of learning new vocabulary whereas visuo-spatial span did not.

Neither visual nor verbal span correlated with how many errors the subjects made one week after learning the words. Atkins and Baddeley interpreted this so that poor short-term memory did not cause faster forgetting of vocabulary items learned.

2.4.5 Comparison and Summary

The studies on the connection between nonword repetition and L2 learning reviewed in the previous sections (2.4.1–2.4.4) are a very heterogeneous group, so comparing the results is not easy or even possible and it is not attempted here. Only a brief summary of the major findings and (somewhat intricate) comparisons of some other aspects of the studies such as the subjects' age, languages involved, and the

measures of phonological working memory and L2 learning are presented in the following.

The Results and Subjects. A clear majority of the studies reviewed above focused on learning L2 vocabulary. Only Service (1989), Service and Kohonen (1995), and Dufva and Voeten (1999) attempted to make more general conclusions of the effect of phonological working memory on L2 learning as a whole. A few studies explicitly mentioned that nonword repetition could be used to predict success in L2 learning (Service 1989, Dufva and Voeten 1999, and Atkins and Baddeley 1998), while others merely talked about a clear or significant association between the two. There were also a few studies in which phonological working memory was found to be language-specific (Thorn and Gathercole 1999 and Masoura and Gathercole 1999). What is common to all these studies is that they seem to support the assumption that there is a significant association between phonological working memory and L2 learning, whether it be predictive, language-specific or something indefinable.

Of the studies reviewed above, six focused on children (Service 1989, Service and Kohonen 1995, Dufva and Voeten 1999, Cheung 1996, Thorn and Gathercole 1999, and Masoura and Gathercole 1999) and four on adults (Baddeley 1993, Papagno and Vallar 1995, Papagno, Valentine, and Baddeley 1991, and Atkins and Baddeley 1998). Interestingly, the studies with adults involved subjects who were somehow exceptional and as such the studies were not comparable to the mainstream of phonological working memory studies. One was a case study of a student with a memory deficiency, one studied polyglots, and two of the studies of adults as L2 learners did not use any kind of nonword repetition task. What comes to gender, there was roughly an equal number of males and females as subjects. Papagno

and Vallar (1995) studied females only and Masoura and Gathercole (1999) did not mention the gender distribution of their 45 subjects.

Considering the small number of studies, the diversity of the L1s of the subjects is great. Among the ten studies, there were L1 speakers of five languages (English, Finnish, Italian, Greek and Chinese) and English-French bilinguals. In some studies the subjects' L1 was not mentioned (Baddeley 1993 and Atkins and Baddeley 1998), but in those cases it was presumably English.

The subjects were also learners of a great number of languages; also the length they had studied the languages varied. Most subjects studied English (Service 1989, Service and Kohonen 1995, Dufva and Voeten 1999, Cheung 1996, and Masoura and Gathercole 1999) but there were also learners of French (Thorn and Gathercole 1999). In the case of the polyglots (Papagno and Vallar 1995), the three or more languages they spoke fluently were not mentioned, nor were the circumstances in which they had learned those languages. There were also other cases (e.g. Atkins and Baddeley 1998) where there was no specific mention of the subjects' possible language learning experience, or it was described but unnecessary for the study, since the measure of L2 learning to which working memory capacity was compared was not a result of any orthodox language learning process. In several cases L2 learning was measured with a learning experiment, such as the speed of learning paired-associates of familiar and new words (Papagno and Vallar 1995), for example. In such cases the subjects' knowledge of L2 was not essential as long as they were not familiar with the language they were required to learn in the experiment.

In sum, the subjects of the studies differed in several respects. An important addition to the differences listed above is that also the number of subjects varied greatly between studies: in Baddeley's (1993) case study there was only one subject with a memory deficiency and six normal controls, and Papagno and Vallar (1995) studied 10 female

polyglots. On the other extreme, Dufva and Voeten's (1999) subjects comprised of 160 children.

The Measures of Phonological Working Memory. Besides the subjects, the studies also varied greatly in the types of tasks the subjects took part in, what was tested and how. The focus is narrowed to comparing the measures of phonological working memory and L2 learning.

All but two of the studies used some form of the nonword repetition task. The two studies that did not, involved adults. In Atkins and Baddeley (1998), verbal short-term memory was tested by having the subjects repeat lengthening sequences of digits, phonologically similar letters, and phonologically dissimilar letters. These three span tests were done with both auditorily and visually presented stimuli. In Papagno, Valentine, and Baddeley (1991) the instrument used to study phonological short-term memory was articulatory suppression. It eliminated the use of the phonological loop while learning new vocabulary and thus enabled the observing of the effect of phonological sort-term memory on L2 learning.

The other eight studies reviewed here used some form of nonword repetition to measure the subjects' phonological working memory. The basic and most often used type of nonword repetition task was the one where the subjects heard nonwords and repeated them aloud immediately one by one. There were an equal number of nonwords of each syllable length. For example, in the Children's Test of Nonword Repetition used by Masoura and Gathercole (1999) there were ten nonwords of each length between two and five syllables (nonwords sounding like English) or two to six syllables (sounding like Greek). Similarly, Thorn and Gathercole (1999) had their subjects repeat 40 English and 40 French nonwords that were between two and five syllables in length. The nonword task in Service (1989) and Service and Kohonen (1995) was otherwise similar, but there were only ten words

per task. Half of the words had two syllables, half had four. There was a test with Finnish-sounding nonwords as well as in English, but the Finnish one proved too simple for the subjects. Possible reasons for this are discussed below. Dufva and Voeten (1999) modeled their nonword task after Service (1989), but only tested with English nonwords.

Cheung (1996), Baddeley (1993), and Papagno and Vallar (1995) used slightly more complicated versions of the repetition task. In fact, they all tested nonword span. Cheung's (1996) subjects repeated nonwords in sequences starting with only one nonword and proceeding until a maximum length of seven nonwords per sequence. There were two trials at each sequence length and all the nonwords throughout the task had two syllables. The task was stopped if the subject failed to repeat both of the sequences of a certain length correctly.

Baddeley (1993:132) called his nonword span task "an adult version of the task developed by Gathercole and Baddeley (1989)". The nonwords were repeated one by one, unlike in the task described above. There were three nonwords of each syllable length starting with one and lengthening until the subject got all three wrong. The span task used by Papagno and Vallar (1995) was very similar, only for them three consecutive errors were enough to stop the task, even if all errors did not occur within the same syllable length. Furthermore, Papagno and Vallar (1995) started from two syllables up.

The subjects in these three nonword span studies were older than in the five studies with more basic nonword repetition testing. Baddeley (1993) and Papagno and Vallar (1995) studied adults and the mean age of Cheung's (1996) subjects was 12.2 years. Age was probably the reason a span task was chosen to test phonological working memory in these cases. Other than that, the researchers seldom gave reasons for choosing the type of nonword task they did, much less

pondered about all the different possibilities for testing phonological working memory or constructing a nonword task.

In some of the studies the nonwords were presented live, but in most cases the stimuli were taped. It is easy to imagine that there are differences between the two methods, and they need to be considered when evaluating the results or constructing a test of one's own (see a discussion on this in Gathercole and Baddeley 1997 and Bowey 1997, for example). In most cases the repetition attempts were taped, too. That way the scoring could be done or revised later. Not all researchers made it clear whether the stimuli or repetition attempts were taped or live.

Methods for scoring the repetition attempts were also various and not clear in all studies. Some used second raters and calculated the inter-rater reliability (e.g. Service 1989), most did not. Generally, a native speaker of the language of the task listened to the repetition attempts and gave points for desired performance. Some gave points for all syllables produced correctly, most gave a point per correct nonword. In the nonword span cases, the score was the number of syllables in the longest nonword, although Baddeley (1993) reported three different scores for his nonword span task. In Cheung's (1996) study the score was the length of the longest sequence (i.e. number of nonwords) where the subject got both trials right. Also, what counted as an error varied somewhat. Usually all phoneme additions, omissions, replacements, or ordering the phonemes differently were counted as an error. Not all, however, made it explicit what counted as an error (e.g. Papagno and Vallar 1995).

Yet another point on which the studies can be divided in several categories is the language the nonwords imitated. The two studies that tested adults with nonword repetition (Baddeley 1993 and Papagno and Vallar 1995) did not make it clear which language the test was supposed to mimic. Both studies only stated that the nonwords were

such that the subjects could pronounce them without too much effort, but it does not have to mean that they sounded like the subjects' L1, which in Baddeley's (1993) case was not stated explicitly, anyway. It was not discussed which features of the nonwords made pronouncing them easy.

Furthermore, in these two cases it was not clear which other languages the subjects could speak or had studied. Apparently it was considered irrelevant, since the nonword tasks were not compared to any actual L2 knowledge but to an L2 learning experiment. In these cases the significance of the relationship found between the tasks and language learning remains questionable, since it could not really be said what was being compared to what, especially in the light of the studies pointing to language-specificity of the phonological memory – language learning relationship (Thorn and Gathercole 1999 and Masoura and Gathercole 1999).

Four of the studies compared success in repeating L2 nonwords to success in learning that language (Service 1989, Service and Kohonen 1995, Dufva and Voeten 1999, and Cheung 1996). In all four cases the language under scrutiny was English. Even these studies differed in how they measured the subjects' L2 knowledge, but that is expanded on below.

Thorn and Gathercole (1999) as well as Masoura and Gathercole (1999) tested both L1 and L2 nonword repetition. They were specifically interested in studying how phonological memory in a particular language is connected to the knowledge of that same language as compared to other languages. In fact, also Service (1989) – and hence also Service and Kohonen (1995) – conducted the nonword repetition task in the subjects' L1 as well as in English. However, the Finnish nonword tasks were excluded from the study because the subjects' performance on them was nearly perfect. Service's (1989) Finnish nonword stimuli deserves a closer look, since her subjects produced

such perfect repetitions in the L1 but the subjects in Thorn and Gathercole (1999) and Masoura and Gathercole (1999) did not.

There are a few aspects of Service's (1989) study that are problematic and most certainly have had an effect on the results. Service (1989:179) provided the lists of nonwords used in her study in Appendix 2. A quick glance at them raises questions about the criteria according to which Service had chosen these particular old Finnish words. Many of the nonwords contained various suffixes which had to make the nonwords much easier to repeat than if they had been completely new to the subjects (e.g. *vieroaja, mujamainen, jamauminen*). Some of the nonwords were clearly verbs with the appropriate endings and that must have aided in repeating them (*sähtyröidä, jongerrella, kämmertyä, kaidehtia, laipiottaa*). If parts of the nonwords are familiar to the subjects, it must have an effect on what repeating the nonwords is actually measuring: phonological memory or something else. The whole idea behind repeating nonwords is to eliminate the influence of the subjects' previous knowledge of vocabulary and access the phonological working memory skills. Including familiar elements in the stimuli works against this.

It also remains unclear why Service (1989) decided to use only two- and four-syllable nonwords. Since the subjects were already nine years old during the first testing and the nonword lists remained similar for the next three years, it seems likely that the nonwords were much too short (that is, easy) to yield any differences between the subjects. For example, Gathercole and Baddeley (1989) tested English-speaking children with English nonwords that were up to four syllables long, while the subjects were only 4 and 5 years old.

The Measures of L2 Knowledge. It is surprising how many different views of L2 learning can be found in the eight studies that needed a language measure to compare the nonword task scores to. It remains unclear how

much thought the researchers have put to their conception of language learning, but judging by the tasks they have chosen to measure knowledge of an L2 with, most have not thought about it much.

As pointed out in section 2.3 (Nonword Repetition and L1 Learning), most research on phonological working memory and language learning has focused on the relationship between nonword repetition scores and knowledge of L1 vocabulary. This is also the case with L2 learning. It may be unfair to criticize studies that purposefully focus on vocabulary learning for their narrow point of view, but it seems necessary to mention something even in brief.

From a linguist's point of view, it seems that most researchers have cut a few corners in their interpretation of L2 knowledge and the role of vocabulary. No doubt, knowing the vocabulary is essential in using an L2, or any language, and learners' vocabulary knowledge is very much reflected in other areas of language knowledge, but using vocabulary as the sole measure of knowledge of an L2 seems unnecessarily limiting. Researchers appear to be very eager to accept vocabulary as the one aspect of knowledge of an L2 worth studying in the context of nonword repetition, but they rarely give convincing evidence to support their choice.

Whether measuring vocabulary or any other aspect of language knowledge, it was not made explicit in most of the studies what the reason for using the particular tasks was, i.e. if the purpose was to measure L2 achievement or ability. For the most part, it seems that the researchers were not aware themselves which of the two they were studying. In some cases this could be deduced from the tasks used. For example, Masoura and Gathercole (1999) used the British Picture Vocabulary Scale, which is an achievement test, whereas Cheung (1996) taught his subjects new L2 vocabulary, which clearly tests the subjects' more general L2 learning ability instead of any particular knowledge or achievement in the language.

There are, however, many aspects in the language learning measures adopted in the studies that tell something of the researchers' view of L2 learning. First of all, the studies can be divided in two on the basis of the language the learning of which was studied. Again, the two studies done with adults (Baddeley 1993 and Papagno and Vallar 1995) differ from the others. In them the subjects were required to learn new vocabulary to a language they had never studied before. In all of the other studies the subjects' knowledge was measured in a language that they had been studying for various lengths of time. In addition, Thorn and Gathercole (1999) and Masoura and Gathercole (1999) also measured knowledge of the L1 for comparison.

The ways to measure the actual or theoretical ability to learn an L2 were various. Starting with the three studies not focusing merely on vocabulary learning, Service (1989) held a rather simplistic conception of L2 learning. What to Service constituted knowledge of English was the school grade, which apparently covered the children's efforts in the English class for one particular semester (the spring two and a half years after the first testing) as evaluated by the English teacher. Service (1989:145) mentioned that language learning was assessed with "normal language tests that had been carefully planned for the purposes of the educational research project the subjects were participating in", but as far as nonword repetition accuracy goes, Service's (1989) main argument seemed to be that repetition scores from four years correlated with one single English grade. She did not attempt to clarify what the grade really consisted of (clearly defined subskills in English, grades in exams, attendance, active participation…?). However, the grade did correlate with the total and subtests (written production and listening and reading comprehension) of two English tests taken by the late-start students and based on this Service (1989:58) suspected that the teachers gave the grades strictly according to the subjects' success in those tests.

In Service and Kohonen (1995), English proficiency was measured in a much more diverse way. Beside the grade, there were communicative and traditional tests aimed to measure a variety of different things about language knowledge. There was also a vocabulary test in which the subjects had to translate 40 words from English into Finnish and vice versa.

Dufva and Voeten (1999) assessed their subjects' knowledge of English with three tasks: an active vocabulary measure, a communicative task and a listening comprehension task. All of the tasks together took an hour, so they were not very broad, but this seems justified since the subjects were young (third-graders) and had only been learning English for one school year.

The rest of the studies focused on vocabulary measurements, but they too can be divided into two groups. Thorn and Gathercole (1999) and Masoura and Gathercole (1999) studied their subjects' knowledge of L2 (and L1) vocabulary with a few simple tasks. Either the subjects had to point to pictures corresponding to L2 words they heard or produce the right words for pictures shown to them (Thorn and Gathercole 1999). In a translation task the subjects heard an L1 word and answered with the corresponding word in the L2 and vice versa (Masoura and Gathercole 1999). It needs to be emphasized that these were words that the subjects could have known because they knew the language, at least to a certain extent. Also there could be no certainty of how the subjects came to know the words in the first place. They either knew them at the moment of testing or they did not.

What Papagno and Vallar (1995), Baddeley (1993), and Cheung (1996) did was to teach their subjects completely new words and observe how well and how fast they learned them. This seems like a very theoretical measure of vocabulary knowledge. It was not always even clear or evidently relevant which L2 the subjects knew, if any. The learning experiments gave a score that told the researchers how well

the subjects could learn new L2 vocabulary, but they were words chosen by the researchers in a language chosen by the researchers learned in a method chosen by the researchers, and thus seemed to be very little connected to any real life experience or knowledge of the subjects. Cheung's (1996) study is a borderline case in the sense that the subjects were students of English and they had to learn three English words in the experiment, but since the words were, naturally, new to the subjects, they might as well have been from any language.

3 EMPIRICAL STUDY

3.1 Research Design, Questions, and Hypotheses

As should be evident from section 2.4.5 above, there is a great number of differences between the studies and there is virtually no way of comparing them. Yet, in all of the studies it is found that there is a connection between phonological memory and L2 learning. This can either raise suspicion: How can they get the same result when they are barely testing the same thing? Or it can be extremely convincing: No matter how you look at it, there is an undeniable connection between phonological working memory and L2 learning. In any case, the confusion calls for a study that takes into consideration the factors found problematic in the previous studies and the present study is an attempt to do that, at least to some extent.

Phonological working memory is claimed to be essential for the long-term learning of language, L1 as well as L2 (Gathercole and Baddeley 1993a:70–73). In other words, if there are problems with the phonological working memory, there are problems in learning languages. Studying this memory system may help in easing possible future problems or already existing language difficulties. One way to operationalize phonological working memory is by nonword repetition (e.g. Gathercole and Baddeley 1989). Language learners' success in repeating unknown words sounding like real language depends on the proficiency of their phonological working memory, and not their knowledge of that particular language. Nonword repetition tests have been used quite extensively to study the relationship between phonological working memory and language learning, but so far results on Finnish subjects repeating Finnish-sounding nonwords have not

been reported. It has been studied, but the stimuli have been too easy for the subjects (Service 1989).

The main interests in the present study were to construct and conduct Finnish nonword repetition tests that yield interpretable results as well as to measure the participants' knowledge of English and compare it to both Finnish and English nonword repetition. There were sufficient grounds for believing that with proper stimuli it would be possible to get results with Finnish nonwords from Finnish children. Another similar test needed to be carried out later to study possible (or inevitable, according to Gathercole, Willis, Emslie, and Baddeley 1991, for example) development in phonological memory. The longitudinal design of the research project – and the present study along with it – made it possible to carry out such comparisons between the participants' performance in the second and fifth grade of comprehensive school.

Since only a few studies exist on the issue of language-specificity of phonological working memory (namely Thorn and Gathercole 1999 and Masoura and Gathercole 1999), it was also of interest and required tests of its own: an English nonword repetition test and a measure of knowledge of English. Finally, the tests were also used to see if the results would support the usual finding that the level of L2 knowledge correlates with the nonword repetition score in that language, in this case English.

The general research questions and the specific research hypotheses are listed below.

1. Is children's knowledge of an L2 connected to their success in repeating nonwords sounding like that L2?

   Hypothesis: There is a positive correlation between the 5th grade English nonword repetition score and the 5th grade English score.

2. Is phonological working memory language-specific?

   Hypotheses:

   - There is no correlation between the 2nd grade Finnish nonword repetition score and the 5th grade English score.
   - There is no correlation between the 5th grade Finnish nonword repetition score and the 5th grade English score.
   - There is no correlation between the 2nd grade Finnish nonword repetition score and the 5th grade English nonword repetition score.
   - There is no correlation between the 5th grade Finnish nonword repetition score and the 5th grade English nonword repetition score.

3. Are the same children good at repeating Finnish nonwords in the fifth grade as in the second grade?

   Hypothesis: There is a positive correlation between the 2nd grade Finnish nonword repetition score and the 5th grade Finnish nonword repetition score.

Both the questions and the hypotheses are based on or inspired by the previous research on the subject reviewed in section 2.4. As to the first research question, at least Service (1989), Dufva and Voeten (1999), Service and Kohonen (1995), and Cheung (1996) have studied the connection between phonological working memory and L2 learning using nonword repetition tests. The repetition scores were compared with the subjects' L2 knowledge and it was found that success in one goes hand in hand with success in the other.

The second research question concerns the language-specificity of phonological working memory. Thorn and Gathercole (1999) found some evidence to suggest that with the same subjects, success in nonword repetition tests in one language depends on knowledge of that language, and if the nonword stimuli are based on other languages,

it is the knowledge in those languages that the repetition results depend on. That is why two of the four research hypotheses do not expect there to be a correlation between knowledge of English and repeating Finnish nonwords. If there really is a separate phonological working memory for each language – or the phonological working memory at least performs differently for each language – the nonword repetition tests with different language stimuli should not correlate. The participants being very young and not far in their English studies, their level of knowledge of English is not at all comparable with their knowledge of Finnish and therefore their English nonword repetition results should not be similar to their Finnish nonword repetition results. The two latter research hypotheses are based on this line of reasoning. Naturally, these hypotheses are based on the assumption that the hypothesis of the first research questions is supported: If there is not a positive correlation between English nonword repetition and knowledge of English, then there is no sense in claiming that phonological working memory is language-specific just because Finnish nonword repetition does not correlate with knowledge of English either, or with English nonword repetition, for that matter.

The third research question and the hypothesis stem from an often repeated finding that phonological working memory develops as children get older (e.g. Gathercole and Baddeley 1993a:25–31 and Gathercole, Willis, Emslie, and Baddeley 1991:365). Since the two nonword repetition tests used in the present study were rather different, it is difficult to say anything about the development in the sense of how much better the children have become at repeating nonwords. However, if both Finnish nonword repetition tests test the same thing and the scores on the two tests correlate, the development of all the participants has been similar: They all have developed as much or as fast in the time between the tests despite the level they started from.

3.2 Participants

The participants were 16 L1 Finnish speakers attending school in central Finland. At present they are in the fifth grade and 11 or 12 years old. Half of them are boys and half girls. All of the participants chose English as their first L2 and began to learn it in the third grade. They were eight or nine years old at the time. The same English teacher teaches all of the participants but they are divided into three groups with other children.

3.3 Timing of Data Collection

Two nonword repetition tests and a vocabulary task were carried out to find answers to the research questions. The results of an earlier Finnish nonword repetition test were also used, so altogether there were three nonword repetition tests: two Finnish ones and one with English-sounding stimuli. There were two and a half years between the earlier test and the rest of the measures. Finnish nonword repetition test 1 was carried out in the second grade whereas Finnish nonword repetition test 2, English nonword repetition test, and the measures of knowledge of English were taken in the fifth grade. The English measures included a vocabulary test constructed particularly for the present study, but also measures originally meant for other purposes were used to assess the participants' knowledge of English. These included parts of an English exam given by the English teacher and a preposition task by another member of the research team.

The timing of the tasks was not haphazard. To test the idea of language-specificity, the Finnish nonword repetition results were compared to English nonword repetition and measures of knowledge of English. For these purposes it was important that the data from the

second Finnish nonword repetition task, the English nonword task and the measures of English were all from approximately the same time, in this case the end of the fall semester in the fifth grade. The stages of data collection are listed in Table 1.

Table 1. Stages of data collection.

Time	Test or task
May 2000	Finnish nonword repetition test 1
October 2002	Finnish nonword repetition test 2
October 2002	English nonword repetition test
November 2002	Preposition task (Sentence formation 2)
December 2002	English exam
December 2002	Vocabulary task

3.4 Finnish Nonword Repetition Test 1

The first Finnish nonword test took place at the end of the second grade, in May 2000. The participants were eight or nine years old at the time and all 20 were available for testing. They had not begun to study an L2 yet. It is important to note that the author of the present paper was not involved in the construction of Finnish nonword repetition test 1 or its use in collecting data. This was done by other members of the research group before the involvement of the present author in the research project. The data were, however scored again by the present author together with the other two nonword tests to ensure the same criteria and equipment were used and the scores were comparable in that respect.

Design of the Test. The nonwords used in the task were not developed especially for this task, but they were old Finnish words that are no

longer in use, selected from an old Finnish dictionary, *Suomen kielen keräilysanasto 1* (Ekman (ed.) 1899). The nonwords were chosen from among two-syllable words. Longer nonwords were made by combining the old words. On the basis of previous research (i.e. that of Service 1989), who got no results with five two-syllable and five four-syllable Finnish nonwords) a test list of 15 nonwords was constructed: five with two syllables, five with four syllables, and five with six syllables. The order of the nonwords in the test was random. There was also a practice list of six nonwords, two of each syllable length. The nonwords are listed in Appendix 1 in the actual order of testing. The entire test was taped with eight second pauses between the stimuli to allow for repetition. The nonwords were read on tape by a female native speaker of Finnish who lived in the same area as the participants. Included in the beginning of the tape were spoken instructions in a male voice.

Procedure. The participants were tested individually in an empty classroom. The experimenter gave general instructions before the test began. The participants were told that they would hear funny words that sounded like Finnish, but they probably had never heard them before. They should try to repeat the words and it did not matter if they were not quite sure what they had heard.

The participants wore headphones through which the taped instructions and the stimuli could be heard. The participants spoke into a microphone and their repetition attempts were taped for later scoring. The attempts were scored by the present author so that one point was given for a correctly repeated nonword and no points were given, if at least one phoneme was wrong, misplaced or omitted. The maximum score was 15.

The criterion for correct / incorrect phonemes was their ability to differentiate meaning in Finnish in general. For example, in Finnish there is only one *s*-sound as far as differentiating meaning is concerned,

so if the participants produced a sound different from that in the nonword stimulus, but still an *s*-sound, it was counted as the right phoneme. Consistent deviations from the norm, such as an inability to pronounce the letter *r,* were not scored as errors. If it was impossible to say whether the participants had produced the correct phoneme or not, the attempt was systematically scored as correct. Self-correction was allowed.

3.5 Finnish Nonword Repetition Test 2

The second Finnish nonword repetition test took place in the fifth grade, in October 2002, about two and a half years from the first test. The participants were ten or eleven years old and there were only 16 of them left.

Design of the Test. Since the age of the participants is a factor in phonological working memory development, the two Finnish nonword tasks had to be different in stimuli as well as in structure. The structure of Finnish nonword repetition test 2 was modeled after Baddeley (1993). Since the participants were relatively old to be repeating nonwords sounding like their L1, it was apparent that the basic version of nonword repetition would be too easy for them. In order for the test to be sensitive enough for the older population, longer nonwords were needed. Baddeley's (1993) version starts with short nonwords and proceeds to the limits of the participants' capacity by adding one syllable at a time. There are only three nonwords of each syllable length, so the test does not get too tiresome even when the participants proceed very far.

The maximum length of nonwords was set at eight, which was two syllables more than in the second grade. The minimum was set at

two syllables, the same as in the second grade. Since there were three nonwords of each syllable length, a total of 21 nonwords had to be created for the test. The nonwords were constructed based on existing guidelines for possible structures of different syllable length words (Lapsen kielen kehitys ja dysleksia –projekti / The Jyväskylä Longitudinal Study of Dyslexia (JLD)). A number of possible structures were chosen from the extensive lists in the guidelines so that the nonwords would be natural-sounding, phonotactically Finnish (e.g. which vowels can go together, which consonants can be used, and how consonants can be clustered) and still have a variety of different consonant quantities, diphthongs, and clusters.

The nonwords had to seem as if they were nominative nouns, so certain word endings were avoided on purpose. The nonwords could not have any suffixes to make sure that they were all in that respect equally difficult and no syllables were offered free for the participants. A number of two-, three-, and four-syllable nonwords were constructed and they were combined in different ways to make up the longer (5–8 syllables) nonwords. There was an attempt to do most of the selection stages objectively at random. However, changes had to be made a few times based on the author's own judgment because the nonwords did not come out wordlike enough. Composing a very natural-sounding eight-syllable nonword was rather complicated. The nonwords are listed in Appendix 2. This time there were no practice items following the example of Baddeley (1993).

The Finnish nonwords were read on a minidisk by a female native speaker of Finnish, the same one as in the previous test. To get the best possible result, the nonwords were read several times in a list and in sentences to avoid a list intonation. The most suitable version of each nonword was chosen for the final taped version of the test, and six-second intervals were left between nonwords to allow time for repetition.

Again, in the beginning of the test tape there were instructions for the participants (in addition to the ones given orally before beginning the actual test). The instructions were only slightly adapted from the first Finnish nonword repetition test. They, too, were read on a minidisk by the same male native Finnish speaker as in the first test. A pilot test was carried out with a female adult speaker of Finnish to make sure the instructions were clear enough, there was enough time for repetition and to test the equipment, as well as to see how much time it took to carry out the whole task with the live instructions.

Procedure. The test was slightly modified from Baddeley (1993) so that it was not disrupted once the participant missed all three nonwords of a certain syllable-length, but all participants were given a chance to repeat all of the nonwords. This was done for three reasons. First, the testing conditions were such that it would have been impossible to score the repetitions attempts reliably enough on the spot to confidently stop the task in the middle. Second, it allowed two different ways of scoring: The strict Baddeley (1993) version, where the score is the number of syllables in the longest correctly repeated nonword before missing all three nonwords of a certain length, and a more traditional scoring of one point per correct repetition, where the possible maximum score for all participants is the same. Third, the English nonword repetition test had to be done the same way as the second Finnish one, and since it was expected that the shortest English nonwords might be difficult for the participants to catch and thus to repeat, many might end up scoring zero points in the Baddeley system.

The testing was done the same way as in the second grade. About a half of the participants were tested first and the other half one week later. The scoring criteria were also the same as in the second grade. However, this time there were two maximum scores. In the traditional scoring system the maximum score was 21, one point per

one correct repetition. In the Baddeley (1993) version of scoring the maximum was eight, since the score was the number of syllables in the longest nonword that was repeated correctly until all three nonwords of certain length were repeated wrong.

3.6 English Nonword Repetition Test

The English nonword repetition was done in October 2002, in the fifth grade, within a few weeks from Finnish nonword repetition test 2. All sixteen remaining participants were available for this test.

Design of the Test. For reasons of comparability, the English nonword test carried out in the fifth grade also followed the slightly modified pattern of testing that was used in Finnish nonword repetition test 2. In fact, the two nonword tests carried out in the fifth grade were very much alike, only there were fewer nonwords in the English test, three of each syllable length from one to six, a maximum total of 18 nonwords. In constructing the test, the participants' level of knowledge of English had to be taken into account as well, and since the participants had been studying English for little less than five school semesters, a maximum of six syllables was considered suitable.

A native English speaker was not available for the construction of this test, so the nonwords were not created specifically for this task. Instead, stimuli were adopted from previous studies, in which native speakers have been involved and the nonwords have complied with English phonotactics (Gathercole, Willis, Emslie, and Baddeley 1991 and Gathercole 1995). The one-syllable nonwords were taken from Gathercole, Willis, Emslie, and Baddeley (1991) and the ones with two to five syllables were taken from Gathercole (1995). The six-syllable

nonwords were made by combining shorter nonwords. The stimuli are listed in Appendix 3.

Once the list of nonwords had been constructed, a male native speaker of English read them on tape. A person with a British rather than American accent was chosen because most of the material used in the participants' English class is spoken with a British accent. The procedure was the exact same as in the second Finnish nonword repetition test. Only the wording of the instructions (read in Finnish by the same person as in the two previous tests) was a bit different. The English nonword repetition test was piloted together with the Finnish one.

Procedure. The procedure was exactly the same as in the second Finnish nonword repetition test. The English test was done a few weeks after the second Finnish one and again all participants were tested within a week from each other. The scoring criteria were somewhat different. Clearly wrong phonemes were not accepted (like /p/ for /b/), but the participants were allowed to sound Finnish in pronouncing the nonwords, for example a Finnish [s] for [ʃ].

3.7 Knowledge of English

An important part of the present study was to use a more comprehensive measure to assess the participants' knowledge of English, one that is as diversified as feasible under the circumstances of this study and not based on vocabulary alone. Unlike in previous studies, there was an attempt to form a clear conception of L2 learning and base the language measures on that.

The conception of language adopted in the present study comes from Bachman and Palmer (1997). They use 'language ability' as a

general term, which they then divide into strategic competence and language knowledge. Strategic competence refers to the metacognitive strategies needed in using language. Language knowledge consists of pragmatic and organizational knowledge. Organizational knowledge in turn divides into textual and grammatical knowledge. What was feasible to include in the present study were knowledge of vocabulary and syntax, which in Bachman and Palmer's (1997) categorization are part of grammatical knowledge. This is only a tiny portion of what Bachman and Palmer (1997:66–75) include in language ability or even in language knowledge. However, using language in the tasks measuring language knowledge requires strategic competence as well. One limitation of the measures of language knowledge used in the present study is that mostly written tasks were used. Furthermore, what was understood as language knowledge in the present study was to some extent dictated by the participants. What fifth-graders could be expected to do was very limited, and normal school work could not be disturbed too much for the purposes of this study.

To have a workable model of language knowledge is only a starting point. The tasks and tests used to measure knowledge of L2 have to be paid attention to as well. The purpose of measurement defines the types of tasks that should be used. In short, the focus of interest may be in what the participants know about the language (achievement), or it may be in their more general ability to use and learn languages (proficiency). The intention in the present study was to use different tasks to measure the participants' knowledge as it was at the moment of measuring. This measurement would then signify what the participants have learned in the two and a half years of studying English, their success in learning the language. The aim was not to predict their future success in English studies or ability to learn languages. For these reasons the tasks used to measure knowledge of English should be viewed mostly as achievement tests, and not as test

of English proficiency. The decision to measure achievement was also the reason for choosing the term 'knowledge' instead of 'ability', for example. The terminology regarding this division in L2 performance assessment is quite varied. See, for example, McNamara (1996) for a summary of views of the concept and terminology.

A separate vocabulary task (section 3.7.1) was included mainly because the bulk of previous research (see sections 2.3 and 2.4) has found it to be positively correlated with nonword repetition and even suspect that the connection between phonological working memory and L2 learning in general may be due to the mediating influence of vocabulary (e.g. Service 1992). Knowledge of syntax was measured with six tasks from the exam given to the participants by their English teacher and a preposition task carried out for the purposes of the research project. Without a doubt the tasks claimed here to measure knowledge of syntax measure much of other kinds of knowledge as well. They at least had the potential to tap the participants' textual and sociolinguistic knowledge (the listening and reading comprehension tasks especially). However, at the present point of learning English, after five semesters, the texts were not very long and complicated, at least in the latest exam. Therefore, it seemed fair to assume that such tasks, too, measured mostly knowledge of syntactic structures in addition to vocabulary.

3.7.1 Vocabulary Task

The vocabulary exams the English teacher gives in class only cover a few texts in the English textbook and the learners study those words particularly for the exams. To get a vocabulary score that somehow reflected a broader knowledge of English vocabulary, a separate vocabulary task had to be constructed. A written constructed answer

format task was chosen partly to save time. In this task the participants simply wrote down the Finnish words that corresponded to a list of English words and vice versa. The participants could have been tested individually and orally, and perhaps this would have been more in consensus with the oral nonword repetition tests, but this would have taken a lot of the participants' time.

Had an oral vocabulary test been chosen, a number of decision would have had be made regarding the presentation of stimuli (live or taped), possible visual cues if the experimenter was present in the testing situation, scoring (on line or from tape), how much time should have been allowed for the answers, and whether the participants could go back and correct their answers. Furthermore, such a test format would undoubtedly have been new and strange to the participants as compared to a paper-and-pencil version. Therefore, it was assumed that the participants were more likely to perform to the best of their ability in a familiar written test format and suffered less from possible inhibitions. The written format was clearly much simpler to carry out as well.

Vocabulary was tested with isolated words to minimize the effect of any other kind of knowledge on the vocabulary score. A constructed answer format was chosen because it had been found the most suitable for this type of testing (Takala 1984). The participants were encouraged to write down something, if they had any idea of the correct word and even if they were not sure of the correct spelling. It was settled that 60 items would suffice to give a reliable estimate of the participants' knowledge of vocabulary. Even though Takala (1984) found no difference between the active and passive English vocabularies of Finnish learners of English, the separation was still made in testing the vocabulary, mostly because there were no grounds for preferring one or the other. Thus, the participants were asked to

translate 30 items from Finnish into English and 30 from English into Finnish.

Design of the Task. The decisions made by Takala (1984) in selecting test items were also the guidelines in the present study. The items were selected at random from the vocabulary lists in the back of the English textbook currently used in the participants' class, *Yes Adventures* (Westlake, Lintunen, Pitkänen, and Satamo 1997). The lists contained words and phrases from all the books previously used, including workbooks. There were separate lists for translations from Finnish into English and vice versa. All items were numbered and 30 items were selected from each of the two lists at random using the Research Randomizer (www.randomizer.org). It was decided in advance which items would be rejected. Those included adverbs that were derived from adjectives regularly, inflected forms of verbs and nouns, inflected forms of irregular verbs, and articles. Items from texts not yet studied were excluded. However, irregular verbs, proper names, phrases, idioms and structural words, for example, were included. The items were arranged in an alphabetical order for the vocabulary task (Appendix 4).

There was a pilot test with an adult female speaker of Finnish to check that the (oral) instructions were sufficient, what possible problems might arise for the test taker or the experimenter in scoring, and to approximate the time needed for the task. No problems arose during the piloting of the test.

Procedure. The vocabulary test was given to all the participants during the same day in December 2002 at the beginning of their English class. Since the participants were divided into three different groups, they took the test at different times. The whole group of students in the class always took the test, even though many of them were not participants

in the research project. Oral instructions were given before handing out the test papers. There were no instructions on the paper itself. The students were instructed to take as much time as they needed and to write something down if they had the slightest idea of the correct translation equivalent. The test took about ten minutes in each of the three groups.

One point was granted for the correct translation, zero for anything else. Thus the maximum score for the vocabulary test was 60. Also other translations besides those offered in the textbook were accepted. In the Finnish translations of English stimuli, inflecting the Finnish word did not matter (e.g. *seuraa* for *follow* was accepted). In the English translations of Finnish items, minor misspellings did not reduce points (e.g. *woter* meaning 'water'). More serious misspellings did not matter either, if it was obvious from the answer that the participant knew what the word sounded like, but could not spell it (e.g. *beasy* meaning 'busy'). The exact correct graphological form was not required from such young and novice learners.

3.7.2 Other Tasks

Six of the tasks in the English exam given by the teacher were used as measures of the participants' knowledge of English: Listening comprehension, Reading comprehension, Grammar tasks 1–3, and Sentence formation 1. All but Listening comprehension were scored again with slightly different criteria for the purposes of the present study. As a general rule, points were not reduced for spelling mistakes if it was clear which word was meant.

The listening comprehension task was not taped or a longer text of any kind; instead the teacher read ten English questions aloud twice. On the exam paper there were 12 one or two word answers and the

participants had to place the number of each question in front of the correct answer (Appendix 5). Two answers were left blank. The questions did not necessarily have anything to do with the participants themselves or even general knowledge, but were such that there was only one answer that fit the question semantically and grammatically, for example, *What language do you speak?* Correct answer: *Spanish.* One point was given for each correct answer, so the maximum score for the listening comprehension task was ten.

In the reading comprehension task the participants read a short English story written in diary-fashion, recounting the events of one school day (Appendix 6). There were ten questions on the text in Finnish and the participants answered them in Finnish. One point was given for each correct answer, so the maximum score was 10 points.

Three of the exam tasks were structurally rather similar grammar tasks. They were all cloze tests in which the participants filled in the blanks with the correct verb in a correct form. In Grammar task 1 there were ten separate English sentences each missing the verb (Appendix 7). There were 12 different verbs to choose from and the participants also had to inflect the verbs to agree with the subjects in the sentences. Half a point was given for the correct choice and half for the correct inflection. The maximum score for the task was thus ten. In Grammar task 2 there were, again, ten separate English sentences with blanks (Appendix 8). The participants had to fill them with the one of the given options: third person *–s*, *do*, *does*, *don't* or *doesn't*. One of the gaps remained blank. One point was given for each correct choice making the total ten points. Grammar task 3 was a short story told in the first person (Appendix 9). There were ten blanks and the participants filled them by choosing between the verbs *am*, *is*, *are*, *have*, and *has.* The maximum total was ten points, one point per correctly filled gap.

In the last task of the exam, Sentence formation 1, there were two pictures, a boy and a girl (Appendix 10). The participants were asked to

write about the children based on Finnish cue words (e.g. *ei koskaan tuhma*, *tappelee siskonsa kanssa*). Four sentences on each picture were expected. Five points were given for each correct sentence, so the maximum score for the task was 40. Full points were given if the sentences corresponded to the cue words and were grammatically correct (e.g. correct word order and verb inflection). One grammatical error, for example, reduced one point: A sentence like *He help his mom* was worth four points because the third person *–s* was missing. On the other hand, points were given if there was anything correct about the answer: One point was given for an answer like *No naughti*, when the expected answer was *He is never naughty*.

The additional preposition task, Sentence formation 2, which was carried out for the purposes of the research project was partly constructed by the English teacher as well (Appendix 11). The participants had had a practice exam with a task testing their knowledge of English prepositions. The picture from that task was used and the participants were asked to give ten answers to questions about the location of certain objects in the picture. The participants had to answer with a full sentence (e.g. Question: *Where is the chair?* Correct answer: *The chair is behind the tree.*) The basic sentence structure in all of the ten questions and expected answers was the same. The present author scored the sentences considering the word order, use of articles, prepositions and verbs. Six points were given for the right answer in the form of a grammatically correct sentence, so the maximum score for the task was 60. Communicativeness of the answer was considered and rewarded in the sense that a prepositional phrase instead of a whole sentence (e.g. *Behind the tree*) gave as much as four and a half of the maximum six points. Had the questions been asked in an authentic situation, mere prepositional phrases would have sufficed very well, they would have been correct and perhaps even more natural answers than complete sentences.

3.8 Data Analyses

The data from all the tests and tasks described in sections 3.4–3.7 above were entered into SPSS 11.0 for statistical analysis. The composite score of knowledge of English was formed by adding together the scores on the eight tasks (Vocabulary task, Listening comprehension, Reading comprehension, Grammar tasks 1–3, and Sentence formation 1 and 2). Since the maximum scores in the different tasks were not the same, some tasks have more weight in the composite score. The vocabulary task and the sentence formation tasks gave the most points, but they also required more work on the participants' part. Because they were longer and thus more demanding tasks, it did not seem unreasonable that they also were more emphasized in the composite score of knowledge of English. To test the research hypotheses, Spearman rank-order correlations were calculated for the nonword repetition tests and knowledge of English. Spearman correlations were used instead of Pearson because there were so few participants and Spearman, being a nonparametric test, gives more reliable results in such situations. When calculating Spearman correlations the variables do not have to come form a bivariate normal distribution, and the data do not have to be so strictly continuous. A one-tailed test of significance was used instead of a two-tailed one, because based on previous research (e.g. Service 1989) it could be expected that there would either be a positive correlation or no correlation at all.

4 RESULTS

4.1 Descriptive Statistics

Descriptive statistics of the tasks used to measure knowledge of English in the present study are given in Table 2. The table gives some idea of what the tasks were like in relation to each other, for example, which tasks were emphasized more in the composite score, and which tasks appeared to be easier for the participants. Possible reasons for the data behaving as they do are discussed in chapter 5.

Table 2. Descriptive statistics of the measures of English.

Knowledge of English	n	k	min	max	mean	s.d.
Composite score	15		66.75	194.50	129.97	39.21
Vocabulary	16	60	18.00	52.00	32.25	10.34
Listening comprehension	16	10	2.00	10.00	6.56	2.34
Reading comprehension	16	10	.50	10.00	6.34	2.90
Grammar task 1	16	10	3.50	10.00	6.89	1.75
Grammar task 2	16	10	1.00	9.50	5.56	2.41
Grammar task 3	16	10	4.50	10.00	7.41	1.86
Sentence formation 1	16	8	5.50	37.50	22.25	9.37
Sentence formation 2	15	10	18.00	58.00	42.47	12.85

n = number of participants
k = number of items

All 16 participants took part in all other tasks but Sentence formation 2. The vocabulary task had the most items, but the participants scored the highest average of points on Sentence formation 2. The vocabulary task, Sentence formation 1 and Sentence formation 2 were the three tasks to weigh most in the composite score of knowledge of English.

The participants' scores on all of the tasks used in the present study are listed in Appendix 12. The participants are ordered according

to the composite score of knowledge of English starting with the highest scorer.

4.2 English Nonword Repetition and Knowledge of English

The first research question was whether children's success in the English nonword repetition task was connected to their knowledge of English. Correlations were calculated between the scores on the English nonword repetition test and the composite score of knowledge of English, as well as each language measure separately (Table 3). As stated in the research hypothesis, positive correlations were expected. One participant was missing from the preposition task, so scores were only available from 15 participants in that task and therefore also in the composite score of knowledge of English and the corresponding tests of correlation.

Table 3. Correlations between the language measures and the English nonword repetition test.

	English nonword repetition test	
	Traditional score	Baddeley score
Knowledge of English		
Composite score[a]	.765**	.389
Vocabulary task	.596**	.311
Listening comprehension	.426*	.118
Reading comprehension	.644**	.282
Grammar task 1	.568*	.039
Grammar task 2	.512*	.055
Grammar task 3	.469*	.321
Sentence formation 1	.691**	.472*
Sentence formation 2[a]	.797**	.369

[a]n = 15
*p < .05; **p < .01

The Spearman correlation coefficients (?) are listed in Table 3 separately for the two ways in which nonword repetition attempts were scored. The traditional score means the total number of nonwords repeated correctly. The Baddeley score is the number of syllables in the longest correctly repeated nonword (as in Baddeley 1993).

The null hypothesis (H_0) in a correlation test is that there is no correlation between the two variables studied. Rejecting H_0 means accepting an alternative hypothesis, which in this case means assuming a positive correlation. The asterisks in Table 3 (and in following tables as well) result from a one-tailed test of significance. One asterisk ($p < .05$) after a correlation coefficient signifies a 95% chance that a mistake is not made if H_0 is rejected, i.e. a positive correlation is assumed. Two asterisks mean the same, only with 99% certainty ($p < .01$). To put it simply, the asterisks mark statistically significant positive correlations.

Repetition of English nonwords (traditional score) and knowledge of English appeared to be significantly correlated (? = .765, $p < .01$). This result supports the research hypothesis and indicates that phonological working memory is connected to the success in L2 learning. At least the measures of vocabulary, reading comprehension, and the two sentence formation tasks correlated separately with nonword repetition. As a statistical tool, calculating correlations for small samples is somewhat precarious and the results must be interpreted very cautiously unless ? is big enough to yield a p-level close to zero. Even though the test of significance takes the small number of cases (participants) into consideration, the coefficients with one asterisk should not be taken at face value. These weaker correlations can be examined further from scatterplots shown in Figure 3.

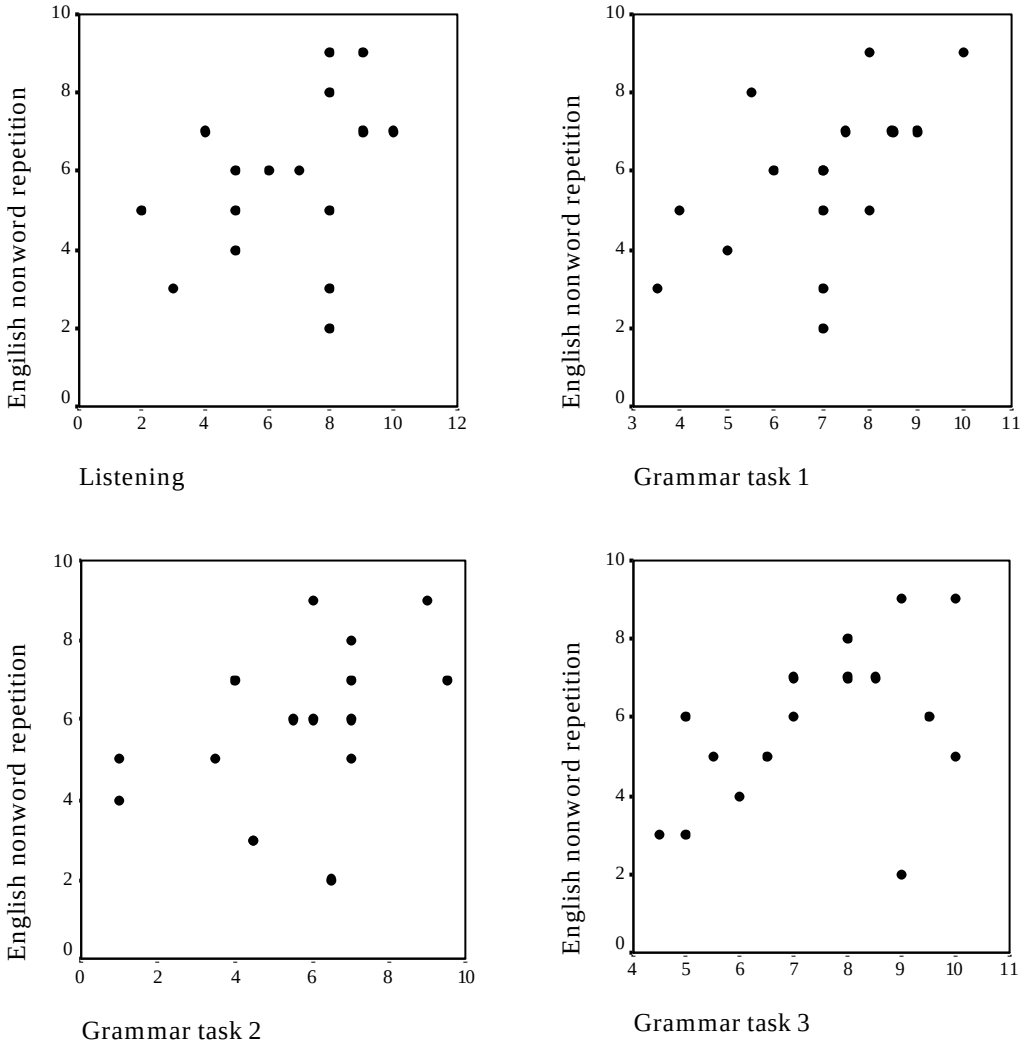


Figure 3. Scatterplots showing the English nonword repetition score as a function of Listening comprehension and Grammar tasks 1– 3.

For Spearman ? to be 1.0 (perfect positive correlation) the correlation between two variables does not have to be linear, which is the case with Pearson correlation. A perfectly linear correlation would show as a straight diagonal line formed by the cases in the scatterplot. For ? to equal 1.0 it is sufficient that the rank-order is the same for the two variables being compared, i.e. the variables further right are also higher up in the coordinates. It is rather obvious from the scatterplots in Figure 1 that even though there appears to be some correlation between English nonword repetition and the three language measures, it is not very clear. There is some rising tendency from left to right in the plots,

but the cases are quite spread out. This confirms the conclusion that with such few participants, a correlation which is significant at a .05 level is not yet a clear sign of the two measures being closely connected.

The repetition scores given in the more strict method following Baddeley's (1993) example did not correlate with the measures of knowledge of English for the most part. There was only a weak correlation between the Baddeley score and the sentence formation task of the English exam (? = .472, $p < .05$). It was apparent that the strict scores are not usable for the purposes of the present study and they are henceforth not reported.

4.3 Language-specificity of Phonological Working Memory

The second research question was about the possible language-specificity of phonological working memory. If it indeed were language-specific, scores on Finnish nonword repetition tests should not correlate with measures of knowledge of English. Nor should the Finnish nonword repetition scores correlate with scores on English nonword repetition. The exact hypotheses were: There is no correlation between the second grade Finnish nonword repetition score and the fifth grade score of knowledge of English. There is no correlation between the fifth grade Finnish nonword repetition score and the fifth grade English score. There is no correlation between the second grade Finnish nonword repetition score and the fifth grade English nonword repetition score. There is no correlation between the fifth grade Finnish nonword repetition score and the fifth grade English nonword repetition score. The correlations were calculated and are given in Table 4.

Table 4. Correlations between Finnish and English nonword repetition and knowledge of English.

	Finnish nonword repetition	
	test 1	test 2
English nonword repetition	.416	.248
Knowledge of English		
Composite score[a]	.501*	.301
Vocabulary task	.356	.148
Listening comprehension	.262	.096
Reading comprehension	.328	.143
Grammar task 1	.388	.322
Grammar task 2	.272	-.060
Grammar task 3	.539*	.580**
Sentence formation 1	.439*	.348
Sentence formation 2 [a]	.459*	.223

[a]n = 15
*p < .05; **p < .01

For the most part the correlation coefficients support the four research hypotheses. Neither Finnish nonword repetition test 1 or 2 seem to correlate with English nonword repetition (? = .416 and ? = .248, respectively), so two of the four hypotheses are supported. There appear to be some weakly significant correlations between Finnish nonword repetition test 1 and some English measures (the composite score, Grammar task 3, and the two sentence formation tasks), but they are not very convincing evidence against the research hypotheses. Even the scatterplot of the strongest of the four correlations indicates no clear relationship (Figure 4).

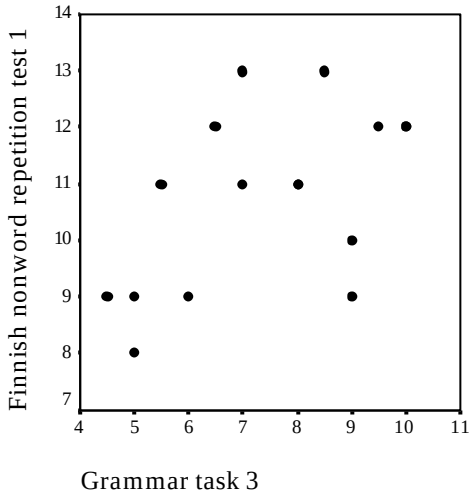


Figure 4. Scatterplot showing the Finnish nonword repetition test 1 scores as a function of scores on Grammar task 3 (? = .539, p < .05).

On the contrary, the correlation between Finnish nonword repetition test 2 and Grammar task 3 is quite significant (? = .580, p < .01), but it is not enough to cause the composite score to correlate with the repetition test. Altogether, these results support the four research hypotheses and indicate that Finnish nonword repetition is not connected to English nonword repetition or knowledge of English. As a consequence – since it has already been shown in the results to the first research question that English nonword repetition and knowledge of English correlate – phonological working memory appears to be language-specific.

4.4 Development of Phonological Working Memory

The third and final research question was whether the same children succeeded well in the Finnish nonword repetition task in the second and fifth grade. Essentially this meant inspecting the possible development of the phonological working memory.

The research hypothesis presumed that there would be a positive correlation between the scores on Finnish nonword repetition test 1 and

2, i.e. second and fifth grade. The Spearman rank-order correlation coefficient for the tests was .512, which is statistically significant (p=.021). Therefore, the research hypothesis is technically supported, but has to be interpreted with caution because of the small number of the participants. Again, a scatterplot helps to illustrate the alleged correlation (Figure 5).

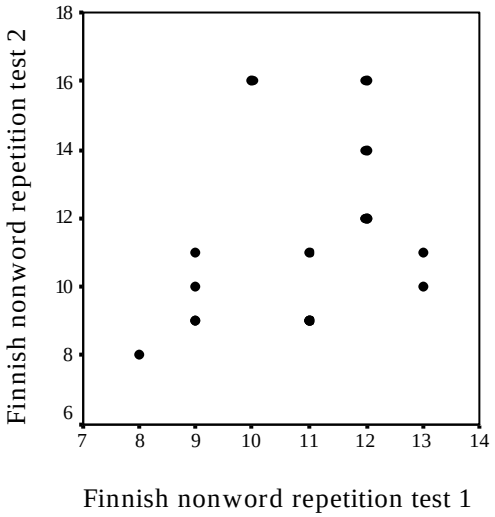


Figure 5. Scores on Finnish nonword repetition test 2 as a function of scores in Finnish nonword repetition test 1.

The rather week correlation is apparent in the scatterplot displaying the test 1 scores in relation to test 2 scores. The cases (participants) are spread out widely instead of forming an ascending line from the lower left corner of the coordinates to the upper right corner, so a high score in the second grade does not necessarily correspond to a high score in the fifth grade.

Hypothetically, if one case scoring rather low (10) in the second grade and high (16) in the fifth grade was excluded from the data, Spearman ? would rise to .640 (p = .005). However, there are no real grounds from dismissing the case in question as an outlier, so this serves only to illustrate how sensitive the coefficient is to exceptional cases.

4.5 Reliability and Validity

The issues of reliability and validity are important to consider in any research, but especially in studies like the present one, which uses quite a few tests and tasks, quantification and statistical analysis. Since all results and conclusions are based on scores given on the different tasks, it is critical to ensure the reliability and validity of the tasks. The scores as well as the results and conclusions based on the scores are of no consequence, if the tasks do not measure what they are claimed to measure or if they could well give very different results every time they are used. In addition to trying to ensure test reliability and validity beforehand, there are several ways to examine them after the fact. There are at lest four commonly known methods for estimating reliability: test-retest, parallel test, interrater reliability, and internal consistency. Validity can also be studied in different ways. Depending on the test, it is possible to focus on content validity, predictive validity or construct validity. For a more extensive discussion on the issues of reliability and validity and the methods to assess them, see, for example, Hatch and Lazaraton (1991), Hatch and Farhady (1982), and Scholfield (1995). As regards the present study, the reliability and validity of the nonword repetition tests and measures of knowledge of English are considered separately below.

Nonword Repetition Tests. The emphasis in the present study was on careful test construction and planning which allegedly ensured test reliability beforehand. The nonword repetition test is by nature such that most ways of testing reliability cannot be applied to it. In theory it would have been possible to carry out retests, but the intention was to use as little of the participants' time as possible. The same applied for using parallel tests. In addition, it would have been difficult to ensure

comparability of the tests, since so many factors make up the nonword repetition test and affect its difficulty, for example. If the nonwords of a test and parallel or retest were very much alike, there could be some effect of learning, and if they were very different, it would be difficult to estimate which test was easier or more difficult. The three different nonword repetition tests which were used in the present study could not function as retests or parallel tests to each other, since they mimiced different languages or took place several years apart. One further measure of test reliability, internal consistency, was not relevant in a nonword repetition test, since the stimuli were intended to vary in difficulty and the two later nonword repetition tests got systematically more difficult towards the end.

One possibility for studying the reliability of nonword repetition tests would be to compare the stimuli and the participants' responses phoneme by phoneme to see which phonemes or phoneme combinations are too easy or too difficult. The problem with this method is that unless it is very clear that the poor quality of the stimuli causes the participants to produce faulty repetitions, it is difficult to decide what to do with such easy or difficult phonemes or items. Perhaps there is really nothing wrong with the stimuli, they just strain phonological working memory a little more or a little less. If, however, there are phonemes that are difficult to hear, as the [l] in *aalsannokuurste* in Finnish nonword repetition 2, it might be best to exclude such items from the data, even though some participants manage to repeat them correctly regardless. Hearing is a natural and necessary part of a nonword repetition test, but if the stimuli are very weak, the focus of the test shifts too strongly away from the phonological working memory component. It is impossible to remember and repeat correctly something that was not heard correctly in the first place, so in order for the nonword repetition test to actually measure phonological working memory, it is critical that the stimuli be as clear as possible. As for the

present study, no items were excluded from the data, even though there was at least one borderline case where the stimulus was rather weak.

As for the characteristics of the test and the testing situation, there were not many factors that might have weakened the reliability of the tests. As L2 learners the participants have come across situations where they have had to repeat words unknown to them, so the situation in the nonword repetition tests was not totally new or strange to them. The tests were also done in familiar school surroundings and they did not take very much time, so fatigue should not have been an issue either. A longer test is usually more reliable, and perhaps a few more nonwords could have been repeated in each syllable length category, but it was also possible that repeating meaningless words would have gotten boring quickly causing the participants' performance to weaken.

Since the stimuli were nonwords, understanding what the participant said could not be used as criteria for correct repetition. The scorer had to listen to every single sound to hear if anything was added, omitted, or misplaced. In general, scoring nonword repetition tests is very difficult and subjective, because it is based on what the scorer hears, just as the participants' performance is partly based on what they hear, not just on their phonological working memory capacity. What is heard and scored depends to some extent on the equipment, but the fact that the scorer expects to hear certain sounds may affect the score as well. For these reasons it is good to have a second opinion and compare the two scores. For the present data, a type of interrater reliability was calculated to estimate the reliability of repetition accuracy judgments. Even though scores given by only one scorer were used to calculate correlations in the present study, another scorer also listened to and judged the repetition attempts to enable reliability assessment. The measure used for the consistency between two raters was Cohen's kappa coefficient (?). The interrater reliability was very

good in all three tests: ? = .839 for the second grade Finnish nonword repetition test, ? = .850 for the fifth grade Finnish nonword repetition test, and ? = .750 for the English nonword repetition test. These coefficients were all significant at the .000 level, so the scores can be considered reliable.

The validity of the nonword repetition tests is also difficult to assess. In a way, the tests are experimental and the present research is just another attempt to validate or invalidate the test type as a measure of phonological working memory. The results of the present study coincided with those of previous research, which could be interpreted as a sign of some validity.

Also the practicality of a test is important to consider, when deciding on its usefulness to measure certain things. The nonword repetition test can be considered a very practical measure of phonological working memory inasmuch as it does not take a very long time, and it could even be carried out in a language lab with all subjects taking it simultaneously. It does not require very sophisticated technology, and since the repetition attempts are taped, scoring can be done carefully without a hurry and by several people.

The Measures of Knowledge of English. The composite score of knowledge of English was composed of scores on eight different language tasks. Six of the tasks (Listening comprehension, Reading comprehension, Grammar tasks 1–3, and Sentence formation 1) came from the English exam given as a part of regular English classes. Two of the tasks were part of the research project and only one of them (Vocabulary task) especially designed for the present study. The reliability of the composite language measure was good when measured in terms of internal consistency: The Cronbach's alpha coefficient for the eight tasks was .846.

All of the tasks included in the composite score of knowledge of English required the participants to write down their answers. There were no tasks requiring the participants to answer in speaking, or, for example, read out text. Therefore, there was no need to judge spoken language, which is always more unreliable as seen above in the case of the nonword repetition tests. Since the English measures were all done in writing and scoring criteria were clearly defined with no room for interpretation, there was no such concern for the correctness and reliability of the scores as in the case of the nonword repetition tests.

Even though all the tasks were rather short, there were so many of them that also retests or parallel tests were out of the question. In a sense the three grammar tasks as well as the two sentence formation tasks could be considered parallel tests to each other, since they measured at least partly the same things. The Spearman rank-order correlation coefficient for the two sentence formation tasks was .630, which was significant at the .05 level (**). The three grammar tasks correlated with each other equally significantly: Grammar task 1 and Grammar task 2: ? = .662**; Grammar tasks 2 and 3: ? = .592**; Grammar tasks 1 and 3: ? = .634**.

The reliability of exam tasks is a two-fold issue. On the one hand, the test situation, the types of tasks, and the environment were familiar to the participants. Furthermore, the history of using these types of tasks to assess L2 knowledge is very long, they are used widely, and the people who created the tasks in this case were the same as those making the books used in class, and could be considered professionals in this field. This all adds to the reliability of the tasks. At the same time, they are reasons for these language measures being very practical. On the other hand, the situation in any ordinary school language exam is such that only certain parts of vocabulary and grammar, for example, are focused on. The participants know this and study those particular things especially for an exam. How they then perform on the exam

could be interpreted in many ways. The results may reflect true knowledge and understanding of the language, or be a product of very intensive studying which may lead to the things being forgotten quickly after the exam.

Internal consistency measures were calculated for the vocabulary task to examine its reliability. 15 of the total of 60 words did not separate the participants at all, since all of the participants got the same score, one or zero, on them. Cronbach's alpha coefficient for the remaining 45 items was .947. Among the 45 items there were still a few that were not very reliable. These were items that otherwise poorly performing participants got right, or such that high scorers got them wrong. In other words, success on these items was based on chance or something else, but very likely not on knowledge. Removing these items would have made the internal consistency of the task somewhat better.

As to the validity of the measure, the eight individual tasks could be thought of as valid measures of the particular aspects of language knowledge they claim to measure. They are fairly short, simple and standard. In the case of the two sentence formation tasks, it was a matter of decision what aspects of grammar or vocabulary were checked and scored, and the decision was to take everything possible into consideration. The sentence formation tasks were the only ones to reward the participants for communicativeness, since some points were given for conveying even a part of the message or all of it, but with insufficient grammar.

If the content validity of the individual tasks as English measures is not questionable, the validity of the composite measure is quite a different matter. The scope of language knowledge covered by the eight measures used is so narrow that the composite score hardly represents a very comprehensive estimate of the participants' knowledge of English. However, this composite measure of knowledge of English

was as valid as was possible under the circumstances. Several tasks were included without straining the participants too much and the content of the tasks fit the participants and the purposes of the study. There was nothing too new or surprising for the participants, and they cooperated well. It could be assumed that the participants saw the exam tasks as important to do well and performed to the best of their ability, since it was a tool to assess their school work. Even though the participants knew the tasks done for research purposes would not affect their English or any other grade, they seemed to take the tasks seriously and indeed did their best.

5 DISCUSSION

English Nonword Repetition and Knowledge of English. The present study replicated the results from numerous previous studies: The positive correlation between the composite score of measures of L2 knowledge and the English nonword repetition score was statistically significant, ? = .765**. Since the correlation was quite strong, it seems apparent that phonological working memory capacity is connected to how well these Finnish fifth-graders had learned English as an L2.

Looking at the measures of knowledge of English separately, some of them correlated clearly more strongly with English nonword repetition than others. One aspect seems to be that the measures which correlated more strongly with nonword repetition required more of the participants own production of language. Sentence formation 2 shared the strongest correlation with English nonword repetition (? = .797**), and it was also a task that had a high maximum score and points were gained easily in this task. These aspects may explain part of the high correlation. Grammar tasks 1–3 and Listening comprehension only required the participants to choose from available alternatives, and they did not correlate as strongly with nonword repetition. An interesting point to notice is that Listening comprehension shared the weakest correlation with English nonword repetition (? = .426*) even though it was the only task of the measures of knowledge of English that involved spoken language as did the nonword repetition test. Perhaps another similar study with more participants, different types of tasks to measure the same aspects of knowledge of English, and then using, for example, regression or factor analysis would clear up the relationship between these types of variables, and between the different kinds of variables and nonword repetition.

Compared to the four studies reviewed above that focused on children as L2 learners, the nonword and English knowledge measures used in the present study were intentionally somewhat different. Yet the results were very similar. Three of the four previous studies (Service 1989, Service and Kohonen 1995, and Dufva and Voeten 1999) only used two- and four-syllable nonwords in the repetition tasks, and Cheung (1996) only used two-syllable nonwords although he connected them in sequences of a variety of lengths. Evidently such short nonwords were not a problem in the English (L2) version of the nonword repetition test: They were not too easy, and yielded results which could be analyzed further. However, since the aim of the present study was to get results with a Finnish nonword repetition test as well, and the English version had to be similar for reasons of comparability, there was a much wider range of syllable-lengths in the stimuli of the present study. Surprisingly, this turned out to benefit the participants, since they often missed the short English nonwords but improved on the longer ones.

Service and Kohonen (1995) measured knowledge of English quite broadly, but ended up claiming that there is no connection between nonword repetition and other English measures besides vocabulary because partialling out vocabulary knowledge with regression analysis led to nonword repetition and the other English measures no longer correlating. One of the studies focused mainly on vocabulary learning (Cheung 1996), and another only had one other language task besides vocabulary knowledge (Dufva and Voeten 1999). Despite the discrepancies in modes of measurement, all of the four previous studies confirmed the connection between phonological working memory and knowledge of L2 English, so the present results fit in perfectly.

A further interesting result was the lack of correlation between measures of knowledge of English and the more strict nonword repetition scores modeled after Baddeley (1993). The strict score worked

for Baddeley (1993) to separate the subject with a memory deficit from the controls, but apparently it did not separate the participants of the present study enough. This ability to separate the participants from each other seems one likely reason for the strict score not correlating with knowledge of English. When ranked according to scores on the English nonword repetition test, the traditional score divided the 16 participants into eight different ranks, with maximum three participants getting the same rank (ranks 3–5). The strict Baddeley score only assigned the participants to six different ranks with as many as six participants getting the third best rank. The rankings of the participants varied very much between the two scoring methods. The fact that there were a different number of ranks in the two kinds of scores blurs the comparison a little, but to give some idea of the differences, only four participants got the exact same rank in both scores. Within the rest of the participants, the differences were great. Most ranked higher in the strict score, a few ranked lower. This puts the term 'strict score' into a strange light! The range was from ranking two places lower to ranking four places higher. It seems a small wonder that the traditional score and the Baddeley score even correlated (? = .688**), even though it is only natural that very good or poor performance on a task will inevitably show in the rank regardless of scoring method.

One obvious difference between the two scoring systems was that if the participant missed all three of the one-syllable nonwords in the very beginning of the test, the strict score was automatically zero and the rank was the worst possible. In the traditional way of scoring those participants still had a chance. There were three such cases, and they ended up scoring two, three, and six points in the traditional method, which entitled them to ranks four, seven, and eight (of the eight different ranks). The one-syllable English nonwords were very difficult and the participants as a group were not very successful in repeating them. They scored altogether 17 points on the one-syllable

nonwords, whereas the two-syllable nonwords were the easiest of all (30 points altogether).

The difficult one-syllable nonwords caused some participants to rank very poorly in the Baddeley score of English nonword repetition. In that sense it was a strict scoring method. On the other hand, it was a very lenient score, since even one correct repetition per syllable-length gave the full points for that length. Knowing six of the eighteen nonwords would have given the full points in the Baddeley method, if only each of them was from a different syllable-length. This fact must have obscured the differences between the participants somewhat, since the full score was given on each syllable length irrespective of the number of correctly repeated nonwords: One nonword per syllable-length was as good as three.

Language-specificity of Phonological Working Memory. Previous research has indicated that phonological working memory may be language-specific (section 2.4.2). It is not clear whether this is because there are supposed to be completely different memory systems for different languages, or if the phonological working memory just functions a bit differently depending on the language used. In the present study language-specificity of phonological working memory was understood so that nonword repetition in a certain language is not connected to nonword repetition performance in other languages or knowledge of languages other than the one that the nonwords sounded like. Thus, the hypotheses connected to the second research question presumed that as a sign of language-specificity, neither of the Finnish nonword repetition tests should correlate with the English nonword repetition test or the measures of knowledge of English. The results complied with that fairly clearly. Finnish nonword repetition test 2 did not correlate with English nonword repetition or the composite score of knowledge of English. Finnish nonword repetition test 1 did not correlate with English

nonword repetition either, and it correlated only weakly (.501*) with the composite score of knowledge of English. Based on these results it can be concluded that phonological working memory appears to function language-specifically, i.e. L2 success contributable to phonological memory is not necessarily similar in all languages.

There appeared to be some correlation between the Finnish nonword repetition tests and some of the English measures. Grammar task 3 and the two sentence formation tasks correlated with Finnish nonword repetition test 1, but only weakly (.05 level of significance). These correlations may be ignored as too small to matter. However, the correlation between Finnish nonword repetition test 2 and Grammar task 3 was fairly strong, .580**. On the surface Grammar task 3 was not very different from the other two grammar tasks. They all required the participants to fill in blanks by choosing from the options given. The only thing that really set it apart from the others was that the blanks were in a short coherent story and not in separate sentences like in the other two grammar tasks. Nevertheless, since Grammar task 3 was the only measure of English that correlated with this test of nonword repetition, it alone did not justify abandoning the assumption of language-specificity.

The maximum score in all of the three grammar tasks was ten points, so the raw scores can be used to compare the tasks with each other. All of the tasks dealt with verb choice and inflecting them in the right person. They were also structurally similar in the sense that they all required filling in the blanks with one of the options given. In brief, the tasks were very similar. Descriptive statistics of the three grammar tasks indicated that Grammar task 3 was the easiest, because it had the highest mean of all participants. It also had the smallest range, so in a sense the participants performed more uniformly on Grammar task 3 than on the other two. However, it is irrelevant in terms of correlation

how well the participants as a group performed on any of the tasks. What matters is the rank-order of the participants.

The way language-specificity was understood in the present study is somewhat different form the two previous studies. Thorn and Gathercole (1999) measured their subjects' phonological memory and vocabulary knowledge in two languages. To them language-specificity meant that phonological memory (measured in a particular language nonword repetition test) reflects the vocabulary knowledge of that particular language. The present study used a more comprehensive measure of language knowledge than just vocabulary, comparing nonword repetition to a composite score of several language measures. Furthermore, the present study did not report any data regarding the participants' L1 knowledge, so unlike in Thorn and Gathercole (1999), L1 measures were not available for studying language-specificity. The Finnish nonword repetition tasks were only compared to English tasks and not expected to correlate.

Since L1 knowledge was not measured, L1 and L2 knowledge could not be compared. It is apparent that the participants would have been much more knowledgeable in tests of knowledge of their L1, Finnish, than they were in measures of knowledge of English. Being fifth-graders, their L1 knowledge is very good, but five semesters of a few weekly lessons of English has not yet made them very proficient L2 users.

Nonword repetition, on the other hand, was measured in the two languages, but it is questionable whether the results on Finnish nonword repetition 2 and the English nonword repetition test are comparable one-to-one. There were different numbers of nonwords and even the lengths differed, but if they are compared regardless, the participants did considerably better on the Finnish test (see Table 5).

Table 5. Descriptive statistics of the English nonword repetition test and the Finnish nonword repetition test 2.

	mean	min	max	R	s.d.
Finnish nonword repetition test 2	11.00	8	16	8	2.082
English nonword repetition	5.75	2	9	7	2.477

The mean of all participants in repeating Finnish nonwords was 11.00, whereas it was only 5.75 for the English nonwords. The range of the scores in each task was about the same, but on a totally different level: The scores on Finnish nonword repetition test 2 varied from eight to sixteen, whereas the scores on the English nonword repetition test varied from two to nine. These statistics point to Finnish nonword repetition being the easier test, which was only to be expected with native speakers of Finnish.

Yet another difference is that Thorn and Gathercole (1999) compared groups with different language backgrounds, whereas the participants of the present study were a very homogenous group in their L1 and L2 knowledge.

Also Masoura and Gathercole (1999) assessed their subjects vocabulary and phonological working memory in both L1 (Greek) and L2 (English), but they were interested in the formal context as L2 learning environment. Thus the subjects were in a rather similar situation as in the present study, but the interpretation of language-specificity differed. In Masoura and Gathercole's (1999) study the vocabulary knowledge of both languages correlated with nonword repetition of both Greek and English sounding nonwords and with each other. So far this does not sound very language-specific. However, when partial correlations were calculated, L2 vocabulary knowledge and the composite nonword repetition score correlated, while L1 vocabulary was partialled out, but when the situation was reversed, L1 vocabulary and nonword repetition did not correlate. Since there

seemed to be a connection between phonological working memory and L2, but not L1, Masoura and Gathercole (1999) drew the conclusion that phonological working memory was language-specific.

The two previous studies and the present study are perhaps too different from each other to justify saying anything definite about language-specificity of phonological working memory. On the other hand, the three different techniques and understandings – including the present study – have all pointed to language-specificity in one way or another, so perhaps there is some truth to it. Further research from all or some of these points of view might shed much needed light on the issue.

Development of Phonological Working Memory. The third research question dealt with the development of phonological working memory. The results showed a significant positive but weak correlation between Finnish nonword repetition tests 1 and 2. This indicates that if there has been development in the phonological working memory, it has been somewhat similar for all the participants, and those who were good nonword repeaters in the second grade are still good in the fifth grade. As such this result is in line with previous research. However, as shown in the scatterplot in Figure 5, the correlation is not very strong, and far reaching conclusions should not be based on that alone.

There are several possible reasons for the weak correlation between the two Finnish nonword repetition tests. If the participants' phonological memory developed similarly and equally fast and the two tests both measured the same thing, the correlation should have been strongly positive. Since it was not, something in the memory system or the tests is not as expected. It could be that the small number of participants is partly responsible for the results. If there were a few participants whose phonological working memory had developed early, and some who were behind in development in the second grade,

they would be likely to develop very differently within the two and a half years that were between the two tests. The early bloomers would not develop much and the slower ones would probably catch up. In such a small sample of participants a few such “odd” cases could interfere with results rather severely. It is another question if such development can be considered odd. It is completely natural for children of this age to vary to a certain extent in a number of physical as well as mental or cognitive attributes, why not in phonological memory?

On the other hand, Finnish nonword repetition test 1 was quite different from test 2 for reasons explained above in section 3.5. It is entirely possible that the two tests both measured phonological working memory but were so different in form and content that comparing the results does not give a completely accurate picture of the development that has taken place.

Based on previous research (see e.g. Gathercole and Baddeley 1993a), it is quite clear that phonological working memory develops during childhood and reaches adult level at around the age of twelve. What is less clear is how exactly this development happens and which factors affect it. Therefore, it is difficult to say if there has been development and what kind it has been solely based on a correlational study of two variables. More sophisticated measures are needed to make such judgments.

6 CONCLUSION

The present study examined children's phonological working memory with Finnish and English nonword repetition tests. The repetition results were then compared to each other and to measures of knowledge of English to see whether phonological working memory and L2 knowledge are connected, whether phonological working memory appears to be language-specific, and what the development of the participants' phonological working memory has been like between grades two and five.

Strengths of the Present Study. Perhaps the most important new information that the present study has to offer stems from the Finnish nonword repetition tests. Evidently it is possible to construct Finnish nonword repetition tests which actually yield results that enable further analysis. It is not a simple matter, however. There are many issues to consider both linguistically and methodologically, and the age of the participants needs to be taken into account, too. The properly functioning Finnish nonword repetition tests enabled comparisons between English and Finnish nonword repetition tests for the first time, and this in turn was essential for assessing the language-specificity of phonological working memory. The Finnish nonword repetition tests could also be used to examine the development of phonological working memory to an extent by comparing the Finnish nonword repetition scores from the second and fifth grade.

Also the more comprehensive than average measure of knowledge of English can be considered a strength of the present study. It had some limitations which will be discussed below, but it was quite diverse considering the circumstances. Several aspects of language knowledge were measured and there were different types of tasks. Furthermore, measures were not taken at the same time. Instead, they

were carried out within two months on three different occasions and in three slightly different situations, so they do not measure performance at exactly the same moment in time.

Reliability and validity of the tasks and scoring have also been paid attention to in the present study, but not at the participants' expense. As a rule, the more tasks are used to measure a particular construct, and the longer those tasks are, the better the reliability of the measure. However, as the measures get to be longer and more numerous, they also get more straining and demanding on the participants' concentration, which may in turn affect the reliability of the measure. The measures used in the present study were designed so that they were as valid and reliable as possible, but the participants' abilities and comfort were put first. Moreover, the students' daily schoolwork had to be taken into account.

Importance of the Results. The results of the present study gave additional support for there being a relationship between L2 knowledge and phonological working memory, as measured by nonword repetition. There was also some indication that phonological working memory may be language-specific, and that the development of the phonological working memory between the second and the fifth grade was similar for these children.

Although these results can be useful from L2 teaching and learning point of view, they mostly serve to clarify the concept of phonological working memory. Especially new information about phonological working memory in different kinds of L2 contexts is important for verifying or discrediting the alleged features of phonological working memory. There are countless combinations of learners' L1s and L2s and so far the interaction between the different languages and how it reflects on the functioning of phonological working memory is not very clear. If phonological working memory is

as essential for long-term language learning as claimed, it is important to know its structure and functioning as thoroughly as possible. This way it may be possible to predict, prevent or ease some language learning problems, and develop such learning methods that suit language impaired learners or make learning even easier for "normal" learners.

The present results seemed to indicate that phonological working memory and L2 learning are connected and that phonological working memory development was similar for the participants between the second and the fifth grade. If further research supports these results, it means that children with poorer than average phonological working memory remain so, at least in the near future, and they are doomed to being poorer than average L2 learners. As a consequence, children diagnosed to have poor phonological working memory should be supported in their language studies and possibly with developing their memory, since results indicate that they will not catch up on their own or be able to use compensating strategies at this age and stage of learning.

Limitations of the Present Study. The measures of L2 knowledge were a strength but also the most obvious weakness of the present study. Intentions were honorable: Unlike in most of the previous studies, knowledge of English was to be measured as widely as possible with thought-out measures, which rested on an accepted conception of language knowledge. However, reality set in in the form of common limitations. The participants could only be bothered so much, and there was no time and not enough personal knowledge to construct or carry out numerous language tasks. Even if the participants had been available for extensive testing, it would have taken too long to construct language tests which tapped into all imaginable aspects of English knowledge and at the same time were suitable for these particular

participants. Language testing is a discipline of its own, and getting acquainted with vast amounts of research would have been essential for proper test construction. In the future, it might be fruitful to look into research on language testing to get a better picture of what can be measured, what absolutely needs to be measured, and what might the best way be. It is unlikely that it will be possible to carry out a very thorough language assessment in any research project which is not especially focusing on language testing, but even the smaller tasks should be well-grounded.

For language-specificity studies it would have been important to get an estimate of the participants' L1 knowledge in addition to L2 assessment. This would have made assessments more reliable and given different possibilities for looking into the issue. The present study could only rely on measures of English (L2) and nonword repetition tests in L1 and L2 for speculations on language-specificity. A wider array of measures, including measures of knowledge of Finnish, will be available in the sixth grade when the participants take part in national school achievement tests, and the matter may be returned to at that point in time.

For studying the development of phonological working memory, it would be a good idea to try and construct two Finnish nonword repetition tests that are as similar as possible, yet suitable for children of different ages. This is undoubtedly a difficult task, but worth attempting. This way studying the development would be easier and more reliable. In the case of the present study, future aspirations were not yet clear while the second grade Finnish nonword repetition test was being constructed. Once it was decided that another Finnish test as well as an English one were needed in the fifth grade, there were several options for the format of the repetition tests. The one used in the second grade would still have been suitable for the English test, but too easy for the Finnish one. Since it was more important for the two fifth

grade tests to be comparable than for the second grade Finnish test and the fifth grade English test to be comparable, the old format was rejected, and the two Finnish tests did not turn out to be very comparable.

Had there been more time, a number of issues could have been studied and measured about the nonwords. Based on previous research, at least the wordlikeness of the nonwords would have been interesting to assess and it might have added to the interpretation of the results of the nonword repetition tests. It would also have been interesting to experiment with different aspects of the tests. For example, the influence of instructions could have been explored to some extent. While constructing the tests an interesting question arose, which had actually already been brought up by Wells (1995). It had to do with the wording of the instructions and whether the participants should be told to repeat or mimic the nonwords they heard on the tape. 'Repeat' was chosen because it had already been used in the second grade test, but it would have been interesting to try and find out if mimicking would have brought better results in the English nonword repetition test or if it would not have made any difference. As it was, most of the children did not attempt to imitate the accent of the British man speaking on the tape, but used their own way of speaking English, which, at this stage still, sounded quite Finnish.

Due to lack of knowledge but mostly due to the small number of the participants, the statistical analysis methods used in the present study were fairly limited. This could not be helped, however, since the present study was part of a longitudinal case study, and most of the research in the project is qualitative in nature anyway. If a similar study were carried out with a larger group of participants, a wider variety of statistical methods could be applied and the results could be considered more reliable. With only 16 participants, it would not have been

reasonable to use some methods which otherwise might have been very useful in revealing what lie behind the alleged relationships.

As is evident in the above, the present study leaves something to be desired, despite the high goals set in the beginning. However, most of these issues are not due to poor planning or lack of effort, but dictated by the circumstances.

Suggestions for Further Research. First and foremost, more research on phonological working memory in an L2 context is needed in any form. Perhaps the most obvious suggestion for further research rising from the present study is that more participants and longer tasks to measure knowledge of L2 be used in similar studies in the future.

In Finland as in many countries it is common that children start studying several L2s quite early. This gives many opportunities for studying the language-specific nature of phonological working memory. If a particular subject group's nonword repetition and language knowledge were measured in several languages, the results could then be compared in many ways and this could easily clarify the concept of phonological working memory or reveal something new about it. Results might show a clear division between how nonword repetition is connected to L1 knowledge and knowledge in the L2s. Perhaps there might appear great differences between the different L2s in how closely they are connected to nonword repetition in the respective languages or some languages may share very similar profiles in this respect. There may also be differences due to the age when learners start studying the particular languages. A study with subjects learning several L2s would help solve or at least clarify some of these issues.

Methodologically it might be interesting and useful to compare different types of nonword repetition tests. The CNRep format is clearly the most widely used, but there are several other possibilities and

probably also room for new or slightly modified test formats. A brief glance at previous research (e.g. section 2.4) might help give ideas of what has been done and how the tests could be developed further. Such experimentation may be essential when trying to find a test format that best suits a certain group of subjects, since age and language learning history are likely to affect subjects' phonological working memory and nonword repetition success.

The third suggestion goes beyond nonword repetition tests. Whether there turns out to be one or several phonological working memory systems, there appears to be a close connection with this concept of memory and success in learning languages. If problems in this memory system can be identified through a nonword repetition test, there also need to be some further steps taken to help learners with these kinds of problems. It may be up to psychologists or language teachers, but suitable language learning methods need to be developed or old methods analyzed to see which of them best serve learner needs.

Considering the results of the present study as well as the great variety of methods used and viewpoints taken in previous research, there most definitely are grounds for further research on the issue of phonological working memory and nonword repetition tests as such but especially in the context of L2 learning.

BIBLIOGRAPHY

Alanen, R. 2000. Vygotsky, van Lier ja kielenoppiminen: sosiokulttuurinen viitekehys kielellisen tietoisuuden ja vieraan kielen oppimisen tutkimuksessa. In P. Kalaja and L. Nieminen (eds.), Kielikoulussa – kieli koulussa. AFinLA Series 58. Jyväskylä: AFinLA, 95–120.

Alanen, R., in press. A sociocultural approach to young language learners' beliefs about language learning. To appear in P. Kalaja and A.M.F. Barcelos (eds.), *Beliefs about SLA: new research approaches*. Dordrecht: Kluwer Academic Publishers.

Aro, M. 2001. Lasten käsityksiä englannin ja suomen kielistä. Unpublished Pro Gradu Thesis. University of Jyväskylä.

Atkins, P.W.B, and A.D. Baddeley 1998. Working memory and distributed vocabulary learning, *Applied Psycholinguistics* 19, 537–552.

Avons, S. E., C. A. Wragg, L. Cupples, and W. J. Lovegrove 1998. Measures of phonological short-term memory and their relationship to vocabulary development, *Applied Psycholinguistics*, 19, 583–601.

Bachman, L.F., and A.S. Palmer 1997. *Language testing in practice.* Oxford: Oxford University Press.

Baddeley, A. 1986. *Working memory*. Oxford: Clarendon Press.

Baddeley, A.D. 1992. Working memory, *Science* 255, 556–559.

Baddeley, A. 1993. Short-term phonological memory and long term learning: a single case study, *European Journal of Cognitive Psychology* 5:2, 129–148.

Baddeley, A.D., and G. Hitch 1974. Working memory, in G. H. Bower (ed.), *The psychology of learning and motivation: advances in research and theory 8*, New York: Academic Press, 47–89.

Bishop, D. V. M., T. North, and C. Donlan 1996. Nonword repetition as

a behavioural marker for inherited language impairment: evidence from a twin study, *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 37:4, 391–403.

Bowey, J. A. 1996. On the association between phonological memory and receptive vocabulary in five-year-olds, *Journal of Experimental Child Psychology*, 63, 44–78.

Bowey, J.A. 1997. What *does* nonword repetition measure? A reply to Gathercole and Baddeley. *Journal of Experimental Child Psychology*, 67, 295–301.

Brady, S., E. Poggie, and M. M. Rapala 1989. Speech repetition abilities in children who differ in reading skill, *Language and Speech*, 32:2, 109–122.

Cheung, H. 1996. Nonword span as a unique predictor of second-language vocabulary learning, *Developmental psychology*, 32:5, 867–873.

Dollaghan, C. A., M. E. Biber, and T. F. Campbell 1995. Lexical influences on nonword repetition, *Applied Psycholinguistics*, 16, 211–222.

Dufva, H., in press. Beliefs in dialogue. To appear in P. Kalaja and A.M.F. Barcelos (eds.), *Beliefs about SLA: new research approaches*, Dordrecht: Kluwer Academic Publishers.

Dufva, H., R. Alanen, and K. Mäntylä 2001. Ei kissa ole mikään sana: lasten näkökulmia kieleen, in H. Sulkala and L. Nissilä (eds.), *XXVII Kielitieteen päivät Oulussa 19.–20.5.2000*. Acta Universitatis Ouluensis Humaniora B 41. Oulu: University of Oulu, 35 – 45.

Dufva, M., and M. J. M. Voeten 1999. Native language literacy and phonological memory as prerequisites for learning English as a foreign language, *Applied Psycholinguistics*, 20, 329–348.

Ekman, E.A. (ed.) 1899. *Suomen kielen keräilysanasto 1*. Helsinki: Suomalaisen kirjallisuuden seura.

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.

Gathercole, S. E., and A. D. Baddeley 1989. Evaluation of the role of phonological STM in the development of vocabulary in children: a longitudinal study, *Journal of Memory and Language*, 28, 200–213.

Gathercole, S. E., and A. D. Baddeley 1993a. *Working memory and language*. Hove: Lawrence Erlbaum Associates.

Gathercole, S. E., and A. Baddeley 1993b. Phonological working memory: a critical building block for reading development and vocabulary acquisition?, *European Journal of Psychology of Education*, 8:3, 259–272.

Gathercole, S.E., and A.D. Baddeley 1997. Sense and sensitivity in phonological memory and vocabulary development: a reply to Bowey (1996), *Journal of Experimental Child Psychology*, 67, 290–294.

Gathercole, S. E., C. Willis, and A. D. Baddeley 1991. Nonword repetition, phonological memory, and vocabulary: a reply to Snowling, Chiat, and Hulme, *Applied Psycholinguistics*, 12, 375–379.

Gathercole, S. E., C. Willis, H. Emslie, and A.D. Baddeley 1991. The influences of number of syllables and wordlikeness on children's repetition of nonwords, *Applied Psycholinguistics*, 12, 349–367.

Gathercole, S. E., C. S. Willis, H. Emslie, and A.D. Baddeley 1992. Phonological memory and vocabulary development during the early school years: a longitudinal study, *Developmental Psychology*, 28:5, 887–898.

Gathercole, S. E., C. S. Willis, A. D. Baddeley, and H. Emslie 1994. The children's test of nonword repetition: a test of phonological working memory, in S.E. Gathercole, and R.A. McCarthy (eds.),

Memory Tests and Techniques, Hove: Lawrence Erlbaum Associates,103–127.

Gathercole, S. E., G. J. Hitch, E. Service, and A. J. Martin 1997. Phonological short-term memory and new word learning in children, *Developmental Psychology* 33:6, 966–979.

Gathercole, S. E., E. Service, G. J. Hitch, A-M. Adams, and A. J. Martin 1999. Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship, *Applied Cognitive Psychology*, 13, 65–77.

Hatch, E., and H. Farhady 1982. *Research design and statistics for applied linguistics*. Cambridge: Newbury House.

Hatch, E., and A. Lazaraton 1991. *The research manual: design and statistics for applied linguistics*. New York: Newbury House Publishers.

Lehto, J. 1996. *Työmuistin yhteys tekstin tiivistämiseen, ongelmanratkaisuun ja koulumenestykseen*. Helsingin yliopiston kasvatustieteen laitoksen tutkimuksia 149. Helsinki: Helsingin yliopisto.

Masoura, E. V., and S. E. Gathercole 1999. Phonological short-term memory and foreign language learning, *International Journal of Psychology*, 34:5/6, 383–388.

Mayringer, H., and H. Wimmer 2000. Pseudoname learning by German-speaking children with dyslexia: evidence for a phonological learning deficit, *Journal of Experimental Child Psychology*, 75, 116–133.

McNamara, T. 1996. *Measuring second language performance*. London: Longman.

Metsala, J. L. 1999. Young children's phonological awareness and nonword repetition as a function of vocabulary development, *Journal of educational Psychology*, 91:1, 3–19.

Michas, I. C., and L. A. Henry 1994. The link between phonological

memory and vocabulary acquisition, *British Journal of Developmental Psychology*, 12, 147–163.

Muter, V. 1998. Concurrent and longitudinal predictors of reading: the role of metalinguistic and short-term memory skills, *Reading Research Quarterly*, 33:3, 320–337.

Papagno, C., and G. Vallar 1995. Verbal short-term memory and vocabulary learning in polyglots, *The Quarterly Journal of Experimental Psychology*, 48A:1, 98–107.

Papagno, C., T. Valentine, and A. Baddeley 1991. Phonological short-term memory and foreign-language vocabulary learning, *Journal of Memory and Language*, 30, 331–347.

Scholfield, P. 1995. *Quantifying language: a researcher's and teacher's guide to gathering language data and reducing it to figures.* Clevedon: Multilingual Matters Ltd.

Service, E. 1989. *Phonological coding in working memory and foreign-language learning.* Helsinki: Helsingin yliopisto.

Service, E. 1992. Phonology, working memory, and foreign-language learning, *The Quarterly Journal of Experimental Psychology*, 45A:1, 21–50.

Service, E., and V. Kohonen 1995. Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition?, *Applied Psycholinguistics*, 16, 155–172.

Snowling, M., S. Chiat, and C. Hulme 1991. Words, nonwords, and phonological processes: some comments on Gathercole, Willis, Emslie, and Baddeley, *Applied Psycholinguistics*, 12, 369–373.

Takala, S. 1984. *Evaluation of students' knowledge of English vocabulary in the Finnish comprehensive school.* Kasvatustieteiden tutkimuslaitoksen julkaisuja 350. Jyväskylä: Kasvatustieteiden tutkimuslaitos.

Taylor, H.G., D. Lean, and S. Schwartz 1989. Pseudoword repetition

ability in learning-disabled children, *Applied Psycholinguistics*, 10, 203–219.

Thorn, A. S. C., and S. E. Gathercole 1999. Language-specific knowledge and short-term memory in bilingual and non-bilingual children, *The Quarterly Journal of Experimental Psychology*, 52A:2, 303–324.

van Bon, W. H. J., and J. M. L. van der Pijl 1997. Effects of word length and wordlikeness on pseudoword repetition by poor and normal readers, *Applied Psycholinguistics*, 18, 101–114.

Weismer, S. E., J. B. Tomblin, X. Zhang, P. Buckwalter, J. G. Chynoweth, and M. Jones 2000. Nonword repetition performance in school-age children with and without language impairment, *Journal of Speech, Language and Hearing Research*, 43:4, 865–878.

Wells, B. 1995. Phonological considerations in repetition tests, *Cognitive Neuropsychology*, 12:8, 847–855.

Westlake, P., M-L. Lintunen, E-L. Pitkänen, and S. Satamo 1997. *Yes Adventures.* Helsinki: WSOY.

Appendix 1. Stimuli Used in Finnish Nonword Repetition Test 1

THE PRACTICE LIST

nulpposalukaaska

leiha

säpämytkä

köhnä

ruunivenka

tätyjäkkikämmy

THE TEST LIST

kaarmu

varelallo

sähykölsähäivi

täkkövörrisääte

torkopuokki

öystilönni

hauri

tätyhyylä

leisokeneponno

uppamesalemsa

arho

toho

naste

haasovinko

rotisireviippi

Appendix 2. Stimuli Used in Finnish Nonword Repetition Test 2

2	ruuppa
	vusti
	tanni
3	uoraste
	tuilasmo
	uroula
4	piuttuikere
	lärskätelo
	kitsamellu
5	suorskisynäyppö
	jiurunauppama
	aalsannokuurste
6	kuossakeetitarsko
	uispauttuhaahuumu
	vokkohotoiheltsu
7	mauhtotuirstoesniikki
	mauppohuottiallamo
	vuoksattiisisuilasmo
8	jiippakonnomeittomella
	ottiisiettaaskelane
	lumponeetriökysättäkkö

Appendix 3. Stimuli Used in the English Nonword Repetition Test

1	sep
	clird
	tull
2	diller
	hampent
	bannow
3	commerine
	brasterer
	skiticult
4	pennerriful
	blonterstaping
	empliforvent
5	versatrationist
	sepretennial
	altupatory
6	prindlefenneriser
	bannifertrumpetine
	smipdefermication

Appendix 4. Vocabulary Task

Käännä suomeksi.

1. an area _____
2. a beat _____
3. borrow _____
4. a class _____
5. cold _____
6. crawl _____
7. Egypt _____
8. a farmhouse _____
9. a field _____
10. floorball _____
11. follow _____
12. a guitar _____
13. a hotel _____
14. Lapland _____
15. last _____
16. later _____
17. a moped _____
18. a noise _____
19. of course _____
20. a panther _____
21. a screen _____
22. a seal _____
23. a sink _____
24. sleepy _____
25. some _____
26. thank you _____
27. walk _____

28. a way _____
29. who _____
30. a year _____

Käännä englanniksi.

1. avain _____
2. bussipysäkki _____
3. eteinen _____
4. haiseva _____
5. hauska _____
6. heitto _____
7. hyytelö _____
8. joukossa _____
9. jäätelö _____
10. karhu _____
11. katsoa _____
12. kiireinen _____
13. kyllä _____
14. leikata _____
15. makuuhuone _____
16. mitä _____
17. narsissi _____
18. näin _____
19. paloauto _____
20. reikä _____
21. satelliitti _____
22. syvä _____
23. terävä _____
24. tie _____

25. Tyyni valtameri _____
26. ukkonen _____
27. vesi _____
28. vihko _____
29. voimala _____
30. vuode _____

Appendix 5. Listening Comprehension

Numeroi vastaukset kysymyksiin 1–10 kuulemassasi järjestyksessä.

___ carrots
___ in France
___ dangerous
___ meat
___ Spanish
___ Copenhagen
___ polite
___ after school
___ Oslo
___ on Wednesday
___ gym
___ Friday

Appendix 6. Reading Comprehension

Lue tarina ja vastaa kysymyksiin suomeksi.

Dear Readers,

Today is Wednesday the 3rd of October. I got up at seven o'clock and I'm getting ready for school. To really wake up I need to wash my face with cold water and brush my teeth. Then I have my breakfast. My father always makes some breakfast for me. This morning I'm having a bowl of porridge and a glass of orange juice. They are very healthy and they make me strong.

After breakfast I go to school by bike. In that way I get some fresh air, and my school is not very far from my home. In the schoolyard I meet my best friend. We talk about last night's TV programme, which was very interesting. The programme was about nature.

We are a bit tired because we also studied for a test. Today we have a geography test. We had to study about Europe. We have to know everything about Sweden, Denmark and Norway. In the afternoon we'll have my favourite subject, needlework. Our teacher is friendly and nice and I enjoy learning to knit sweaters and sew shirts. But before that is our lunch hour. Today we're going to have vegetable soup and chicken salad. Yum! And buttermilk to drink. Hmm!

The last lesson is for a quiz. Interesting. The quiz is about American Indians. We have to make up the funniest Indian name. The winner gets a little toy skunk as a prize!

In the drama club in the afternoon we have to be people in the mirror world. A good person has to be bad and a quiet person has to be loud and noisy. It's very difficult to be something you're not in the real world. For me anyway.

Love, Bobbie

1. Mitä Pat tekee heti herättyään?
2. Millaisen aamupalan hän nauttii?
3. Miksi hän menee kouluun pyörällä?
4. Mistä hän keskustelee ystävänsä kanssa?
5. Minkä aineen kokeet hänellä on tänään?
6. Mistä pitää tietää paljon?
7. Mikä on hänen mieliaineensa koulussa?
8. Mitä hän syö lounaaksi?
9. Mikä on palkintona tietokilpailussa?
10. Mikä on vaikeaa näytelmäkerhossa?

Appendix 7. Grammar Task 1

Täydennä alla olevat verbit lauseisiin oikeassa muodossa. Kaksi verbiä jää yli.

collect	*understand*	*live*	*play*	*use*	*try*
hate	*speak*	*want*	*behave*	*ask*	*drink*

1. Eddy _____ a new computer.
2. Does Alice ever _____ football?
3. We don't _____ stamps.
4. Molly never _____ badly.
5. My mum doesn't _____ me.
6. My cat just _____ mice.
7. We _____ in a new house.
8. My little brother often _____ stupid questions.
9. Don't _____ to be funny.
10. Do you ever _____ English?

Appendix 8. Grammar Task 2

Lisää viivoille –s –pääte tai sana do, does, don't, doesn't. Yksi kohta jää tyhjäksi.

1. Sometimes my teacher ask _____ stupid questions.
2. Why _____ you want to be naughty?
3. We drink _____ coffee in the morning.
4. Ronnie _____ speak Italian.
5. What _____ Molly think about fast food?
6. Why _____ we have to get up so early?
7. _____ your cat eat cat food?
8. I _____ understand maths.
9. My dad _____ laugh at my jokes.
10. The Indians _____ live in teepees any more.

Appendix 9. Grammar Task 3

Täydennä alla oleviin lauseisiin am / is / are / have / has.

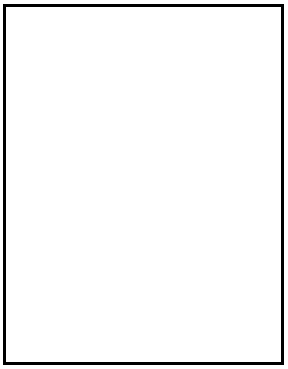
My name _____ Kate O'Malley. I _____ from Dublin, Ireland. I _____ not got a pet, but my friend Ian _____ got a tortoise. His name _____ Alfie. Alfie's legs _____ very short and thick. Alfie _____ ten years old, but still very lively. What pet _____ you got? _____ your pet got long ears? What do you think _____ the nicest animals in the world?

Appendix 10. Sentence Formation 1

Kuvassa on poika, Kind Ken. Kirjoita hänestä englanniksi vihjeitten perusteella.

ei koskaan tuhma — auttaa äitiään

tekee kotitehtävät — hänellä on paljon ystäviä



_____

_____

_____

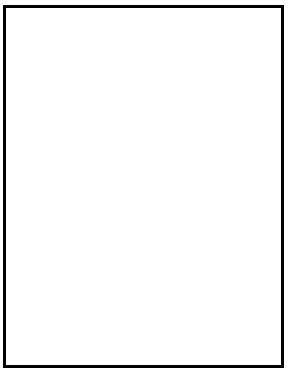
_____

_____

Kuvassa on tyttö, Awful Annie. Kirjoita hänestä englanniksi vihjeitten perusteella.

ei kuuntele isäänsä — syö paljon karkkeja

tappelee siskonsa kanssa — on hyvin tyly



_____

_____

_____

_____

_____

Appendix 11. Sentence Formation 2

Vastaa kysymyksiin kuvan perusteella. Muista vastata kokonaisella lauseella.

1. Where is the chair?

   _____

2. Where are the owls?

   _____

3. Where is the cat?

   _____

4. Where are the sandwiches?

   _____

5. Where are the shoes?

   _____

6. Where is the dog?

   _____

7. Where is the sun?

   _____

8. Where are the bottles?

   _____

9. Where is the umbrella?

   _____

10. Where are the birds?

    _____

Appendix 12. Participants' Scores on All Tests and Tasks

Name	*Composite Score of knowledge of English*	*Vocabulary task*	*Listening comprehension*	*Reading comprehension*	*Grammar task 1*	*Grammar task 2*	*Grammar task 3*	*Sentence formation 1*	*Sentence formation 2*	*Finnish nonword repetition test 1*	*Finnish nonword repetition test 2*	*English nonword repetition test*
Sanna	194.50	52	10	10	9.00	9.50	8.50	37.50	58.00	13	10	7
Helen	179.00	45	8	9	10.00	9.00	10.00	30.50	57.50	12	12	9
Rauli	176.00	44	9	9.50	8.00	6.00	9.00	34.50	56.00	10	16	9
Valtteri	164.00	42	9	9.00	8.50	7.00	8.00	31.50	49.00	11	9	7
Aku	152.50	44	8	8	5.50	7.00	8.00	27.00	45.00	11	9	8
Jonne	142.25	35	8	6.75	8.00	7.00	10.00	27.00	40.50	12	16	5
Maija	134.75	23	6	7.75	7.00	6.00	9.50	24.00	51.50	12	14	6
Eeva	134.50	29	7	6	6.00	5.50	7.00	18.50	55.50	13	11	6
Emma	126.00	33	8	9	7.00	6.50	9.00	20.50	33.00	9	9	2
Jari	116.50	31	8	4.50	7.00	4.50	4.50	16.00	41.00	9	9	3
Matti	110.50	22	5	5.50	7.00	7.00	5.00	13.50	45.50	8	8	6
Annika	89.50	22	5	4	7.00	3.50	6.50	17.50	24.00	12	12	5
Sakari	86.00	26	5	.50	5.00	1.00	6.00	5.50	37.00	9	11	4
Maria	76.75	18	3	3.75	3.50	4.50	5.00	21.00	18.00	9	10	3
Seppo	66.75	22	2	1.25	4.00	1.00	5.50	5.50	25.50	11	9	5
*Veera**	-	28	4	7.00	7.50	4.00	7.00	26.00	-	11	11	7

**The participant does not have a score on Sentence formation 2(preposition task), and therefore there is no composite score either.*