JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Kanerva, Kaisa; Kiistala, Ilkka; Kalakoski, Virpi; Hirvonen, Riikka; Ahonen, Timo; Kiuru, Noona

Title: The feasibility of working memory tablet tasks in predicting scholastic skills in classroom settings

Year: 2019

Version: Accepted version (Final draft)

Copyright: © 2019 John Wiley & Sons, Ltd

Rights: In Copyright

Rights url: http://rightsstatements.org/page/InC/1.0/?language=en

Please cite the original version:

Kanerva, K., Kiistala, I., Kalakoski, V., Hirvonen, R., Ahonen, T., & Kiuru, N. (2019). The feasibility of working memory tablet tasks in predicting scholastic skills in classroom settings. Applied Cognitive Psychology, 33(6), 1224-1237. https://doi.org/10.1002/acp.3569

Kanerva Kaisa (Orcid ID: 0000-0001-5686-2396)

The feasibility of working memory tablet tasks in predicting scholastic skills in classroom settings

Kanerva, Kaisa[1,2]; Kiistala, Ilkka[2]; Kalakoski, Virpi[3]; Hirvonen, Riikka[2]; Ahonen, Timo[2] & Kiuru, Noona[2]

[1] Department of Medicine, University of Helsinki, Finland

[2] Department of Psychology, University of Jyväskylä, Finland

[3] Finnish Institute of Occupational Health, Finland

Author Note:

Correspondence concerning this article should be addressed to Kaisa Kanerva

University of Helsinki, Siltavuorenpenger 1 A, P.O. Box 9, FI-00014 University of Helsinki

Tel.: +(358) 41-536-4747

E-mail: kaisa.kanerva@helsinki.fi

Orcid ID: https://orcid.org/0000-0001-5686-2396

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgements:

This study forms part of the STAIRWAY-From Primary School to Secondary School Study (Ahonen & Kiuru, 2013-2017). The study was funded by grants from the Academy of Finland(#266851, 294970).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/acp.3569

This article is protected by copyright. All rights reserved.

Accepted Article

Abstract

Cognitive assessment in natural group settings facilitates data collection, but poses threats to the validity. In this study, tablet-based WM tasks, the Counting Span and Reading Span were used in predicting 12-year-old children's (*N*=837) scholastic skills and fluid intelligence in a classroom with environmental noise. WM tasks had excellent internal consistency, correlated with scholastic skills and accounted for more of the variance in cognitive performance (GPA, fluid intelligence, scholastic skills) compared to individually administered (*N*=190) Digit Span task. Furthermore, the multilevel analysis revealed that compared to the classrooms with no noise, when naturally occurring speech or non-speech types of environmental noises were present during assessment, WM scores or the reliability estimates were not lower. In contrast, when both types of noises were present, the relationships between some of the WM and achievement scores were even stronger. Thus, assessments in natural classroom contexts may promote revealing the individual differences in WM.

Introduction

Working memory (WM) refers to the limited memory capacity to store and manipulate information for brief periods of time in an ongoing task (Baddeley, 1986). Individuals vary in their WM capacity (Klein, & Fiss, 1999), and these capacity differences are related to many cognitive skills (Engle, 2001). The WM capacity predicts such basic academic abilities as mathematical skills and reading comprehension that require controlled attention (Engle, & Kane, 2004). Since WM is an essential cognitive capacity underlying success in academic learning, we need reliable, valid and practical methods for the use of educational and clinical psychological assessment and for research purposes (Conway et al., 2005).

Complex span tasks have proved to be successful in assessing WM capacity; performance levels in these tasks are stable and have high correlations with other cognitive skills (Conway et al., 2005; Hitch, Towse, & Hutton, 2001; Redick, Heitz, & Engle, 2007; Tillman, 2011). Traditionally, WM assessments have been conducted either with a paper-and-

Accepted Article

pencil test or using computer-based tests in a laboratory setting. However, for most children and adolescents digital devices are a natural interface used in their leisure time and they also tend to find the devices easy and motivating to use. Administering the WM span tasks in a mobile tablet environment also features another practical utility: it allows testing the participants in a group setting and saves resources when collecting multiple assessments at once.

Although mobile digital devices suggest advantages for data collection in many fields related to psychology, the eligibility of the mobile assessment tasks involving new presentation and response methods has to be carefully evaluated (Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016; Moore, Swendsen, & Depp, 2017). In this study, we investigate the utility of two widely used WM span tasks adapted to a touchscreen tablet interface, focusing on their feasibility in group test situations in school settings.

Working memory assessment

The *complex span* tasks (*working memory span* tasks) are a set of different memory tasks in which the participant stores items in his/her memory (memory component) while conducting a processing task in between memory items (processing component). Various different memory and processing components have been adopted in studies since Daneman and Carpenter (1980) introduced their first Reading Span task. In the original Reading Span task, reading sentences was used as a processing component and retaining the last words of the sentences as a memory component. Other widely used processing components have been tasks of counting dots (Counting Span; Case, Kurland, & Goldberg, 1982) and verifying arithmetic operations (Operation Span; Turner & Engle, 1989). Commonly used memory components, in turn, have involved unrelated words (Daneman & Carpenter, 1980), digits (Turner & Engle, 1989), or letters (Barrouillet, Bernandin, & Camos, 2004). In children, two commonly adopted complex span tasks are the Counting Span task and the Reading Span

This article is protected by copyright. All rights reserved.

task, which were both also used in the present study.

It has been suggested that due to their requirement of concurrent processing and storage, these tasks tap some domain-general, executive-attentional functions of working memory (Engle, Tuholski, Laughlin, & Conway, 1999), and that this explains why these tasks predict higher-order cognition so well. Different processing components have been shown to systematically predict higher-order cognition to a fairly similar degree. However, there are other aspects that affect the tasks' predictive use. Earlier research suggests that complex span tasks predict higher-order cognitive performance better when the processing component is attention-demanding in nature (Kanerva & Kalakoski, 2016; St. Clair-Thompson, 2007) and when the presentation is computer-paced rather than self-paced (Lépine, Barrouillet, & Camos, 2005). For example, by varying the cognitive cost of the processing component of the Counting Span task, St. Clair-Thompson (2007) demonstrated that the task with most demanding processing component was best predictor for scholastic skills (literacy and numeracy) in children and adults. Using tablets in WM assessments allows for controlling for the pace of the processing and memory components, which is essential for reliable assessment.

Assessing cognitive performance with mobile digital devices

Tablets and smartphones are increasingly being adopted for data collection in various research areas in psychology (Brouillette et al., 2013; Timmers et al., 2014), including in developmental studies (Frank et al., 2016; Semmelmann et al., 2016). Comparability studies between touchscreen and paper-and-pencil methods for data collection in cognitive assessments have demonstrated that children prefer the touchscreen tablet mode over the paper-and-pencil mode (Piatt, Coret, Choi, Volden, & Pisanz, 2016), and using a touchscreen is an overall reliable method for collecting behavioral data (Frank et al., 2016). For example, Piatt and colleagues (2016) showed that the accuracy of the number-line estimation did not

Accepted Article

differ between the tablet-based and paper-and pencil versions of the task when tested with 9- to 13-year-old children. Dufau and colleagues (2011) demonstrated that the latencies in a lexical decision task were highly correlated and similarly distributed whether touchscreen or traditional methods were used.

Earlier research indicates that many aspects of cognitive performance in children can be assessed with mobile devices. Typically, WM is assessed individually in laboratory settings, but mobile devices allow for easily portable assessment in groups of participants, such as in school classrooms. Despite of the apparent advantage of mobile digital devices in collecting cognitive data, the field is new and research is still sparse. It is not clear whether WM can be reliably assessed with mobile devices in a classroom. In this study, we address the question of whether or not the concept being assessed, that is, working memory, can indeed be reliably assessed with modern touchscreen devices in a classroom.

Assessing WM in group situations

In some recent studies group-based assessments of WM have been conducted in classrooms, typically by computers (e.g., De Neys, d'Ydewalle, Schaeken, & Vos, 2002; Van de Weijer-Bergsma, Kroesbergen, Prast, & Van Luit, 2015; St. Claire-Thompson, 2014). Those studies have demonstrated that WM capacity can be reliably assessed in a classroom environment. Studies also indicate, that there is some variation in the strength of the relationship between WM and school skills in comparing school classes (Van de Weijer-Bergsma et al., 2015; Van de Weijer-Bergsma, Kroesenbergen, Jolani, & Van Luit, 2016). However, little is known about the nature of classroom differences and the factors that may contribute to the variation in relationships between WM and cognitive achievement in classrooms when assessed in groups of students.

This article is protected by copyright. All rights reserved.

In a group situation, two factors could potentially limit the validity of the assessment of WM capacity. The first is the self-reliance of the tasks, which may lower their validity: when the experimenter does not guide the participants through the cognitively demanding WM tasks, some participants may have difficulties in completing the tasks regardless of their WM capacity. This would result in low performance levels in such WM span tasks and lower the associations between WM and individually administered tasks conducted in a laboratory. The second potential threat for the validity of group assessments, especially in a classroom context, is the environmental noise or other environmental distractions during assessments (Elliot, 2002). Distractions of the environment may affect the performance or a task's ability to tap the underlying cognitive construct (Redick et al., 2012), and they may also contribute to the relationship between WM and scholastic skills.

Previous research has demonstrated that even brief exposure to irrelevant sound can be detrimental to the cognitive performance of children and adults (Sörqvist, 2010). Irrelevant auditory distractions, verbal or non-verbal affecting WM are referred to as irrelevant sound effects (Salamé & Baddeley, 1982). Any extraneous stimulation may constitute a potential source of distraction (Wetzel & Schröger, 2007) and the effect is present even when the memory task is presented visually. Furthermore, children are more susceptible to the disruptive effect of irrelevant sound than adults (Elliott, 2002; Elliott & Briganti, 2012; Elliot et al., 2016). Both verbal and non-verbal auditory distraction has been documented to disrupt memory performance (Tremblay, Nicholls, Alford, & Jones, 2000), also in classroom settings (Hygge, 2003), but some research indicates a more detrimental effect for speech-based distraction compared to non-speech-based distraction (Salame & Baddeley, 1982). In the context of the feasibility of the mobile, group assessment of WM in school, the important question is whether the environmental noise affects the reliability of the WM assessment and/or the relationship between WM and other cognitive tasks, such as scholastic skills and

Accepted Article

fluid intelligence.

The present study and aims

In our study, we present a touchscreen method for assessing working memory (WM) in group settings with two widely used, reliable WM span tasks, that is, the Counting Span (Case et al., 1982) and Reading Span (Daneman & Carpenter, 1980) tasks. We studied the feasibility of the touchscreen version in assessing 12-year-old children's working memory. We first evaluated the internal consistency of the touchscreen WM span tasks and evaluated their potential in predicting cognitive achievement scores (fluid intelligence, arithmetic skills, reading skills and grade point average). Second, we evaluated the correlations between performance in WM tasks conducted in a classroom setting and WM tasks conducted individually in a laboratory. Digit Span (Wechsler, 2010) was chosen as the individual assessment measure, because it is considered a well documented, valid indicator of storage and processing functions of working memory. In this task an increasing number of digits are presented, and the participant is required to recall the sets first in their presentation order and then in the reverse order. The combined score is assumed to reflect storing and processing functions of WM, which are also critical in the complex span tasks. We investigated whether the WM tablet tasks administered in the classroom predict fluid intelligence and scholastic skills similarly compared to Digit Span task conducted in the laboratory. Thirdly, to describe the naturalistic aspects of the testing situation, we documented the environmental noises in the classroom during WM testing, scored as speech or non-speech noise and took them into account when modeling the relationship between the children's working memory and cognitive achievement.

Accepted Article

This article is protected by copyright. All rights reserved.

Methods

Participants

The participants in the present study were part of a larger longitudinal study focusing on individual and environment-related factors that promote students' learning and school well-being during the transition from elementary school to lower secondary school (Authors, 2013-2017). The sample of the present study consisted of 837 early adolescents (461 girls, 376 boys). The mean age of the participants was 12 years (M = 12.3, SD = 0.4) and the ages of the participants ranged from 10 to 14 years. A total of 96% of the participants had Finnish as their mother tongue, 2% were bilingual (Finnish and some other language), and 2% of the adolescents either spoke a language other than Finnish as their mother tongue or did not report what their mother tongue language was.

The schools can be viewed as a good representation of the schools in Finland since they were located in both rural and urban areas. The majority of the families, 75%, were nuclear families, 13% were single-parent families, 12% were blended families, and 0.7% were other types of families. Compared to Finnish families with underage children (Official Statistics of Finland, 2016b), single-parent households were underrepresented and two-parent households overrepresented in our sample. A total of 4% of the participating parents were not educated beyond compulsory schooling (i.e., nine years of basic education), 29% had completed upper secondary education, 40% had a Bachelor's degree or vocational college degree, and 27% had a Master's degree or higher. Compared to the same-age Finnish population (Official Statistics of Finland, 2016a), the parents of the present sample were slightly more educated: parents with a college degree or higher were overrepresented, whereas parents with an upper secondary school diploma or lower were underrepresented.

There were 63 testing sessions in total in which children were assessed with WM span tasks and the environmental noise was reported. A total of 22 children were absent from

Accepted Article

school at the time of the group testing sessions and subsequently seven additional sessions were arranged for these participants to make up the missed testing sessions; however, since the conditions were dissimilar from the original group testing sessions they had missed, featuring a considerably smaller number of children (1 to 5) in the class, these participants were ultimately removed from the analyses. The mean number of children in the remaining 56 group testing sessions was 19, while the specific number of children in each testing session varied from 6 to 28.

A subsample (N = 190) of the participants were also tested individually. Eight of these participants were absent from school at the time of the group testing sessions and where thus dropped from the analyses. Because one aim of the broader longitudinal study was to investigate adolescents' learning difficulties, the subsample consisted of students from the following three groups: 1) students (n = 62) with difficulties in arithmetic fluency (i.e., students scoring in the lowest 16th percentile (~ -1 SD) in a test of arithmetic fluency, 2) students (n = 62) with difficulties in reading fluency (i.e., students scoring in the lowest 16th percentile (~ -1 SD) in a test of reading fluency, and 3) students (n = 66) without difficulties in arithmetic fluency or reading fluency who were matched in regards to IQ and gender with the students in the other two groups. In the present study, the focus was on the Digit Span task that was individually administered for all 190 students of the subsample.

Parents' written consent and student assent was required for children's participation. The parents were advised to discuss the study with their offspring to ensure the children's own willingness to participate. Teachers of the participating classrooms gave their written consent for the data collections to be conducted during the lessons. The research plan of the longitudinal project has been evaluated by the local Ethics Committee. This committee has given an *ethical statement* verifying that the study does not pose any ethical concerns.

Accepted Article

This article is protected by copyright. All rights reserved.

Procedure

The group assessments were conducted in the classrooms during regular school hours. Two trained research assistants who were present for all of the assessment sessions administered the tests and questionnaires to the students. The data collections took place in Grade 6 (Fall and Spring semesters). Fluid intelligence, arithmetic skills and reading skills were assessed in the Fall semester and the WM tasks were assessed in the Spring semester. In addition, the Digit Span task was individually administered to the subsample of students (n = 190) in the Spring semester. This was conducted in an ambulatory laboratory built inside a caravan that was moved from one school to another and parked in the schoolyard for the time of the individual assessments.

The WM tablet tasks were conducted successively starting with the Counting Span task and immediately followed by the Reading Span task. After the completion of the WM span tasks, the participants were instructed to independently move on to completing a questionnaire they were given. Administration of the WM span tasks lasted approximately 20 minutes. The participants were carefully instructed to stay quiet during the WM testing and to ask for help by raising their hand if they were to have any problems with the tasks.

Materials

Working memory span tasks (Grade 6, Spring semester). Two WM span tasks, Counting Span and Reading Span, were conducted using the touchscreen interface of an Android tablet (Samsung Galaxy Tab3, GT-P5210; Display: 10.1 inches; resolution: 1280 x 800; orientation: landscape) with OpenSesame Runtime for Android (version 2.8.3) (Mathôt, Schreij, & Theeuwes, 2012). The WM span tasks used in this study are available at *https://github.com/kiistala/working-memory-span-tasks-for-tablet*.

The Counting Span task was modified from the version originally introduced by Case and colleagues (1982). The task consists of counting yellow dots (against a black

Accepted Article

background) in number sets of 3, 4, 5, 6, 7 or 8 (processing component) presented on a black tablet screen and storing the number of dots in each set in one's memory (storing component). The task started with the presentation of two sets of dots, and after three trials the amount of sets increased by one. The dots were presented in a randomly determined pattern on the screen. The sets of the amounts of dots were randomly selected from five possible sets of each set size. There were no regularities in the sets, such as consecutive or alternating numbers. After the presentation of all sets of dots, the participants entered the recalled number of digits on the tablet screen by touching the numbers on the screen in the correct serial order. The selected numbers would appear on the screen for confirmation and by touching the "OK" button the next trial was presented.

The Reading Span task was modified from the original version by Daneman and Carpenter (1980). The task consisted of reading sentences, judging whether they are true or false (processing component), and storing the unrelated letters in one's memory (storing component). On the tablet screen, a sentence first appeared followed by two boxes with TRUE and FALSE -texts. After the participant responded, an unrelated letter appeared. The task started with the presentation of two sentences (and two letters to remember), and after three trials the amount of sentences increased by one. The sentences and letters were presented on the tablet's screen and the participant responded to each set of letters by pressing the letters on the touchscreen in the correct serial order. The sentences were short *true* or *false* statements, such as "A rabbit is green." Half of the statements were true and the other half false. The letters used were: H, J, K, L, M, P, R, S, and V, which are common consonants in Finnish language.

In both tasks, the items were presented only visually, that is, on the tablet's screen. The participants gave their responses by touching the relevant items on the screen, and the responses were recorded in a log file. The participants had the possibility to clear a current

Accepted Article

This article is protected by copyright. All rights reserved.

response by pressing a “C” button, and the next item would appear after touching the “OK” button on the screen.

All instructions regarding the WM span task were presented on the tablet’s screen. In both tasks, the testing started with two practice trials. The Counting Span task did not start until the participant responded correctly to the practice trials. After that, there were three trials consisting of two sets of processing and storage tasks. The amount of sets increased if a participant recalled at least one of the three sets correctly and this continued until the participant would not recall any of the three sets correctly, at which point the task would be discontinued.

Fluid intelligence (Grade 6, Fall semester). The Raven Standard Progressive Matrices (Raven, Raven, & Court, 1998) were used in assessing non-verbal intelligence. Raven’s test consists of diagrams with one part missing. Participants are asked to select the correct part that would complete each design, and the test increases in difficulty. In our study, only half of the items were used and alternating items were selected to be presented. All participants completed all selected items. Responses were scored as correct or incorrect and the maximum score was 30.

Arithmetic skills. A basic arithmetic test (Aunola & Räsänen, 2007; Räsänen, Salminen, Wilson, Aunio & Dehaene, 2009) was used to assess fluency in arithmetic skills. In this speeded group-administered test, each participant was required to complete as many arithmetic operations as possible within a three-minute time limit. The sixth-grade form of the test consists of 10 additions and 11 subtractions. Seven tasks include both additions and subtractions or multiplication and division problems to be solved. The score was the total number of correct answers, the maximum possible being 28.

Accepted Article

Reading skills. Three tasks were used in assessing reading skills. *Reading fluency* was assessed with a word chain task (Holopainen, Kairaluoma, Nevala, Ahonen, & Aro, 2004) in which the participant had to separate words with a vertical line. One hundred words were written in four-word clusters without any spaces in between the words. The time limit for the completion of each task was 90 seconds and the score was the number of correctly separated words. *Spelling* was assessed with a task in which the participant had to find spelling errors among one hundred written words (Holopainen et al., 2004). The participants' task was to mark (with a vertical line) as many spelling errors as possible within 3.5 minutes. The possible types of spelling errors were: a missing letter, a wrong letter, or an extra letter. The score was the number of correctly identified errors. *Reading comprehension* was assessed with a modified Finnish version of the Salzburg Reading Test (Landerl, Wimmer, & Moser, 1997; translated into Finnish by Sini Huemer). It consists of 36 sentences that are each either *true* or *false*. The participants were instructed to read each sentence and mark it as either *true* or *false*. The time limit for each task was 90 seconds. The score was the number of correctly judged sentences.

School achievement (grade point average, GPA). School achievement was seen as represented by each participants' self-reported grade point average of the previous Spring semester (Grade 5). The participants were asked to report their grade point average from the previous Spring semester. Children's self-reported school grades have been shown to correlate .86 with their actual grades from the school registers (Authors, 2013-2017).

Accepted Article

This article is protected by copyright. All rights reserved.

Digit Span task. The Digit Span task is a subtest of the WISC-IV (Wechsler, 2010). In this task, children were asked to repeat auditorily presented digit strings in their correct order, first forward and then backward. Standardized WISC-IV administration of the Digit Span task according to the test manual was followed and the standard scores of the Digit Span task were obtained.

Environmental noise in the classroom during WM assessment. During the WM testing sessions, the research assistants were instructed to document all environmental distractions, noise and other sudden sounds that they observed. There were two research assistants in every session. The observing and documenting of the environmental noise were done over the total session in which both tasks were administered. The research assistants were instructed to describe what caused the sound or noise. Afterwards, the verbal descriptions were coded by another researcher according to four categories: “no noise,” “speech noise,” “non-speech noise,” and when applicable, “both speech and non-speech noise.” The noise was scored as speech noise when verbal auditory noise was reported, for example the teacher (in nine sessions) or students (in 11 sessions) asked something or the central radio announcement started (in two sessions). The noise was scored as non-speech noise when non-verbal auditory noise was reported, for example the school bell rang (in five sessions) or a piano was being played in the neighboring classroom (in two sessions). However, in some cases the nature of the noise was not that clear, such as unspecified sounds from the school's corridor (in five sessions) or from the neighboring classroom (in two sessions). In above-mentioned cases, the noise was scored as non-speech noise.

Accepted Article

Results

The WM span tasks were analyzed using *partial credit unit* scores (PCU) (Conway et al., 2005). In this scoring method, the mean number of correctly recalled memory items within a list length is calculated and these proportions are then averaged to obtain the score. This scoring method is recommended by Conway and colleagues (2005) based on solid internal consistency. All of these analyses were conducted in addition to using the total scores but since the PCU scores are widely used and recommended, and because the results were similar, we used the PCU scores when reporting the results. For all the analyses, the criterion for statistical significance was set a priori at alpha = 0.05.

The computer code used for data processing and statistical analysis is available online at *https://osf.io/w5qb6/*

The associations between the WM tablet tasks, scholastic skills and fluid intelligence

According to the descriptive statistics reported in Table 1, the WM performance mean scores were reasonable and suggested typical average performance levels. However, it is not possible to directly compare these values with other studies, since other studies have used different languages, scoring, and/or presentation formats.

In order to obtain the reliability estimates for WM span tasks, we combined the first presentation of all the sets of different lengths into a single PCU score, the second presentation into a single PCU score, and the third presentation into a single PCU score. These three subscores for each WM task were used to compute Cronbach's α as a measure of reliability (Conway et al., 2005; Engle et al., 1999). The reliability estimates are reported in Table 1, and they demonstrated high internal consistency for both WM span tasks. This indicates that participants who responded with the correct answer for one set of span tasks tended to respond with the correct answer also on the other sets.

Accepted Article

This article is protected by copyright. All rights reserved.

The correlations between WM span tasks, scholastic skills and fluid intelligence are reported in Table 2.

The associations between the WM tablet tasks, individually administered WM task and cognitive achievement

The Pearson's correlations between performance in the WM tablet tasks administered in a classroom and the individually administered Digit span task were calculated for the subsample of students (N = 182). The correlations were statistically significant: between the Counting Span task and Digit Span task's standard score $r(180) = .44$, $p < .001$, and between the Reading Span task and Digit Span task's standard score $r(180) = .33$, $p < .001$.

To determine if the individually administered WM task accounted for any additional variance in the cognitive achievement of the students beyond the effect of the tablet WM tasks administered in a classroom, we conducted hierarchical regression analyses. These analyses were conducted for each of the three measures of cognitive achievement, which were not used as a criterion in selecting participants in the subsample, namely Raven's Test, reading comprehension and GPA. We entered the classroom and laboratory WM tasks as predictors in separate models (Models 1 and 2) and studied the proportion of explained variance in each cognitive achievement measure (univariate regression analyses) (see Tables 3 to 5). In Models 3, we entered both the WM tablet task (Counting Span or Reading Span) and the laboratory Digit Span tasks to evaluate whether the WM tablet task predicted unique variance in students' cognitive achievement over and above the effect of the

Accepted Article

laboratory Digit Span task.

The results in tables 3 to 5 show that when entered separately, all WM tasks (the group administered WM tablet tasks and individually administered Digit Span task) significantly predicted amounts of variance in the three cognitive achievement measures (Models 1 and 2). However, when the WM tablet task, either Counting Span or Reading Span, was entered into the model together with the Digit Span task (Model 3), it explained unique variance beyond the Digit Span task consistently in Raven's Test, reading comprehension, and GPA. Moreover, in this model the laboratory Digit Span task remained as a significant predictor only for GPA and when explaining reading comprehension additionally with the Counting Span task, but not for Raven's Test or when explaining reading comprehension additionally with the Reading Span task. Furthermore, when comparing Models 2 and 3, the models which included the WM tablet tasks were in all cases statistically significantly superior in explaining the cognitive measures compared to models with only Digit Span task, as shown in the change of R^2 (ΔR^2).

WM performance in the classrooms with environmental noise

The types of environmental noises that the examiners reported in the classrooms during the WM span assessment sessions were classified as: no-noise (28 sessions, 397 participants, mean of the number of participants in a session = 19.54; sd =4.51; median = 20), speech noise (12 sessions, 174 participants, mean of the number of participants in a session = 22.89; sd =4.30; median 24), non-speech noise (14 sessions, 180 participants, mean of the number of participants in a session = 18.68; sd = 3.81; median = 19), and simultaneous speech and non-speech noises (9 sessions, 86 participants, mean of the number of participants in a session = 22.23; sd = 4.76; median = 22.00). The mean scores, standard deviations and the reliability estimates for the WM performances in the span tasks, separately for each type of noise group,

This article is protected by copyright. All rights reserved.

are reported in Table 6. The analysis shows that the performance levels (PCU scores) did not differ by type of noise group. In addition, the reliability estimates were high and similar with each other in all noise groups.

Linear hierarchical models of the relationships between WM and cognitive achievement

In order to examine to what extent the strength of the association between WM span and cognitive achievement varied among the classes, and more importantly, to what extent environmental noise (as a classroom-level variable) moderated this variation, multilevel modeling with random slopes and random intercepts (e.g., Hox, 2002) was used.

Our data have a hierarchical structure in which participants are nested within school classes. Multilevel or hierarchical linear modeling (Luke, 2004) allows investigating whether or not there is an interaction in the relationship between performance in WM span tasks and cognitive achievement scores (Raven's Test, Arithmetic skills, three tests of Reading skills, and GPA) that might be explained by the Level 2 measure, namely, environmental noise. Thus, we focused our analyses primarily on the cross-level interactions of the relations between Level 1 performance scores (WM, cognitive achievement) and the Level 2 variable (noise). The cross-level interactions indicate that, within school classrooms, associations at Level 1 vary as a function of the Level 2 variable. For example, the relation between Reading Span and Reading Comprehension (both Level 1, within-person variables) might change as a function of environmental noise (a Level 2, between-persons variable), the relationship being different with more environmental noise compared to the same situation with less environmental noise.

The linear hierarchical models were carried out with the *lme4* statistical package (version 1.1.12; Bates, Mächler, Bolker, & Walker, 2014). The p-values were obtained by using lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2014; see also Luke, 2017).

Accepted Article

First, to examine to what extent classroom membership influenced the participants' performance in the criterion tasks, we analyzed the ratio of variance in performance between classes to variance in performance within classes. The intraclass correlation (ICC) was analyzed using a multilevel analysis with a two-level structure (Level 1, participant; Level 2, testing session). ICC values of .05, .10 and .15 are considered to be small, medium and large, respectively (Hox, 2002). In our study, the ICCs varied from small to medium (Counting Span: ICC = 0.01; Reading Span: ICC = 0.06; Raven's test: ICC = 0.11; Arithmetic skills: ICC = 0.05; Reading Fluency: ICC = 0.02; Spelling: ICC = 0.04; Reading Comprehension: ICC = 0.06; and GPA: ICC = 0.06).

We ran two sets of hierarchical linear analyses, one set for the Counting Span task and the other set for the Reading Span task. In each set, each of the criterion tasks were used as an outcome in separate models. We entered the WM Span task as a predictor for each criterion task. The predictors were grand mean centered. The strength of the associations between the WM Span task and criterion tasks were allowed to randomly vary between school classes (i.e., random slopes). Also, we explored whether the Level 2 variable noise level would predict random variation between school classes in regard to the strength of the relationship between working memory and cognitive achievement (i.e., cross-level interaction).

Our analyses confirmed that both types of WM span tasks predicted all of the achievement scores (Table 7). In addition, our analyses confirmed that the environmental noise did not significantly affect the intercept of the WM performance. Furthermore, although speech noise or non-speech noise alone did not have an interaction effect in moderating the relationship between the WM span tasks and achievement scores, in the group with both noises the interaction effect was present in the relationship between the Counting Span and Raven's test, Counting Span and Arithmetic skills, and Counting Span and two Reading skills tests, as well as between the Reading Span and Raven's test (see also Figures 1 and 2). These

Accepted Article

This article is protected by copyright. All rights reserved.

effects should be considered with caution, since when multiple testing was corrected for by Bonferroni correction, the only statistically significant ($p < .0083$ [.05 divided with the number of tests, 6]) interaction effect was found when predicting Raven's test with Counting Span task when both types of noise were present.

Discussion

In this study, our aim was to evaluate the feasibility of the touchscreen WM span task in assessing 12-year-old children's working memory capacity in a group setting. In particular, we were interested in whether the tablet tasks predict cognitive achievement, namely scholastic skills and fluid intelligence, and whether they thus have potential in assessment of WM in ecologically valid settings. We used two widely known complex span tasks, which were administered in a classroom setting using touchscreen tablets. Additionally, a small proportion of the sample was individually assessed in a laboratory with a widely used Digit Span task. The environmental noise in each classroom was reported by observing researchers and we were interested in the moderating effects of environmental noise on the relationship between WM span and scholastic skills in order to gain further understanding of the feasibility of such an assessment in a natural classroom setting.

The associations between mobile WM, individual WM assessment and cognitive achievement

The internal consistency, that is, tasks' ability to constantly assess the latent variable, was found excellent in two mobile WM span tasks., It was demonstrated by high correlations between estimates of three presentations of the same list length of the WM span tasks. The high internal consistency is crucial especially for tasks that are used for examining individual differences, because the reliability of the assessment methods limits the correlation that can be observed between the methods (Hedge, Powell, & Sumner, 2017; Nunnally, 1970).

The two WM span tasks in the present study were positively correlated with each

Accepted Article

other and with participants' performance in cognitive achievement measures similarly to many related earlier studies conducted in a laboratory (Alloway & Passolunghi, 2010; Lehto, 1995; Magimairaj & Montgomery, 2012) and group-settings (St. Clair-Thompson, 2014). The Counting Span task correlated slightly more highly with higher-level cognitive tasks such as fluid intelligence and grade point average compared to the Reading Span task. However, the Reading Span task correlated slightly more than the Counting Span task with tasks that are mostly literal, like spelling and reading comprehension. The differences were not large, but they might reflect the different requirements of the processing components of the tasks that involve either counting or reading (Siegel & Ryan, 1989).

The performance in both WM tablet tasks was also associated with performance in the well-known WM task, Digit span task, conducted in an individual laboratory setting. Moreover, when explaining fluid intelligence, both WM tablet tasks administered in a classroom accounted for unique variance beyond the individually administered WM task. These results suggest that WM tablet tasks administered in natural settings such as classrooms may be sufficient methods in capturing WM and that their possible challenges, such as self-reliance, do not impede the assessment when explaining cognitive achievement in this age group.

In sum, the correlations between the WM span, criterion tasks and Digit Span task were in line with previous studies, and the reliability estimates were high. These results are fundamental in providing the assurance that the touchscreen device versions of the tasks were valid in assessing WM capacity in early adolescent children.

The role of environmental noise

To gain further insight into the influence of the context, the classroom, on the WM assessment results, particularly where participants differed, we conducted further analyses using linear hierarchical models to assess how the naturally occurring environmental noise

This article is protected by copyright. All rights reserved.

contributed the association between the participants' WM span and their cognitive achievement. Occurrences of environmental noise were not accompanied with changes in WM performance: the participants in the classrooms that were exposed to speech and/or non-speech noise during the test sessions performed on average as well as the groups not exposed to environmental noise. However, environmental noise did appear to play some role in the relationship between the participants' WM performance and cognitive achievement scores, especially if there were both speech and other types of noises in and/or around the classroom during testing. The correlation between the participants' WM performance and cognitive achievement was higher in the groups exposed to noise than in the groups who were reportedly not exposed to environmental noise. In regard to the Counting Span task, this difference was clear; particularly when we examined the strength of its relation to Raven's test, arithmetic skills and reading skills. These results are in line with and extend those of a study by Sullivan and colleagues (2015), who found there to be a stronger relation between WM span and comprehension in school-aged children when these were exposed to noise compared to when not. In regard to the Reading Span task, the difference was not clear; the only significant interaction was detected in the relationship between the Reading Span task and Raven's test.

Because we did not control for the occurrences of noise, it is worth noting that these results might reflect the amount of noise rather than the type of noise, or there can be some other uncontrolled aspects that were related to those particular school classes. Thus, we restrict our conclusions regarding the type or amount of noise in concluding that the environmental noise, whether speech or non-speech type, was not accompanied with lower association between performance in the WM tablet task and cognitive achievement. Furthermore, since the WM span tasks' reliability estimates were similar in regardless of the environmental noise in both WM tasks, the tasks' ability to consistently catch the individuals'

Accepted Article

performance levels could be considered reliable even under environmental noise.

Our results, although correlational in nature, might reflect individual differences in attentional mechanisms of WM. Previous findings have shown that individuals with high WM are able to resist interference better than individuals with low WM (Sörqvist, Stenfelt, & Rönnberg, 2012). Although different theoretical views on WM recognize its effect on task performance, there is no consensus about the underlying mechanisms. It has been proposed that individuals with high WM are less susceptible to distraction because they can focus (or constrain) their attention better regarding attended targets (Heitz & Engle, 2007), or they have a superior inhibition capacity (Lustig, Hasher, & Zacks, 2007) or because they manage to maintain the goal-directed task set in their working memory even when challenged by stimuli that capture part of their attention (Unsworth & Engle, 2007). It is also possible that individuals with low WM have a more vulnerable rehearsal process, demonstrated by greater irrelevant speech effects when a task's requirements for rehearsal increase (Elliot et al., 2016). In future studies, the theoretical underpinnings of the effects of natural environmental noise could be clarified by addressing the participants with low WM span matched with control participants to find out whether the memory scores and the correlations between WM and academic skills would be different in classroom compared to laboratory settings in these two groups.

Limitations

All of the children completed the two WM span tasks at their own pace and we did not control the occurrences of the environmental noises during the course of the task. Thus, our results reveal the role of noise only at a general level. We cannot define the exact temporal locations of the noises in relation to the tasks. Experimental paradigms are needed if the aim is to show which cognitive phases and processes are disrupted by which noise at what point in time, specifically in regard to the encoding, maintaining and rehearsal of information

This article is protected by copyright. All rights reserved.

(Elliot, 2002). However, in our present study we were interested in the WM performance in an ecologically valid setting with naturally occurring environmental distractions.

In the end, we did not assess the effects of environmental noise on the cognitive achievement testing situations, but only on the WM testing situations. It is possible that some school classes are overall noisier than others and that the participants were exposed to more noise also when completing paper-and-pencil tests on fluid intelligence, arithmetic, and reading. However, we argue that when completing a paper-and-pencil test such as simple arithmetic, it is possible to allocate attention more flexibly when noise occurs (e.g., have a break when the fire engine passes the school). For that reason, the participants would be more resistant to interference during paper-and-pencil tests and the environmental noise would not have affected their performance to that extent (Barrouillet & Camos, 2012).

Despite these limitations, the results obtained in the natural classroom setting reveal that WM span tasks can be sufficiently assessed with touchscreen devices in school environments in this age group. Also, this approach enables cost-effective and practical methods to be used for assessing individual WM capacity.

Conclusions

Our study shows that tablet-based WM span tasks are a feasible method for assessing children's WM capacity and that such mobile paperless testing in classrooms, rather than in a laboratory setting, does not impede the validity of the assessment. The average performance levels in the touchscreen WM span tasks assessed in our present study are comparable to those of previous studies in that the participants' performance in the touchscreen WM span tasks predicted their fluid intelligence and scholastic skills. Furthermore, the environmental noise in the classrooms did not diminish the WM span tasks' feasibility in predicting higher-order cognitive performance. To the contrary, our results suggest that using a classroom group setting when assessing WM ability may offer an ideal testing environment with strong

Accepted Article

predictive validity of WM task. This kind of naturalistic testing environment, the classroom, may in some cases even improve WM span tasks' effectiveness in predicting cognitive achievement. In the future, this claim should be verified experimentally in a controlled setting.

References

Alloway, T. P., & Passolunghi, M. C. (2010). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences, 21*, 133–137. doi:10.1016/j.lindif.2010.09.013

Aunola, K., & Räsänen, P. (2007). *The 3-minutes basic arithmetic test.* Unpublished test material. Jyväskylä, Finland.

Baddeley, A. (1986). *Working memory*. New York: Oxford University Press.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 153*, 83–100. doi:10.1037/0096-3445.133.1.83

Barrouillet, P., & Camos, V. (2012). As time goes by: Temporal constraints in working memory. *Current Directions in Psychological Science*, *21*, 413–419. doi:10.1177/0963721412459513

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48. doi:10.18637/jss.v067.i01

Brouillette, R. M., Foil, H., Fontenot, S., Correro, A., Allen, R., Martin, C. K., ... Keller, J. N. (2013). Feasibility, reliability, and validity of a smartphone based application for the assessment of cognitive function in the elderly. *PloS one*, 8, e65925. doi:10.1371/journal.pone.0065925

This article is protected by copyright. All rights reserved.

Case, R., Kurland, M. D., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33, 386–404. doi:10.1016/0022-0965(82)90054-6

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z, Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769-786. doi:10.3758/BF03196772

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466. doi:10.1016/S0022-5371(80)90312-6

De Neys, W., d Ydewalle, G., Schaeken, W., & Vos, G. (2002). A Dutch, computerized, and group administrable adaptation of the operation span test. *Psychologica Belgica*, *42*, 177–190.

Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario, F. X., ... Ktori, M. (2011). Smart phone, smart science: how the use of smartphones can revolutionize research in cognitive science. *PloS one*, *6*, e24974. doi:10.1371/journal.pone.0024974

Elliott, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition*, *30*, 478-487. doi:10.3758/BF03194948

Elliott, E. M., & Briganti, A. M. (2012). Investigating the role of attentional resources in the irrelevant speech effect. *Acta Psychologica*, *140*, 64–74. doi:10.1016/j.actpsy.2012.02.009

Elliott, E. M., Hughes, R. W., Briganti, A., Joseph, T. N., Marsh, J. E., & Macken, B. (2016). Distraction in verbal short-term memory: Insights from developmental differences. *Journal of Memory and Language*, *88*, 39–50. doi:10.1016/j.jml.2015.12.008

Engle, R. W. (2001). What is working-memory capacity? In H. L. Roediger III, & J. S. Nairne (Eds.), *The Nature of Remembering: Essays in Honor of Robert G. Crowder* (pp. 297-314). Washington, DC: American Psychological Association.

Accepted Article

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *The Psychology of Learning and Motivation*, *44*, 145-199. doi:10.1016/S0079-7421(03)44005-X

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. Journal of Cognition and Development, 17, 1–17. doi:10.1080/15248372.2015.1061528

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 1–21. doi:10.3758/s13428-017-0935-1

Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: individual differences in visual attention control. *Journal of Experimental Psychology: General*, *136*, 217–240. doi:10.1037/0096-3445.136.2.217

Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, *130*, 184–198. doi:10.1037/0096-3445.130.2.184

Holopainen, L., Kairaluoma, L., Nevala, J., Ahonen, T., & Aro, M. (2004). Lukivaikeuksien seulontamenetelmä nuorille ja aikuisille [Dyslexia screening test for youth and adults]. Jyväskylä, Finland: Niilo Mäki Instituutti

Hox, J. (2002). Multilevel analysis: techiques and applications. Mahwah, NJ: Lawrence Erlbaum Associates

Accepted Article

This article is protected by copyright. All rights reserved.

Hygge, S. (2003). Classroom experiments on the effects of different noise sources and sound levels on long- term recall and recognition in children. Applied cognitive psychology, 17(8), 895-914. doi: 10.1002/acp.926

Kanerva, K., & Kalakoski, V. (2016). The Predictive Utility of a Working Memory Span Task Depends on Processing Demand and the Cognitive Task. *Applied Cognitive Psychology, 30*, 681–690. doi:10.1002/acp.3243

Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments, & Computers, 31*, 429–432. doi:10.3758/BF03200722

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2014). LmerTest: Tests for random and fixed effects for linear mixed effect models. R package, version 2.0-3.

Landerl, K., Wimmer, H., & Moser, E. (1997). Salzburger Lese-und rechtschreibtest (SLRT). *Verfahren zur Differentialdiagnose von Störungen des Lesens und Schreibens für die, 1*.

Lehto, J. (1995). Working memory and school achievement in the Ninth Form. *Educational Psychology, 15*, 271–282. doi:10.1080/0144341950150304

Lépine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin & Review, 12*, 165–170. doi:10.3758/BF03196363

Luke, D. A. (2004). *Multilevel modeling* (Vol. 143). London: Sage.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*, 1494–1502. doi:10.3758/s13428-016-0809-y

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a “new view”. In D. S. Gorfein, & C. M. MacLead, (Eds.): *Inhibition in Cognition* (pp. 145–162). American Psychological Association.

Accepted Article

Magimairaj, B. M., & Montgomery, J. W. (2012). Children's verbal working memory: Role of processing complexity in predicting spoken sentence comprehension. *Journal of Speech, Language, and Hearing Research, 55*, 669–682. doi:10.1044/1092-4388(2011/11-0111)

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324. doi:10.3758/s13428-011-0168-7

Moore, R. C, Swendsen, J., & Depp, C. A. (2017). Applications for self-administered mobile cognitive assessments in clinical research: A systematic review. International Journal of Methods in Psychiatric Research, e1562. doi:10.1002/mpr.1562

Nunnally, J. C. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.

Official Statistics of Finland (2016a). Educational structure of population (e-publication). Helsinki, Finland: Statistics Finland. Retrieved from www.stat.fi/til/vkour/2014/vkour_2014_2015_tie_001_en.html

Official Statistics of Finland (2016b). Families. Appendix Table 3: Families with underage children by type in 1950–2014 (e-publication). Helsinki, Finland: Statistics Finland. Retrieved from www.stat.fi/til/perh/2014/perh_2014_2015-05-28_tau_003_en.html

Piatt, C., Coret, M., Choi, M., Volden, J., & Bisanz, J. (2016). Comparing Children's Performance on and Preference for a Number-Line Estimation Task: Tablet Versus Paper and Pencil. *Journal of Psychoeducational Assessment*, *34*, 244–255. doi:10.1177/0734282915594746

Raven, J., Raven, J. E., & Court, J. H. (1998). *Standard progressive matrices*. Oxford, England: Oxford Psychologists Press

Redick, T. S., Heitz, R. P., & Engle, R. W. (2007). Working memory capacity and inhibition: Cognitive and social consequences. In D. S. Gorfein, & C. M. MacLead (Eds.), *Inhibition in*

Accepted Article

This article is protected by copyright. All rights reserved.

Cognition (pp. 125–142). American Psychological Association

Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. Cognitive Development, 24, 450–472. doi: 10.1016/j.cogdev.2009.09.003

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 150–164. doi:10.1016/S0022-5371(82)90521-7

Semmelmann, K., Nordt, M., Sommer, K., Röhnke, R., Mount, L., Prüfer, H., ... Weigelt, S. (2016). U can touch this: how tablets can be used to study cognitive development. *Frontiers in Psychology, 7:* 1021. doi:10.3389/fpsyg.2016.01021

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development, 60,* 973–980. doi:10.2307/1131037

St. Clair-Thompson, H. L. (2007). The effects of cognitive demand upon relationships between working memory and cognitive skills. *The Quarterly Journal of Experimental Psychology, 60,* 1378–1388. doi:10.1080/17470210601025505

St. Clair-Thompson, H. (2014). Establishing the reliability and validity of a computerized assessment of children's working memory for use in group settings. *Journal of Psychoeducational Assessment, 32*, 15–26. doi:10.1177/ 0734282913497344

Sullivan, J. R., Osman, H., & Schafer, E. C. (2015). The Effect of Noise on the Relationship Between Auditory Working Memory and Comprehension in School-Age Children. *Journal of Speech, Language, and Hearing Research, 58,* 1043–1051.

Sörqvist, P. (2010). The role of working memory capacity in auditory distraction: A review. *Noise & Health, 12,* 217–224. doi:10.4103/1463- 1741.70500

Accepted Article

Sörqvist, P., Stenfelt, S., & Rönnberg, J. (2012). Working memory capacity and visual–verbal cognitive load modulate auditory–sensory gating in the brainstem: Toward a unified view of attention. *Journal of Cognitive Neuroscience, 24*, 2147–2154. doi:10.1162/jocn_a_00275

Tillman, C. M. (2011). Developmental change in the relation between simple and complex spans: A meta-analysis. *Developmental Psychology, 47*, 1012–1025. doi:10.1037/a0021794

Timmers, C., Maeghs, A., Vestjens, M., Bonnemayer, C., Hamers, H., & Blokland, A. (2014). Ambulant cognitive assessment using a smartphone. *Applied Neuropsychology: Adult, 21*, 136–142. doi:10.1080/09084282.2013.778261

Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(6), 1750-1754. doi:http://dx.doi.org/10.1037/0278-7393.26.6.1750

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154. doi:10.1016/0749-596X(89)90040-5

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*, 104–132. doi:10.1037/0033-295X.114.1.104

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Prast, E. J., & Van Luit, J. E. (2015). Validity and reliability of an online visual–spatial working memory task for self-reliant administration in school-aged children. *Behavior Research Methods, 47*, 708–719. doi:10.3758/s13428-014-0469-8

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Jolani, S., & Van Luit, J. E. (2016). The Monkey game: A computerized verbal working memory task for self-reliant administration in primary school children. *Behavior Research Methods, 48*, 756–771. doi:10.3758/s13428-015-0607-y

Wechsler, D. (2010). WISC-IV. Wechsler Intellingence Scale for Children -IV. Helsinki:

Accepted Article

This article is protected by copyright. All rights reserved.

Hogrefe Psykologien Kustannus Oy.

Wetzel, N., & Schröger, E. (2007). Modulation of involuntary attention by the duration of novel and pitch deviant sounds in children and adolescents. *Biological Psychology*, *75*, 24–31. doi:10.1016/j.biopsycho.2006.10.006

Accepted Article

Table 1. *Descriptive Statistics for WM Span Tasks, Fluid Intelligence and Scholastic Performance*

	N	*Mean*	*SD*	*Min*	*Max*	*Skew*	*Kurtosis*	Cronbach's α
Counting Span	837	0.45	0.19	0.00	0.97	–0.08	–0.53	0.95
Reading Span	837	0.61	0.18	0.00	0.97	–0.56	0.43	0.93
Fluid Intelligence	823	22.67	3.60	5.00	29.00	–1.16	2.23	0.81
Arithmetic Skills	822	15.49	3.57	0.00	26.00	–0.47	0.88	0.82
Reading Fluency	823	39.78	14.04	1.00	99.00	0.39	0.16	0.88
Spelling	823	33.96	14.82	2.00	99.00	0.51	0.41	0.96
Reading Comprehension	823	16.75	4.81	5.00	36.00	0.62	0.90	0.90
GPA	708	8.33	0.66	6.00	10.00	–0.37	–0.02	

Accepted

This article is protected by copyright. All rights reserved.

Table 2. *The Correlations Between WM Span Tasks, Fluid Intelligence and Scholastic Performance*

	1	2	3	4	5	6	7
1. Counting Span	--						
2. Reading Span	0.44						
3. Fluid Intelligence	0.42	0.36					
4. Arithmetic Skills	0.45	0.34	0.40				
5. Reading Fluency	0.43	0.40	0.27	0.47			
6. Spelling	0.42	0.44	0.26	0.43	0.75		
7. Reading Comprehension	0.27	0.36	0.18	0.39	0.65	0.69	
8. GPA	0.42	0.36	0.38	0.35	0.43	0.44	0.40

Note. All correlations are statistically significant, $p < .001$.

Accepted Article

Table 3. *Regression Analyses (N=182) Explaining Fluid Intelligence (Raven's test) with the Classroom WM Tablet Tasks (Model 1), Laboratory Digit Span Tasks (Model 2), and Both (Model 3)*

Dependent: Raven	Counting Span				Reading Span			
	β(SE)	*t*	*R2*	*ΔR2*	*β(SE)*	*t*	*R2*	*ΔR2*
Model 1								
Classroom WM task	.34 (.07)	4.93***	0.12		.35 (.07)	5.11***	0.13	
Model 2								
Laboratory Digit Span	.23 (.07)	3.11**	0.05		.23 (.07)	3.11**	0.05	
Model 3				*ΔF*(179,1) =				*ΔF*(179,1) =
Laboratory Digit Span	.09 (.08)	1.20	0.13	15.18***	.12 (.07)	1.64	0.14	18.71***
Classroom WM task	.30 (.08)	3.90***			.32 (.07)	4.33***		

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Accepted Article

This article is protected by copyright. All rights reserved.

Table 4. Regression Analyses (N=182) Explaining Reading Comprehension with the Classroom WM Tablet Tasks (Model 1), Laboratory Digit Span Tasks (Model 2), and Both (Model 3)

Dependent: RC	Counting Span				Reading Span			
	β(SE)	*t*	*R2*	*ΔR2*	*β(SE)*	*t*	*R2*	*ΔR2*
Model 1								
Classroom WM task	.24 (.07)	3.38***	0.06		.34 (.07)	4.89***	0.12	
Model 2								
Laboratory Digit Span	.24 (.07)	3.30**	0.06		.24 (.07)	3.30**	0.06	
Model 3				*ΔF*(179,1) =				*ΔF*(179,1) =
Laboratory Digit Span	.17 (.08)	2.07*	0.08	4.51*	.14 (.07)	1.91	0.14	16.38***
Classroom WM task	.17 (.08)	2.12*			.30 (.07)	4.05***		

Note. RC = Reading Comprehension, * $p < .05$, ** $p < .01$, *** $p < .001$

Accepted Article

Table 5. *Regression Analyses (N=148) Explaining GPA with the Classroom WM Tablet Tasks (Model 1), Laboratory Digit Span Tasks (Model 2), and Both (Model 3)*

Dependent: GPA	Counting Span				Reading Span			
	β(SE)	*t*	*R2*	*ΔR2*	*β(SE)*	*t*	*R2*	*ΔR2*
Model 1								
Classroom WM task	.37 (.08)	4.87***	0.14		.39 (.08)	5.04***	0.15	
Model 2								
Laboratory Digit Span	.34 (.08)	4.34***	0.11		.34 (.08)	4.34***	0.11	
Model 3				*ΔF*(146,1) =				*ΔF*(146,1) =
Laboratory Digit Span	.23 (.08)	2.89**	0.19	12.64***	.25 (.08)	3.23**	0.20	16.39***
Classroom WM task	.29 (.08)	3.56***			.32 (.08)	4.05***		

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Accepted Article

This article is protected by copyright. All rights reserved.

Table 6. *The Performance Levels (Mean and SD) and the Reliability Estimates in Two WM Span Tasks in Relation to Each Environmental Noise Group*

		No-noise (N = 397)	Speech noise (N = 174)	Non-speech noise (N = 180)	Both types of noises (N = 86)	
Counting Span (PCU)	*Mean (SD)*	0.45 (0.19)	0.47 (0.19)	0.44 (0.19)	0.46 (0.20)	F (833, 3) = 0.92, p = .43 $\eta 2$ = .003
	Cronbach's α	0.95	0.95	0.95	0.96	
Reading Span (PCU)	*Mean (SD)*	0.61 (0.18)	0.63 (0.18)	0.61 (0.18)	0.60 (0.22)	F (833, 3) = 0.57, p = .64 $\eta 2$ = .002
	Cronbach's α	0.94	0.91	0.92	0.95	

Accepted Article

Table 7. *Fixed Effects Estimates (Standard Errors) for Models Predicting Fluid Intelligence and Scholastic Performance with the Counting and Reading Span Tasks*

Counting Span	Raven's Test	Arithmetic Skills	Reading Fluency	Spelling	Reading Comprehension	GPA
Intercept	22.73 (0.22) ***	15.58 (0.20) ***	40.22 (0.78) ***	34.07 (0.91) ***	17.20 (0.33) ***	8.38 (0.04) ***
Counting Span	1.39 (0.17) ***	1.61 (0.18) ***	6.10 (0.67) ***	6.18 (0.68) ***	1.10 (0.25) ***	0.27 (0.03) ***
Speech Noise	–0.37 (0.40)	–0.19 (0.37)	0.32 (1.42)	0.18 (1.66)	–0.48 (0.60)	–0.11 (0.08)
Non-Speech Noise	0.06 (0.38)	–0.12 (0.36)	–1.92 (1.83)	–0.47 (1.61)	–1.04 (0.58)	–0.09 (0.08)
Both Noises	–0.37 (0.51)	–0.84 (0.48)	–1.71 (1.85)	–1.02 (2.16)	–0.99 (0.78)	–0.13 (0.11)
Counting Span x Speech Noise	0.08 (0.31)	–0.21 (0.32)	–0.47 (1.20)	0.28 (1.22)	0.23 (0.45)	0.03 (0.06)
Counting Span x Non-Speech Noise	–0.32 (0.30)	–0.29 (0.31)	–0.74 (1.18)	–1.44 (1.20)	0.34 (0.43)	–0.03 (0.06)
Counting Span x Both Noises	1.41 (0.39) **	1.10 (0.41) *	1.87 (1.55)	3.45 (1.58) *	1.39 (0.57) *	0.03 (0.08)
Reading Span						
Intercept	22.69 (0.22) ***	15.52 (0.19) ***	40.12 (0.68) ***	33.87 (0.81) ***	17.18 (0.31) ***	8.39 (0.05)

This article is protected by copyright. All rights reserved.

						***
						0.23
	1.24 (0.19)	1.36 (0.24)	6.13 (0.88)	7.20 (0.88)		(0.03)
Reading Span	***	***	***	***	2.03 (0.30) ***	***
	–0.30					–0.09
Speech Noise	(0.39)	–0.07 (0.35)	0.37 (1.23)	0.31 (1.47)	–0.53 (0.56)	(0.09)
			–2.14	–0.55		–0.10
Non-Speech Noise	0.08 (0.38)	–0.14 (0.35)	(1.21)	(1.44)	–1.07 (0.75)	(0.09)
			–0.58			–0.09
Both Noises	0.08 (0.52)	–0.33 (0.47)	(1.63)	0.22 (1.96)	–0.34 (0.53)	(0.12)
Reading Span x	–0.09		–0.87	–0.92		0.01
Speech Noise	(0.35)	–0.52 (0.43)	(1.60)	(1.59)	–0.39 (0.52)	(0.06)
Reading Span x	–0.36		–1.39	–1.13		0.03
Non-Speech Noise	(0.34)	–0.31 (0.42)	(1.56)	(1.55)	–0.38 (0.52)	(0.06)
Reading Span x	0.94 (0.40)					0.01
Both Noises	*	0.39 (0.52)	1.12 (1.87)	0.17 (1.86)	–0.05 (0.63)	(0.07)

Note. * p < .05, ** p < .01, *** p < .001

Accepted

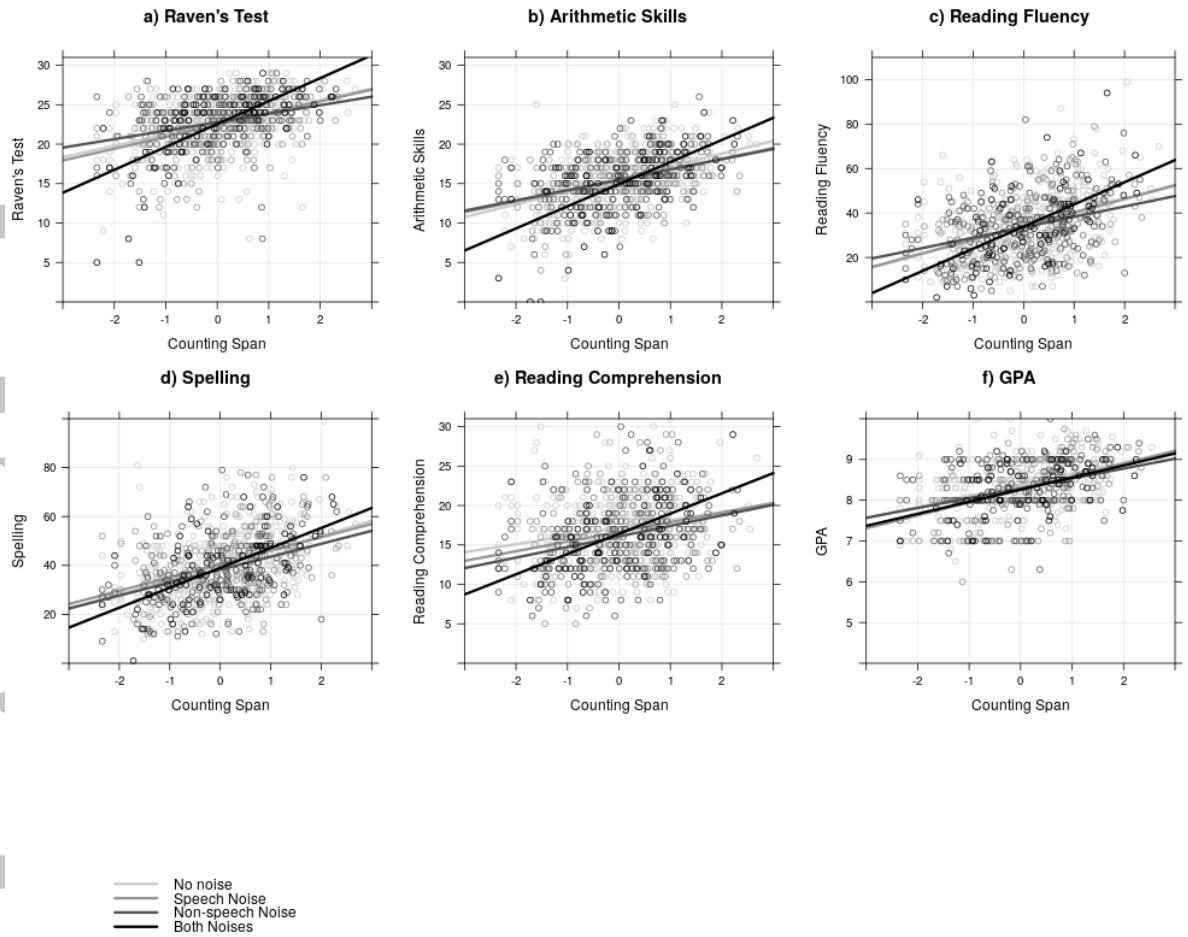

Fig. 1 Scatter plots describing the associations between the Counting Span task and cognitive achievement scores.

Accept

This article is protected by copyright. All rights reserved.

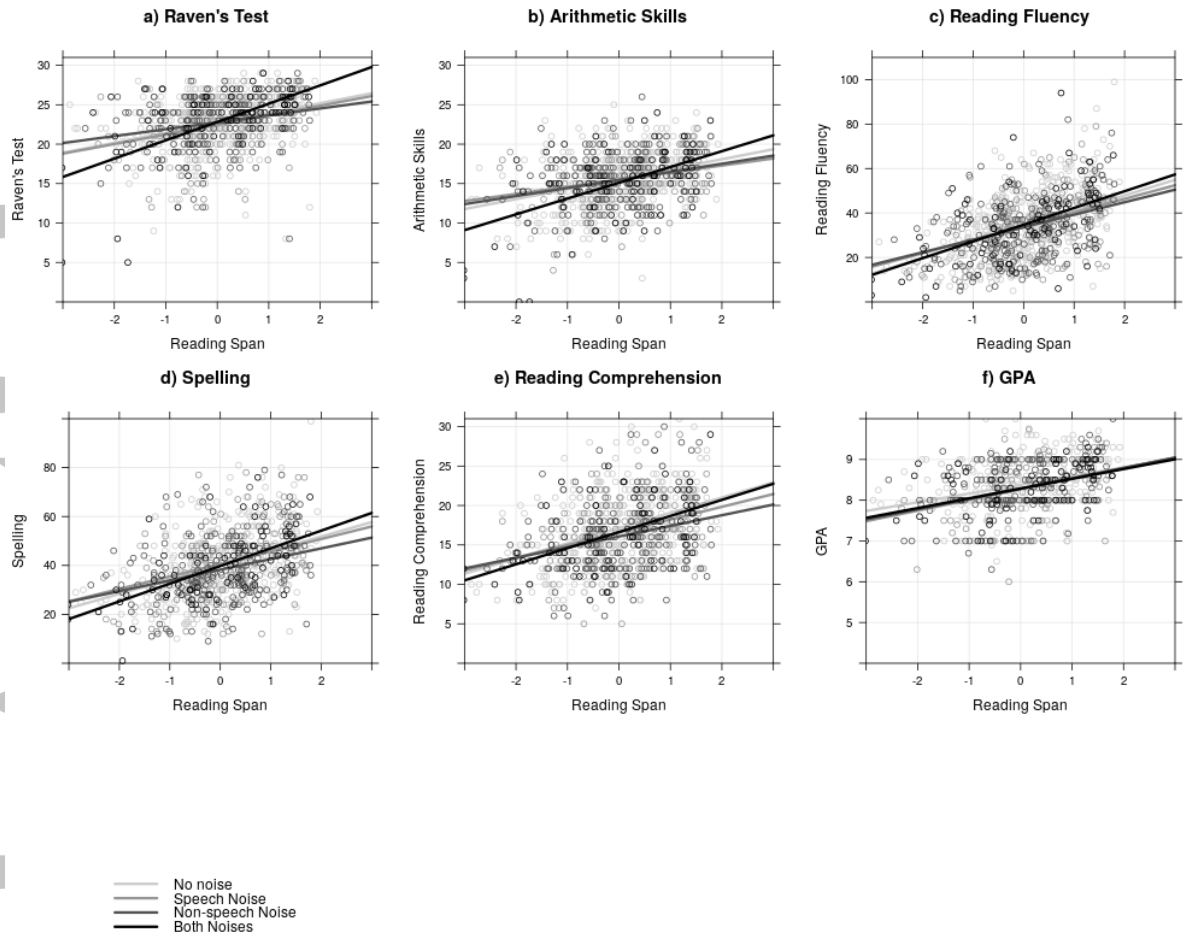

Fig. 2 Scatter plots describing the associations between the Reading Span task and cognitive achievement scores.

Accept