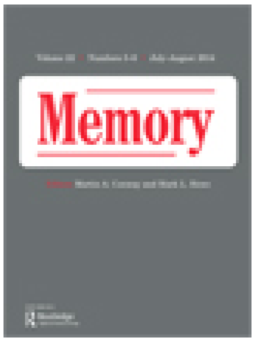

Memory

ISSN: 0965-8211 (Print) 1464-0686 (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Classroom versus individual working memory assessment: predicting academic achievement and the role of attention and response inhibition

Ilona Friso-van den Bos & Eva van de Weijer-Bergsma

To cite this article: Ilona Friso-van den Bos & Eva van de Weijer-Bergsma (2019): Classroom versus individual working memory assessment: predicting academic achievement and the role of attention and response inhibition, Memory, DOI: 10.1080/09658211.2019.1682170

To link to this article: https://doi.org/10.1080/09658211.2019.1682170

MEMORY
https://doi.org/10.1080/09658211.2019.1682170

Routledge
Taylor & Francis Group

Classroom versus individual working memory assessment: predicting academic achievement and the role of attention and response inhibition

Ilona Friso-van den Bos[a,b] and Eva van de Weijer-Bergsma[a]

[a]Department of Pedagogical and Educational Sciences, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, Netherlands; [b]Department of Educational Sciences, Faculty of Behavioural and Management Sciences, University of Twente

ABSTRACT
Working memory (WM) is an important predictor for academic learning and achievement. Typically, children's WM is assessed in controlled testing situations, which might not reflect functioning in typical classroom learning situations with natural distractions. In this study, we compared WM performance in controlled and classroom situations and their predictive value for academic achievement. Also, we examined whether performance differences between situations were moderated by attention or response inhibition. In a within-subjects design, primary school children completed visuospatial and verbal WM tasks in two settings (classroom versus controlled individual setting). First, WM functioning was lower in the classroom setting. Second, attention moderated individual differences in this discrepancy between settings, but response inhibition did not. Third, classroom obtained verbal WM scores were the strongest predictors of academic achievement. Our results indicate that classroom assessment of verbal WM provides a more ecologically valid measurement of WM abilities in a real-life learning situation.

ARTICLE HISTORY
Received 5 March 2019
Accepted 11 October 2019

KEYWORDS
Working memory; attention; response inhibition; academic achievement; classroom assessment

Working memory (WM) is the ability to temporarily store and process information that is needed to perform complex cognitive tasks. It is considered one of the most important predictors of academic skills, such as mathematics (e.g., Alloway & Alloway, 2010; Friso-van den Bos, Van der Ven, Kroesbergen, & Van Luit, 2013; Wu et al., 2017) and reading comprehension (e.g., Alloway & Alloway, 2010; Gerst, Cirino, Fletcher, & Yoshida, 2017; Pham & Hasson, 2014). However, WM is usually assessed in controlled testing situations in the presence of only a test leader, i.e.,: an environment with limited distractions. This does not reflect the typical learning situation in which WM is recruited, in which children participate in a classroom setting with multiple distracters and with peers and a teacher present. In this study, assessment of WM in a controlled testing situation is compared to that in a more ecologically valid classroom situation with all its distracters.

The original model of WM, proposed by Baddeley and Hitch (1974) consisted of three components. The two storage components, the visuospatial sketchpad and phonological loop, are proposed to be responsible for maintaining verbal and visuospatial information, respectively. A third component, the central executive, is thought to be responsible for regulating information in WM, by for example inhibiting the activation of irrelevant information, updating information to contain the most recent and relevant information, and shifting between response sets (also see: Miyake et al., 2000). Although a factor analysis has demonstrated that these three components adequately describe WM performance of children (Alloway, Gathercole, & Pickering, 2006), other studies have demonstrated that a distinction between visuospatial and verbal WM describe children's performance more adequately (Jarvis & Gathercole, 2003; Swanson, 2017). The latter distinction, between visuospatial and verbal WM, is employed in many studies targeting the role of WM in various domains of performance (e.g., Jarvis & Gathercole, 2003; Nath & Szücs, 2014; Pham & Hasson, 2014) and has been adopted in the current study.

There is ample evidence that WM forms a robust predictor of academic achievement in mathematics (for two meta-analyses, see Friso-van den Bos et al., 2013; Peng, Namkung, Barnes, & Sun, 2016) and reading comprehension (Borella & de Ribaupierre, 2014; Pham & Hasson, 2014). In order to solve a mathematics problem, children need to hold the numerical information in mind, manipulate this information, and keep track of intermediate solutions. Visuospatial WM may be recruited during calculation when strategies involve spatial representations of number such as a number line (Geary, Hoard, Nugent, & Byrd-Craven, 2008) and calculations may rely more heavily on verbal WM when strategies involve the production, updating, or

CONTACT Eva van de Weijer-Bergsma e.vandeweijer@uu.nl Department of Pedagogical and Educational Sciences, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, Netherlands

rehearsal of verbal information such as a counting procedure (Hecht, 2002). Likewise, in order to read and comprehend a written text, children need to hold recently read propositions in mind, integrate these with nearby text segments and retrieve information from prior knowledge to form a coherent understanding of the text (Van de Weijer-Bergsma, Kroesbergen, Jolani, & Van Luit, 2016). Verbal WM has been shown to be more strongly related to reading comprehension than visuospatial WM (Savage, Lavers, & Pillay, 2007; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000), although visuospatial WM may be involved in reading through visual imagery, as the processing of verbally conveyed spatial information has been found to interfere with the processing of visuospatial items (Baddeley, 2007).

However, while mathematics and reading activities typically take place in a classroom setting, evidence for the relation between WM performance on the one hand and mathematics and reading performance on the other hand almost exclusively relies on studies conducting WM tests in controlled testing situations. In such situations, distractions are kept to a minimum and the environment is stripped of the external stimuli that put a burden on children in everyday life. Children's task scores, therefore, may not fully reflect their ability to recruit WM resources in the typical learning environment that is the classroom, where distracters are lavishly available. In fact, the ecological validity of neuropsychological tests in general has been questioned as they often do not provide an accurate prediction of a person's functioning in real-world settings (Barkley, 1991; Burgess et al., 2006). The ecological validity of neuropsychological tests relies on two characteristics: verisimilitude and veridicality (Chaytor & Schmitter-Edgecombe, 2003). The concept of verisimilitude refers to the degree to which task-demands resemble the cognitive demands of real-life tasks. Many neuropsychological tests are administered in laboratory (-like) settings that bear little resemblance to real-life situations. The concept of veridicality refers to the degree to which performance on a neuropsychological test is *empirically* related to measures of everyday functioning, such as clinical observations and informant rating (e.g., by teachers and parents). Cognitive tasks may thus be more ecologically valid when administered in authentic real-life settings such as a classroom setting with multiple distracters and with peers and a teacher present.

Research on how distracters affect cognitive performance shows some contradictory results, depending on the type of distracter (Söderlund, Sikström, Loftesnes, & Sonuga-Barke, 2010). In general, children's cognitive performance is hampered in situations where children are confronted with distracters, such as meaningful irrelevant speech sounds, road traffic sounds (for a short review, see Klatte, Bergström, & Lachmann, 2013) or visual distracters (Dolcos, Miller, Kragel, Jha, & McCarthy, 2007). Also, visual distracters affect WM performance negatively, especially when those distracters can be confused with the information to-be-remembered (Dolcos et al., 2007). Moreover, the relation between WM and comprehension of spoken sentences was stronger in the presence of background noise than in a quiet environment (Sullivan, Osman, & Schafer, 2015), which may suggest that children's WM is taxed more heavily for language comprehension in a noisy classroom setting than in a controlled testing situation. Together, these findings indicate that the distribution of WM scores in a controlled testing situation does not adequately reflect the availability of WM resources in a noisy classroom setting. Results of assessing cognitive and academic skills in a controlled testing situation may therefore not be as ecologically valid as typically assumed. Importantly, a recent study demonstrated that assessing WM in 12-year-old children in a classroom setting yielded stronger associations with other cognitive functions than when WM was assessed individually in a quiet environment (Kanerva et al., 2019), suggesting that WM assessed in a situation with distracters present better represents the capacity for performing cognitive tasks. The authors concluded that classroom assessed WM adequately captured performance in WM. However, because different instruments were used in the individual versus classroom setting, a direct comparison between both settings was not feasible. Although these findings suggest that the classroom may provide an ecologically valid setting to assess WM functioning, more research using identical tasks in both settings is needed.

Notably, children's attentional capacities may moderate the detrimental effect of distracters. On the one hand, research shows that children who have difficulties with attention are more vulnerable to distractions than typically developing children (Cassuto, Ben-Simon, & Berger, 2013; Geffner, Lucker, & Koch, 1996). On the other hand, under certain circumstances, background noise can actually improve children's performance. That is, random noise, with moderate intensity and noise level (also called "white noise"), can improve WM performance in children who have difficulties with attention (Söderlund et al., 2010), while it turns out to hamper WM performance in typically developing (Söderlund et al., 2010; Sullivan et al., 2015) or highly attentive children (Helps, Bamford, Sonuga-Barke, & Söderlund, 2014). This differential effect of white noise is theorised to reflect that children with attention difficulties have lower levels of neural noise, while the neural system requires a certain level of noise to reach an optimal state of arousal and alertness to detect signals. According to this theory, adding white noise can bring inattentive children in a more optimal state of arousal, while it is detrimental for children who already have an optimal level (for more information see Helps et al., 2014).

Another cognitive resource (partly) responsible for the differential susceptibility to various distracters may be inhibition, which refers to children's ability to suppress automatic, prepotent, or inappropriate responses (Miyake et al., 2000). The construct is typically divided into response inhibition, which is the suppression of a behavioural response, and cognitive inhibition (also called interference control), which is responsible for supressing attention toward previously activated but non-relevant information and directing attention towards relevant information, both of which have shown to be separable skills (Brydges, Anderson, Reid, & Fox, 2013). Response inhibition (Gilmore, Keeble, Richardson, & Cragg, 2015) and cognitive inhibition (e.g., Laski & Dulaney, 2015) are related to measures of academic achievement. Hypothetically, both behavioural and cognitive responses to stimuli, in this case responses to the distracter that is present in the direct environment of the child, could be suppressed if inhibitory capacities are adequate. Consequently, well-developed inhibitory capacities should allow a child to fully devote the available WM capacity to task-relevant information rather than irrelevant distracters. Indeed, many children with WM difficulties showed high levels of distractibility (Gathercole et al., 2008), which suggests that their WM difficulties may be attributed to a certain extent to an inability to suppress cognitive or behavioural responses to irrelevant stimuli. In the current study, we focussed on response inhibition, which may help a child inhibit behavioural responses to distracters, such as looking up at other children rather than focussing on an academic task. There is a modest base of evidence that demonstrates a relation between response inhibition and WM performance: for example, in children of age 5–6 (but not age 8-9), these constructs were significantly related (Tsujimoto, Kuwajima, & Sawaguchi, 2007), and response inhibition improved along with WM capacity in a WM training for children with ADHD (Klingberg et al., 2005). Moreover, it has been demonstrated in adult subjects with ADHD that response inhibition is related to distractibility, i.e., looking away from a task (Adams, Roberts, Milich, & Fillmore, 2011). If some children are indeed more inclined to look away from tasks in the face of distracters, response inhibition may (partially) predict a drop in task performance between quiet test environments and classroom environments.

In summary, because the presence of distracter stimuli in the environment may affect WM performance of children differentially (e.g., Söderlund et al., 2010), the assessment of WM in a quiet, controlled setting may be less ecologically valid. Therefore, the present study aimed to investigate differential performance and associations of WM tasks conducted in classroom and individual settings as to make inferences regarding ecological validity of testing situations. To be able to make a direct comparison between classroom and individual WM performance and how these are associated with academic performance, the same WM tasks were used in both settings. The research questions were:

1. Do children's scores on visuospatial and verbal WM tasks differ when tasks are conducted individually compared to when they are conducted group-wise in a classroom setting?
2. Is variation in any existing difference in scores related to children's attentional capacities and/or response inhibition?
3. Which setting, classroom or individual testing, yields the best WM predictors for mathematics performance and reading comprehension?

Based on previous findings that the presence of distracter stimuli negatively affected WM performance of typically performing children (Klatte et al., 2013), we expected children's performance on average to be higher in individual sessions than in a classroom setting. Moreover, we expected that the decline in performance in a classroom setting is larger in children who have weaker attentional capacities and response inhibition. Finally, we expected that classroom WM assessment would better predict mathematics and reading performance than individual WM assessment.

Method

Participants

A total of 108 children from five primary schools in the Netherlands participated in the study. All children with missing data on any of the WM measures were excluded from the analyses: two children were absent from school during testing, and seven children had missing data due to technical errors. For one child, the mathematics score was missing. This child's scores were included in analyses in which mathematics was not relevant. Also, scores for attention of 15 children were missing. These children were included in all analyses in which attention was not relevant. The final dataset contained data of 99 children from grade 3 ($n = 43$), 4 ($n = 41$), and 5 ($n = 15$). At the end of data collection, the children were on average 9;6 years old ($SD = 9.45$ months).

Procedure

The procedure and research questions were approved before the start of the study by the Faculty Ethics Review Board of the Faculty of Social and Behavioural Science, Utrecht University. All children were tested during January and February 2017 on WM in two settings: a classroom setting taking place in the classroom in the presence of all children, and an individual setting in which only a trained graduate student was present. In the individual testing session, children also performed a computerised response inhibition task. To preclude testing order effects, half of the children were first tested in the classroom setting and then in the individual setting, and the other half of the children performed the tests in the reverse order. Children were assigned to these testing orders randomly. Time between the two sessions was one (minimum) to two (maximum) weeks.

During the classroom assessment, children first performed the visuospatial WM task and then the verbal

WM task. This took place on a computer in the classroom while other children worked independently on other tasks. Instructions were embedded in the tasks and given through audio-recordings using on-ear headphones. A graduate student was available throughout the session to answer questions. The same procedure was used for individual assessments, but tasks were conducted in a quiet room inside the school. After completing the WM tasks in the individual session, the inhibition task was started. The graduate student would sit at the table next to the child at a 90° angle. During the WM assessment in both settings (classroom and individual) environmental noise was measured over the duration of both the visuospatial and the verbal WM task to check whether auditory distraction differed between settings.

Mathematics achievement and reading comprehension were tested as part of the standard test battery of the schools. The results of the latest assessments in the middle of the academic year (January-February), which are typically administered in a classroom setting, were requested from the teacher. During this period, the teachers were also requested to fill out a rating scale of (in)attentive behaviour for the participating children.

Instruments

Visuospatial working memory

Visuospatial WM was indexed using the Lion game. In this complex-span task, children recall the locations of coloured lions. In a 4 × 4 matrix, eight lions of five different colours are presented consecutively for 2,000 ms each on the screen, after which the child is asked to recall the location of the last lion of a certain colour, announced at the start of the trial. The number of coloured lions (and thus the number of locations) that the child is asked to recall increases from one lion at level 1, to five lions of different colours at level 5. The task starts with two practice items followed by 20 test items across five levels. For each set, a proportion of correctly recalled locations was computed by dividing the number of correctly recalled locations of a set by the number of items within that set. The score of the child was the average proportion of correctly recalled locations across all items. Internal consistency of the Lion game is excellent (α ranges from .86 to .90), test-retest reliability satisfactory, and there are adequate concurrent relationships with other WM tasks. The task is predictive of subsequent mathematics performance (Van de Weijer-Bergsma, Kroesbergen, Prast, & Van Luit, 2015a).

Verbal working memory

The Monkey game was used to measure verbal WM. In this backward word recall task, children are presented with an audio recording containing a spoken string of one-syllable high-frequency words at a rate of one word per 1,000 ms, and asked to recall all of the words in the reverse order by clicking on the corresponding written word in a 3 × 3 matrix. The number of words to recall in the reverse order increases with one word in each level, from two words in level 1 to six words in level 5. Four practice items are followed by 20 test items across five levels. Scoring of the task is similar to that of the Lion game. Internal consistency of the task is excellent (α ranges from .81 to .88), concurrent relationships with the other WM measures good, and performance on the Monkey game was found to be predictive of subsequent mathematics performance and reading comprehension (Van de Weijer-Bergsma et al., 2016).

Teacher-reported attention

The Strengths and Weaknesses of ADHD symptoms and Normal behaviour rating scale (SWAN; Swanson et al., 2012) was designed to measure inattentive, hyperactive, and impulsive behaviour in both typically performing children and children with attentional difficulties. The instrument consists of 18 questions targeting behaviours that are indicative of attentional functioning, based on DSM-5 criteria for ADHD. Behaviours are rated on a 7-point Likert scale, and higher total scores are indicative of better developed attentional capacities. The scale has been demonstrated to have good test-retest reliability and scores high on various indicators of validity (Arnett et al., 2013). A reliability analysis using the current data showed that the internal consistency between items was high, Cronbach's $\alpha = .98$.

Response inhibition

A Go/No-go task adopted from De Weerdt, DeSoete, and Roeyers (2013) and created using E-Prime software was used to measure response inhibition. In this task, the child was asked to respond to 75% of presented trials as quickly as possible by a key press, but to inhibit the response to 25% of the trials, characterised with a different stimulus, by not pressing the key. Each trial started with a fixation cross for 500 ms. Then the stimulus (e.g., a bird) was shown. The stimulus disappeared upon a key press, and remained on the screen for a maximum of 300 ms, after which it disappeared but a response could still be entered. The trials had three modalities, appearing in three sub-sessions: pictures (birds and butterflies), letters (a and m) and numbers (1 and 6). Before test items started, six practice items were presented to make sure that the children understood the task. The practice trials contained 50% go- and 50% no-go trials. The test items consisted of 75% go trials (birds, a's, and 1's; 30 items per modality) and 25% no-go trials (butterflies, m's, and 6's; 10 items per modality). Commission errors, which is the number of times each child responded mistakenly to a no-go trial, were used as a measure for response inhibition. Test-retest correlations of a similar task have been found to be moderate to high (Kuntsi, Andreou, Ma, Börger, & Van der Meere, 2005).

Mathematics achievement

Mathematics achievement was indexed using the Dutch national norm-referenced Cito Mathematics tests (Janssen, Scheltens, & Kraemer, 2005). These tests are conducted twice during each year of primary school: once in the middle of the academic year, and once at the end. Test items primarily consist of context problems, targeting a child's understanding of various age-appropriate domains of mathematics: number and number relationships, addition and subtraction, multiplication and division, mental calculation and estimation, written calculation, use of a calculator, geometry, time and money, and fractions and percentages. Raw scores are converted to ability scores that typically increase over academic years. The test is reliable with α ranging from .91 to .97 and valid (Janssen, Verhelst, Engelen, & Scheltens, 2010). Because two different versions were in use at different schools, ability scores were standardised prior to analysis.

Reading comprehension

Reading comprehension was measured using the Dutch norm-referenced Cito tests of Reading Comprehension (Feenstra, Kamphuis, Kleintjes, & Krom, 2010; Weekers, Groenen, Kleintjes, & Feenstra, 2011). Each test consists of different texts and an accompanying set of 50 (grades 1–4) or 55 (grades 5–6) multiple choice questions. Raw scores are converted to ability scores that make a comparison to national norms possible. The instrument has been found to be reliable (Cronbach's α = .84 to .93) and valid (Feenstra et al., 2010). Because two different versions were in use at different schools, ability scores were standardised prior to analysis.

Environmental noise

The noise level during WM assessment was measured in decibel (dB) using the Sound Metre HQ application for mobile devices, developed by ExaMobile S.A. This free application is available for Android and iOS systems and uses the built-in microphone of a smartphone to measure level of noise. The application provides lowest, highest and average dB measured over a period of time. The average dB over the period during which the two WM tasks were administered was used.

Analyses

To address the first research question, targeting the difference in WM performance between the classroom and individual setting, we conducted two repeated measures ANCOVA's, one for visuospatial WM and one for verbal WM, with classroom and individual setting as a within-subjects factor. In the next step, to address the question whether differences in WM scores between classroom and individual settings are moderated by attention and/or response inhibition, we added the SWAN scores or the number of commission errors on the Go/No-go task respectively as covariates. The interaction between setting and the attention or inhibition measure was an indicator of the extent to which attention and response inhibition related to a difference in scores.

The final research question, targeting the predictive value of classroom and individual WM scores for academic skills, was addressed using dominance analysis. This statistical technique is especially appropriate in the case where multiple predictors share large proportions of variance, known as multicollinearity, such as when various measures are used to index the same trait (Azen & Budescu, 2003; Kraha, Turner, Nimon, Zientek, & Henson, 2012). This is particularly relevant in the current study where identical tasks are used to assess the same ability in different settings. In contrast to traditional regression analysis, where the focus lies on interpreting regression coefficients, in dominance analysis, multiple regression analyses are used, and the explained variance of all possible subset models is examined. Because these multiple regressions contain all possible combinations of predictors, this method takes into account that the association of a certain predictor with the criterion varies depending on the other predictors in the submodel and allows us to determine which of a group of predictors holds the most "importance". A predictor can only be seen as more important than another predictor if its contribution in terms of explained variance to various models is greater than that of the other predictor given the inclusion of every possible subset of predictors (Azen & Budescu, 2003; Budescu, 1993). In this case, a predictor is determined to be completely dominant over another predictor. Because complete dominance is infrequent in social sciences, Azen and Budescu (2003) introduced two more levels of dominance to characterise the relative importance of predictors: conditional dominance is achieved when the average contribution of a predictor is greater than that of another predictor across averages in additional contribution within models containing a given number of predictors. General dominance is achieved when additional contributions of a predictor to any of the models *on average* is greater than that of another predictor (Azen & Budescu, 2003). These levels of dominance are related hierarchically: a predictor that holds complete dominance over another predictor also holds conditional and general dominance, and a predictor that holds conditional dominance over another predictor also holds general dominance. Two dominance analyses were performed: one for predicting mathematics performance and one for predicting reading comprehension. Regression analyses contained all possible combinations of the four WM measures as predictors: performance on visuospatial WM in a classroom and individual session, and performance on the verbal WM in a classroom and individual session.

Results

Descriptive statistics of all variables can be found in Table 1. Correlations between variables can be found in Table 2. The correlation analyses showed that both verbal and visuospatial WM in either setting significantly correlated amongst each other, as well as with measures of academic

Table 1. Descriptive Statistics of Measures of Working Memory (WM), Response Inhibition, Attention, and Academic Performance.

	n	*M*	*SD*	Min	Max
Visuospatial WM: individual	99	.80	.10	.51	.96
Visuospatial WM: classroom	99	.77	.11	.43	.99
Verbal WM: individual	99	.61	.09	.44	.85
Verbal WM: classroom	99	.59	.11	.12	.80
Response inhibition: commission errors	99	5.67	3.47	0.00	15.00
SWAN attention	84	38.33	11.84	15.00	63.00
Mathematics Z (old version)	25	0.00	1.00	−2.69	1.83
Mathematics Z (new version)	73	0.00	1.00	−3.82	2.78
Reading comprehension Z (old version)	52	0.00	1.00	−1.92	2.62
Reading comprehension Z (new version)	47	0.00	1.00	−1.99	2.08

Note. WM scores are percentages of correctly recalled items. Response inhibition commission errors is the number of errors. Mathematics and reading comprehension scores are standardised ability scores.

achievement and attentional capacities as rated by the teacher. The response inhibition measure, however, did not correlate with any of the other measures.

Prior to analyses, effects of testing order were investigated using a repeated measures ANOVA with classroom and individual setting as a within-subjects factor for both visuospatial WM and verbal WM, and testing order (individual first or classroom first) as a between-subjects factor. The analysis showed that testing order significantly affected performance on visuospatial WM, $F(1,97) = 15{,}84$, $p < .001$, partial $\eta^2 = .14$, but not on verbal WM, $F(1,97) = .04$, $p = .85$, partial $\eta^2 < .01$. We therefore decided to include testing order as a covariate in all analyses targeting the difference in performance between individual and classroom testing for visuospatial WM, but not in analyses targeting performance differences in verbal WM.

Also prior to analysis, a paired-samples *t*-test was conducted to compare environmental noise levels in the classroom and individual setting. There was a significant difference between settings, $t(98) = -17.37$, $p < .001$, with lower noise levels in dB in the individual setting, $M = 32.76$, $SD = 4.71$, compared to the classroom setting, $M = 49.60$, $SD = 9.06$.

Differences in performance between settings

The repeated measures ANCOVA for visuospatial WM, with testing order included as a covariate because the testing order affected scores, showed a significant main effect of setting with a medium effect size, $F(1,97) = 6.10$, $p = .02$, partial $\eta^2 = .06$, which indicated that children performed better in the individual setting than in the classroom setting. There was also a main effect of testing order, $F(1,97) = 6.66$, $p = .01$, partial $\eta^2 = .06$, with children who first performed the individual setting obtaining higher scores than the children who first performed the classroom setting, and an interaction between setting and testing order, $F(1,97) = 14.81$, $p < .01$, partial $\eta^2 = .13$.

For verbal WM, we conducted a repeated measures ANOVA, which showed that children scored significantly higher on the WM measures in the individual setting than in the classroom setting, $F(1,98) = 8.05$, $p < .01$, partial $\eta^2 = .08$. This is a medium effect. It should be noted that testing order was not included as a covariate in this analysis because the exploratory analyses did not show a significant effect of testing order.

In a next step, SWAN attention scores were added to the models as a covariate. Results of these analyses indicated that attentional capacities predicted part of the variance in Visuospatial WM scores, $F(1,81) = 6.30$, $p = .01$, partial $\eta^2 = .07$, with testing order included as a covariate because the testing order affected scores. However, attention did not moderate the effect of setting (setting*attention interaction) on Visuospatial WM, $F(1,81) = 0.02$, $p = .89$, partial $\eta^2 < .001$, indicating that attention did not relate to the reported difference in performance. For Verbal WM, on the other hand, attentional capacities both predicted variance in the scores on its own, $F(1,82) = 24.73$, $p < .001$, partial $\eta^2 = .23$, and moderated the effect of setting (setting*attention interaction), $F(1,82) = 7.59$, $p < .01$, partial $\eta^2 = .09$. This means that differences between individual and classroom WM scores were larger for children with lower attentional capacities as rated by the teacher.

When response inhibition was added as a covariate to the analyses, results indicated that inhibition did not predict scores on Visuospatial WM tasks, $F(1,96) = 0.27$, $p = .61$, partial $\eta^2 < .01$, nor did it moderate the effect of setting (setting*inhibition interaction) on Visuospatial WM performance, $F(1,96) = 0.08$, $p = .77$, partial $\eta^2 < .01$. Testing order was included as a covariate because the testing order affected scores on Visuospatial WM. Moreover, response inhibition scores did not predict verbal WM performance, $F(1,97) = 3.19$, $p = .08$ partial $\eta^2 = .03$, nor did it moderate the effect of setting (setting*inhibition

Table 2. Correlations between Measures of Working Memory (WM), Response Inhibition, Attention, and Academic Performance.

	1.	2.	3.	4.	5.	6.	7.
1. Visuospatial WM: individual	–						
2. Visuospatial WM: classroom	.62**	–					
3. Verbal WM: individual	.41**	.35**	–				
4. Verbal WM: classroom	.29**	.33**	.51**	–			
5. Response inhibition: commission errors	−.06	−.02	−.20	−.12	–		
6. SWAN attention	.26*	.32**	.33**	.51**	−.07	–	
7. Mathematics	.33**	.25*	.33**	.35**	−.15	.22*	–
8. Reading comprehension	.29*	.21*	.39**	.39**	−.08	.39**	.61**

Note. Ability scores of Mathematics and Reading comprehension were standardised within each version and merged into single variables. * $p < 05$, ** $p < .01$.

Table 3. Dominance Analysis Predicting Mathematics Performance Using Individually and Classroom Conducted Working Memory Assessments.

		Additional contribution of			
Subset model	R^2	1 (Visuospatial indiv)	2 (Visuospatial class)	3 (Verbal indiv)	4 (Verbal class)
***k* = 0 average**		**.106**	**.063**	**.110**	**.123**
1 (Visuospatial indiv)	.106	–	.004	.047	.073
2 (Visuospatial class.)	.063	.047	–	.067	.080
3 (Verbal indiv)	.110	.043	.020	–	.044
4 (Verbal class)	.123	.056	.020	.031	–
***k* = 1 average**		**.049**	**.015**	**.048**	**.066**
1,2	.110	–	–	.044	.069
1,3	.153	–	.001	–	.037
1,4	.179	–	.000	.011	–
2,3	.130	.024	–	–	.035
2,4	.143	.036	–	.022	–
3,4	.154	.036	.011	–	–
***k* = 2 average**		**.032**	**.004**	**.026**	**.047**
1,2,3	.154	–	–	–	.036
1,2,4	.179	–	–	.011	–
1,3,4	.190	–	.000	–	–
2,3,4	.165	.025	–	–	–
***k* = 3 average**		**.025**	**.000**	**.011**	**.036**
1,2,3,4	.190	–	–	–	–
Overall average		**.053**	**.021**	**.049**	**.068**

Note. *k* is the number of additional variables to the variable listed in the designated row.

interaction) on verbal WM performance, $F(1,96) = 0.10$, $p = .75$, partial $\eta^2 < .01$. This means that response inhibition did not have an effect on either WM performance, or on performance differences between the individual and classroom settings.

Working memory predicting academic achievement

An overview of explained variances of WM measures predicting Mathematics performance can be found in Table 3. The analyses showed that classroom obtained scores for verbal WM *completely* dominated individually obtained scores for both visuospatial WM and verbal WM, and classroom obtained scores for visuospatial WM. This can be inferred from the fact that the additional contribution of classroom verbal WM to each of the subset models forming the basis for comparison is greater than that of each of the other predictors. Moreover, individually obtained scores on visuospatial WM *completely* dominated classroom obtained scores on visuospatial WM, and *generally* dominated individually obtained scores on verbal WM. Finally, individually obtained scores on verbal WM *completely* dominated classroom obtained scores on visuospatial WM. Classroom obtained scores on visuospatial WM did not show dominance of any sort over any other variable. To summarise, it can be tentatively concluded that classroom performance on verbal WM is the best predictor of mathematics achievement, followed by individual performance on visuospatial WM, individual performance on verbal WM, and then classroom performance on visuospatial WM.

An overview of explained variances of WM measures predicting Reading comprehension can be found in Table 4. The analyses showed that classroom obtained verbal WM scores *completely* dominated individually obtained visuospatial WM and verbal WM scores and classroom obtained visuospatial WM scores. Also, individually obtained scores on verbal WM *completely* dominated individually and classroom obtained scores on the visuospatial WM. Finally, individually obtained visuospatial WM scores *completely* dominated classroom obtained visuospatial WM scores. From these data, it can be concluded that classroom verbal WM performance is the best predictor of reading comprehension, followed by individually obtained verbal WM scores, individually obtained visuospatial WM scores on the Lion game, and finally classroom obtained visuospatial WM scores.

Discussion

The current study aimed to compare performance of children on WM tasks in a controlled, individual testing situation to performance in a classroom setting. Consistent with our expectations, the results showed that performance in an individual setting was significantly better than performance in a classroom setting: children recalled more items of both the visuospatial and the verbal WM task in the individual setting, regardless of testing order. Variability in decline in performance was explained by teacher-reported attentional functioning of the children, but not by our measure of response inhibition. Partially in line with our expectations, verbal WM – but not visuospatial WM – functioning in a classroom setting was a stronger predictor of academic performance than WM functioning in an individual setting; a more detailed summary of these findings can be found below.

The finding that children show lower performance on WM tasks in a classroom setting is in line with studies indicating that the cognitive performance of typically performing children is hampered in situations where they are confronted with auditory and visual distracters (Dolcos

Table 4. Dominance Analysis Predicting Reading Comprehension Using Individually and Classroom Conducted Working Memory Assessments.

		Additional contribution of			
Subset model	R^2	1 (Visuospatial indiv)	2 (Visuospatial class)	3 (Verbal indiv)	4 (Verbal class)
***k* = 0 average**		**.082**	**.042**	**.148**	**.153**
1 (Visuospatial indiv)	.082	–	.001	.086	.104
2 (Visuospatial class)	.042	.041	–	.110	.118
3 (Verbal indiv)	.148	.020	.006	–	.052
4 (Verbal class)	.153	.033	.007	.047	–
***k* = 1 average**		**.031**	**.005**	**.081**	**.091**
1,2	.083	–	–	.085	.104
1,3	.168	–	.000	–	.046
1,4	.186	–	.001	.028	–
2,3	.154	.014	–	–	.047
2,4	.160	.027	–	.041	–
3,4	.200	.014	.001	–	–
***k* = 2 average**		**.018**	**.001**	**.051**	**.066**
1,2,3	.168	–	–	–	.047
1,2,4	.187	–	–	.028	–
1,3,4	.214	–	.001	–	–
2,3,4	.201	.014	–	–	–
***k* = 3 average**		**.014**	**.001**	**.028**	**.047**
1,2,3,4	.215	–	–	–	–
Overall average		**.036**	**.012**	**.077**	**.089**

Note. *k* is the number of additional variables to the variable listed in the designated row.

et al., 2007; Klatte et al., 2013). Moreover, this study expands on laboratory-based research outcomes by demonstrating that a decline in cognitive performance also occurs in a classroom setting, where distracters are part of an environment that is not manipulated as part of the study-design.

The finding that attentional capacities as rated by the teacher interacted with the setting in which the tasks were performed, demonstrates that the decline in scores in the classroom setting was sharper for children with lower attentional capacities. This is in line with previous studies showing that children who have more difficulties with attention are also more vulnerable to distractions (Cassuto et al., 2013; Geffner et al., 1996). The interaction effect, however, was only present for scores on verbal WM and not for scores on visuospatial WM, which indicates that a drop in performance was dependent on attention regulation only when it concerned the verbal WM task. The fact that attentional capacities did not relate to the drop in performance in the visuospatial WM task may be due to the fact that visual information during the Lion game was presented briefly and constant visual attention was needed for the task; if children looked away from the screen, as elicited by a classroom situation, they would miss a stimulus and fail a trial. Because constant visual attention is not needed for typical academic tasks, it is also likely not captured in teacher assessments of attentional functioning, which focus on behavioural indicators of attention regulation that are much less subtle than briefly looking away from a task, such as awaiting turns. Also, attention for verbal information is possibly captured more fully by the instrument, through questions about listening to instructions, than attention for visuospatial information, which may have a less obvious role in the teacher assessment.

Additionally, we found that the association of the teacher-rating of attention with both the verbal and visuospatial classroom WM assessments was stronger than with the individual WM assessments. Although this difference only reached significance for the verbal WM task[1], this is in line with the assumption of veridicality (i.e., the degree to which performance is *empirically* related to measures of everyday functioning) and is further confirmation of the ecological validity of classroom WM assessment.

In contrast with attention, our measure of response inhibition could not explain variability in children's decline in WM functioning in the classroom setting, which suggests that it is not the level of response inhibition of the children that determines their vulnerability to distractions in their environment as reported in a previous study (Adams et al., 2011). There are several explanations possible for our findings. First, although the cited study demonstrated a link between response inhibition and distractibility, the measure of inhibition employed in the current study does not directly reflect the type of inhibition that allows a person to block out distracters such as present in the classroom situation. We opted to include a measure of response inhibition in the study with the rationale that children with better response inhibition would be better able to resist a behavioural response to external stimuli, such as looking up at classroom noise. Response inhibition, however, is separable from cognitive inhibition (Brydges et al., 2013), which helps a child keep irrelevant stimuli out of the attentional focus, albeit these stimuli are typically intertwined with the task in cognitive testing. Yet, a measure of cognitive inhibition, which could also be relevant for ignoring classroom noise and visual distracters, may explain variance in the decline in performance in the classroom setting better. Perhaps the behavioural response occurs later in the inhibitory process than the cognitive distraction: some children may be distracted at a cognitive level but are still able to suppress a behavioural response. Moreover, as the response inhibition task was administered in the individual setting, it may have given a less representative measure of the child's inhibitory functioning in the classroom. Perhaps that including a inhibition measure in the classroom setting would have yielded different results. Finally, the response inhibition measure may not have adequately indexed inhibitory capacities of the children in this sample of typically developing children. It should be noted that the number of commission errors made by the children also failed to correlate with WM measures or with measures of academic achievement, which is in sharp contrast with previous studies in which inhibition showed significant relations with other measures of cognitive and academic functioning (e.g., Cragg, Keeble, Richardson, Roome, & Gilmore, 2017; Friso-van den Bos et al., 2013; Miyake & Friedman, 2012). Although a full account of issues with the measurement of inhibition is beyond the scope of the current study, a recent review by Lee and Lee (2019) has listed some compelling observations, such as the relatively small amount of errors made in Go/No-go tasks posing a problem for differentiation. If this task indeed is not able to differentiate between children with varying levels of response inhibition, variability may stem from other factors, such as motivation, sustained attention, or simply from measurement error. If this is the case, the validity of the task is jeopardised, explaining the absence of significant correlations with other variables. To be able to draw conclusions about the role of inhibition in the decline of WM performance in the classroom, future studies should incorporate additional or alternative measures of response inhibition as well as measures of cognitive inhibition.

Furthermore, the finding that verbal WM performance in a classroom setting is a better predictor of academic tasks than other WM predictors, including the individual counterpart of the task is in line with the assumption that the level of verisimilitude is higher when a cognitive test resembles the cognitive demands of real-life tasks (Chaytor & Schmitter-Edgecombe, 2003) and indicates that WM assessment in the classroom setting better mimics the cognitive demands of real-life situations than individual WM assessment, especially when the aim is to predict and explain variation in academic skills. After all, the academic skills for which WM is recruited are usually taught and tested in the setting of a classroom that comes with the distracters that contributed to the dominance structure between WM variables. This finding is in line with the study by Kanerva et al. (2019), who reported that classroom assessed WM was able to explain variance in academic skills, even when individually assessed WM performance was controlled for. However, contrary to our expectation, visuospatial WM as assessed in the classroom in the current study was the least predictive for both mathematics and reading comprehension, as shown by an absence of dominance for this measure.

Moreover, the individual assessments of verbal and visuospatial WM showed patterns of dominance that are in line with previous research showing that both visuospatial and verbal WM are involved in mathematics achievement (e.g., Friso-van den Bos et al., 2013; Peng et al., 2016) and that verbal WM is more strongly involved in reading comprehensions than visuospatial WM (Savage et al., 2007; Seigneuric et al., 2000). More specifically, the dominance pattern for mathematics performance showed that the dominance of individual verbal and visuospatial WM was conditional and thus differed between subset models. However, on average, visuospatial WM was more predictive than verbal WM for mathematics achievement when assessed individually. This implies that, when assessing WM in an individual setting, visuospatial WM will give the strongest prediction in most situations, but adding a verbal WM task will add predictive value. Additionally, the dominance pattern for reading comprehension showed that individual verbal WM displayed complete dominance across all models, which implies that it is a stronger predictor than individual visuospatial WM for reading. This contrast in dominance patterns, with differences between subset models in the prediction of mathematics and complete dominance of verbal WM in the prediction of reading comprehension is consistent with the pattern observed in published studies: the most prominent WM predictor of mathematics performance differs between studies (for reviews, see Friso-van den Bos et al.; Peng et al.), but verbal WM is a very consistent most prominent predictor of reading comprehension (e.g., Savage et al., 2007; Seigneuric et al., 2000).

The finding that *classroom verbal* WM was the most dominant predictor in both mathematics and reading comprehension suggests that specifically for verbal WM, classroom assessments are more ecologically valid compared to individual assessments. Furthermore, it might suggest that auditory distracters had a larger influence than visual distracters in explaining individual performance differences, at least in our study. Specifically, we hypothesise that verbal information in the presence of auditory distracters can briefly be stored in a sensory memory trace and be retrieved for active processing in WM with varying rates of success dependent on age (e.g., Glass, Sachse, & von Suchodoletz, 2008). This process is likely to occur both in WM tasks conducted in a classroom setting and typical academic tasks in a classroom, explaining the dominance of verbal WM in predicting academic performance. The strong associations between academic tasks and verbal WM may also imply that children use more verbal solution strategies (Hecht, 2002; Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2015b), or they may reflect the high frequency of verbal instruction in classroom settings, producing stronger associations with verbal WM.

The finding that the *classroom visuospatial* WM assessment showed an opposite pattern with no dominance over any of the other WM tasks is striking, and task characteristics (e.g., modality) and study design can possibly explain these findings. With regard to task modality, the previously mentioned visual presentation of information to be remembered may explain the finding that individual assessment gives a better reflection of visuospatial WM as involved in academic tasks. While distractions are much less disruptive when information is presented verbally, these distractions may have produced noise in the visual WM task that was irrelevant for predicting scholastic performance, which is typically not hampered by occasionally looking up from one's work. Although Kanerva et al. (2019) also used a WM task with a visual component, they did not report on decreasing task performance when attention was disrupted. In their counting span task, participants were asked to count visually presented dots and store the number of dots in each set in memory. It may be that Kanerva et al. (2019) did not detect any effect of such disruptions because task execution was more self-paced than in the task used in the current study, allowing for brief periods of shifted attention, or because the authors analysed the counting span task in a composite score with a reading span task, which may have levelled out such noise. Also, although the information in the counting span task was presented visually, the number of dots counted was likely stored verbally, making visual disruption of less relevance in their study. Moreover, although Kanerva et al. (2019) also did not manipulate or control the level of noise in their study, their participants were instructed to be as quiet as possible during group assessment, which may have reduced the number and level of distractions compared to our study. Finally, a counting span task may trigger more continuous processing and more focussed attention, while the updating component of our Lion game may have triggered a more passive attentional screening of which colour lions appear, creating a difference in how easily attention is distracted. This suggests that task characteristics and participant instructions need careful consideration when visuospatial WM tasks are used for classroom assessment.

The ecological validity of neuropsychological tests sparks continuous discussion (Barkley, 1991; Burgess et al., 2006). Although the ecological validity of laboratory tests seems to be higher in neurologically impaired samples than in typically performing samples (Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2008), clinical (neuro)psychologists are often perplexed by the inconsistency between clients performance on tests of EF and their everyday EF abilities (Chaytor, Schmitter-Edgecombe, & Burr, 2006). Some children may be more affected by environmental distracters than others, and this may not become visible when using assessments in a laboratory (-like) setting. Indeed, several studies using virtual reality classrooms have shown that attentional functioning is hampered more by distracters in children with Attention Deficit Hyperactivity Disorder (ADHD) than in typically developing children (Adams, Finn, Moes, Flannery, & Rizzo, 2009; Neguţ, Jurma, & David, 2017). Classroom assessment, or a combination of (and comparison between) individual and classroom assessments may give a more complete picture of a child's cognitive functioning in clinical samples. Moreover, a comparison between the two situations may give insight into which practical adaptations may be suitable to support a child's WM functioning in the classroom (e.g., limiting visual or auditory distracters during tasks that rely heavily on a domain of WM that is affected by the testing situation).

It should be noted however, that besides the presences of distracters, other factors could have been of influence on our findings. For example, the presence of a unfamiliar test leader may act as an external regulator of behaviour in the individual setting. This is not typical for classroom settings, where one-on-one supervision is a rare occurrence and activities are typically overseen by the more familiar class teacher. In addition, emotions related to learning, such as anxiety, may also be activated more in the classroom, possibly explaining the stronger relation between classroom WM assessment and academic achievement (Ashcraft & Kirk, 2001; Valiente, Swanson, & Eisenberg, 2012). Future research may focus on unravelling how situational aspects trigger motivation, emotions and self-regulation, and how these – in interaction – influence cognitive functioning in the classroom setting.

Another limitation of the current study is that it did not incorporate rigorous control over the nature, timing and extent of the distracters. Although natural variability in the extent to which distracters are present is a necessary condition for investigating the ecological validity of assessments, it may have increased variability (data noise) in the data in terms of the level or pattern of decline in each individual. Indeed, the variation in environmental noise levels was two times larger in the classroom setting compared to the individual setting. In line with expectations however, the measurement of environmental noise indicates that auditory distracters were not only more variable, but also on average more present in the classroom setting compared to the individual setting. The dB levels in the classroom setting were in line with the typical classroom setting when pupils are engaged in independent work but there is still some noise (Shield, Greenland, & Dockrell, 2010), while the dB levels in the individual setting were comparable to soft whispering and coincide with the levels that are recommended as an optimal level for classroom work (Shield & Dockrell, 2003). However, as participants used on-ear headphones, they most likely experienced lower levels of noise in both settings and the distraction in the classroom setting is still less than typically faced by students, even though the headphones were not designed to cancel out environmental noise (which is the case in *over*-ear headphones). This design aspect of the study may have attenuated the effects. Nevertheless, these findings may be interpreted in the light of current developments in daily classroom practice, where computerised instructions, learning tasks and assessments become incorporated more and more. As the presence of visual distracters may also have varied between participants, we recommend future studies to observe these as well or try to control for them.

In conclusion, the current study showed that WM performance is dependent on the setting in which the construct is tested, with higher performance in controlled testing situations than in a classroom situation, in which WM is typically recruited for learning tasks. Moreover, verbal WM as tested in the classroom setting proved a better predictor of mathematics and reading attainment than verbal WM as tested in a controlled testing situation, supporting the conclusion that classroom assessment is a more ecologically valid method of indexing verbal WM as a predictor of academic attainment. In clinical and educational practice, testing children's verbal WM in the classroom instead of, or in addition to, in individual settings may prove a valuable addition to clinical assessment. Moreover, classroom assessment is a feasible, low-cost way to collect data for large-scale studies. For visuospatial WM on the other hand, individual assessments seem to better index children's capacities. This does not mean that classroom assessments of visuospatial WM have no worth, as the classroom administered task correlated well with both its individually administered counterpart and with academic performance, but that in the case of visuospatial WM, individual assessments holds a small but important advantage. Of response inhibition and attentional capacities, only attention could explain individual variation in the decline in performance when WM was tested in a classroom setting, specifically for verbal WM. Additional research is needed to expose the nature of this decline, and investigate its relation to other cognitive and affective processes.

Note

1. Steiger's Z (Z_H) was used to test whether differences in dependent correlations are statistically significant (Hoerger, 2013; Steiger, 1980), taking into account the covariance between the individual and classroom administered working memory tasks. For verbal WM: $Z_H = 2.04$, $p < .05$; for visuospatial WM: $Z_H = 0.71$, $p = .475$

Acknowledgements

We express our thanks to Liselotte Koen, Klarien Lakerveld, Sanne van Leuken, Rosa Papenborg, and Cora van Raamsdonk for their contribution to the data collection of this study. Moreover, we would like to thank the teachers and students who participated in this study for their time and effort.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Adams, R., Finn, P., Moes, E., Flannery, K., & Rizzo, A. S. (2009). Distractibility in attention/deficit/hyperactivity disorder (ADHD): The virtual reality classroom. *Child Neuropsychology*, *15*(2), 120–135. doi:10.1080/09297040802169077

Adams, Z. W., Roberts, W. M., Milich, R., & Fillmore, M. T. (2011). Does response variability predict distractibility among adults with attention-deficit/hyperactivity disorder? *Psychological Assessment*, *23*, 427–436. doi:10.1037/a0022112

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*, 20–29. doi:10.1016/j.jecp.2009.11.003

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, *77*, 1698–1716. doi:10.1111/j.1467-8624.2006.00968.x

Arnett, A. B., Pennington, B. F., Friend, A., Willcutt, E. G., Byrne, B., Samuelsson, S., & Olson, R. K. (2013). The SWAN captures variance at the negative and positive ends of the ADHD symptom dimension. *Journal of Attention Disorders*, *17*, 152–162. doi:10.1177/1087054711427399

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*, 224–237. doi:10.1037/0096-3445.130.2.224

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, *8*, 129–148. doi:10.1037/1082-989X.8.2.129

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation, Vol. 8* (pp. 47–89). New York: Academic Press.

Barkley, R. A. (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *Journal of Abnormal Child Psychology*, *19*(2), 149–178. doi:10.1007/BF00909976

Borella, E., & de Ribaupierre, A. (2014). The role of working memory, inhibition, and processing speed in text comprehension in children. *Learning and Individual Differences*, *34*, 86–92. doi:10.1016/j.lindif.2014.05.001

Brydges, C. R., Anderson, M., Reid, C. L., & Fox, A. M. (2013). Maturation of cognitive control: Delineating response inhibition and interference suppression. *PLoS ONE*, *8*(7), e69826. doi:10.1371/journal.pone.0069826

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, *114*, 542–551. doi:10.1037/0033-2909.114.3.542

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Laure, M. C., Dawson, D. R., … Channon, S. (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society*, *12*, 194–209. doi:10.1017/S1355617706060310

Cassuto, H., Ben-Simon, A., & Berger, I. (2013). Using environmental distractors in the diagnosis of ADHD. *Frontiers in Human Neuroscience*, *7*, 805. doi:10.3389/fnhum.2013.00805

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*(4), 181–197. doi:10.1023/B:NERV.0000009483.91468.fb

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*(3), 217–227. doi:10.1016/j.acn.2005.12.002

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, *162*, 12–26. doi:10.1016/j.cognition.2017.01.014

De Weerdt, F., DeSoete, A., & Roeyers, H. (2013). Behavioral inhibition in children with learning disabilities. *Research in Developmental Disabilities*, *34*, 1998–2007. doi:10.1016/j.ridd.2013.02.020

Dolcos, F., Miller, B., Kragel, P., Jha, A., & McCarthy, G. (2007). Regional brain differences in the effect of distraction during the delay interval of a working memory task. *Brain Research*, *1152*, 171–181. doi:10.1016/j.brainres.2007.03.059

Feenstra, H., Kamphuis, F., Kleintjes, F., & Krom, R. (2010). *Wetenschappelijke verantwoording begrijpend lezen voor groep 3 tot en met 6 [Scientific account reading comprehension for grades 1 through 4]*. Arnhem: Cito.

Friso-van den Bos, I., Van der Ven, S. H. G., Kroesbergen, E. H., & Van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, *10*, 29–44. doi:10.1016/j.edurev.2013.05.003

Gathercole, S. E., Alloway, T. P., Kirkwood, H. J., Elliott, J. G., Holmes, J., & Hilton, K. A. (2008). Attentional and executive function behaviours in children with poor working memory. *Learning and Individual Differences*, *18*, 214–223. doi:10.1016/j.lindif.2007.10.003

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology*, *33*, 277–299. doi:10.1080/87565640801982361

Geffner, D., Lucker, J. R., & Koch, W. (1996). Evaluation of auditory discrimination in children with ADD and without ADD. *Child Psychiatry and Human Development*, *26*, 169–180. doi:10.1007/BF02353358

Gerst, E. H., Cirino, P. T., Fletcher, J. M., & Yoshida, H. (2017). Cognitive and behavioral rating measures of executive function as predictors of academic outcomes in children. *Child Neuropsychology*, *23*, 381–407. doi:10.1080/09297049.2015.1120860

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *ZDM Mathematics Education*, *47*, 771–782. doi:10.1007/s11858-014-0659-y

Glass, E., Sachse, S., & von Suchodoletz, W. (2008). Development of auditory sensory memory from 2 to 6 years: An MMN study. *Journal of Neural Transmission*, *115*, 1221–1229. doi:10.1007/s00702-008-0088-6

Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory & Cognition*, *30*, 447–455. doi:10.3758/BF03194945

Helps, S. K., Bamford, S., Sonuga-Barke, E. J., & Söderlund, G. B. (2014). Different effects of adding white noise on cognitive performance of sub-, normal and super-attentive school children. *PLoS One*, *9*(11), e112768. doi:10.1371/journal.pone.0112768

Hoerger, M. (2013). Z_H: An updated version of Steiger's Z and web-based calculator for testing the statistical significance of the difference between dependent correlations. Retrieved from http://www.psychmike.com/dependent_correlations.php

Janssen, J., Scheltens, F., & Kraemer, J. (2005). *Leerling- en onderwijsvolgsysteem. Rekenen-wiskunde. Handleiding [Student and education monitoring system. mathematics. teachers guide]*. Arnhem, The Netherlands: Cito.

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS rekenen-wiskunde voor groep 3 tot en met 8 [Scientific account of the mathematics test for grades 1–6]*. Arnhem, The Netherlands: Cito.

Jarvis, H. L., & Gathercole, S. E. (2003). Verbal and non-verbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educational and Child Psychology*, *20*, 123–140.

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*, 1–14.

Klatte, M., Bergström, K., & Lachmann, T. (2013). Does noise affect learning? A short review on noise effects on cognitive performance in children. *Frontiers in Psychology*, *4*, 578. doi:10.3389/fpsyg.2013.00578

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., … Westerberg, H. (2005). Computerized training of working memory in children with ADHD—A randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*, 177–186. doi:10.1097/00004583-200502000-00010

Kraha, A., Turner, H., Nimon, K., Zientek, L. R., & Henson, R. K. (2012). Tools to support interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology*, *3*, 44. doi:10.3389/fpsyg.2012.00044

Kuntsi, J., Andreou, P., Ma, J., Börger, N. A., & Van der Meere, J. J. (2005). Testing assumptions for endophenotype studies in ADHD: Reliability and validity of tasks in a general population sample. *BMC Psychiatry*, *5*, 40. doi:10.1186/1471-244X-5-40

Laski, E. V., & Dulaney, A. (2015). When prior knowledge interferes, inhibitory control matters for learning: The case of numerical magnitude representations. *Journal of Educational Psychology*, *107*, 1035–1050. doi:10.1037/edu0000034

Lee, K., & Lee, H. W. (2019). Inhibition and mathematical performance: Poorly correlated, poorly measured, or poorly matched? *Child Development Perspectives*, *13*, 28–33. doi:10.1111/cdep.12304

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14. doi:10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. doi:10.1006/cogp.1999.0734

Nath, S., & Szücs, D. (2014). Construction play and cognitive skills associated with the development of mathematical abilities in 7-year-old children. *Learning and Instruction*, *32*, 73–80. doi:10.1016/j.learninstruc.2014.01.006

Neguţ, A., Jurma, A. M., & David, D. (2017). Virtual-reality-based attention assessment of ADHD: ClinicaVR: classroom-CPT versus a traditional continuous performance test. *Child Neuropsychology*, *23*(6), 692–712. doi:10.1080/09297049.2016.1186617

Peng, P., Namkung, N., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, *108*, 455–473. doi:10.1037/edu0000079

Pham, A. V., & Hasson, R. M. (2014). Verbal and visuospatial working memory as predictors of children's reading ability. *Archives of Clinical Neuropsychology*, *29*, 467–477. doi:10.1093/arclin/acu024

Savage, R., Lavers, N., & Pillay, V. (2007). Working memory and reading difficulties: What we know and what we don't know about the relationship. *Educational Psychology Review*, *19*, 185–221. doi:10.1007/s10648-006-9024-1

Seigneuric, A., Ehrlich, M. F., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing*, *13*, 81–103. doi:10.1023/A:1008088230941

Shield, B. M., & Dockrell, J. E. (2003). The effects of noise on children at school: A review. *Building Acoustics*, *10*(2), 97–116. doi:10.1260/135101003768965960.

Shield, B., Greenland, E., & Dockrell, J. (2010). Noise in open plan classrooms in primary schools: A review. *Noise and Health*, *12*(49), 225. Retrieved from http://www.noiseandhealth.org/text.asp?2010/12/49/225/70501

Söderlund, G. B. W., Sikström, S., Loftesnes, J. M., & Sonuga-Barke, E. J. (2010). The effects of background white noise on memory performance in inattentive school children. *Behavioral and Brain Functions*, *6*, 55. doi:10.1186/1744-9081-6-55

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251. doi:10.1037/0033-2909.87.2.245

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. doi:10.1044/2015_JSLHR-H-14-0204

Swanson, H. L. (2017). Verbal and visual-spatial working memory: What develops over a life span? *Developmental Psychology*, *53*, 971–995. doi:10.1037/dev0000291

Swanson, J. M., Schuck, S., Porter, M. M., Carlson, C., Hartman, C. A., Sergeant, J. A., … Wigal, T. (2012). Categorical and dimensional definitions and evaluations of symptoms of ADHD: History of the SNAP and the SWAN rating scales. *The International Journal of Educational and Psychological Assessment*, *10*, 51–69. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4618695/.

Tsujimoto, S., Kuwajima, M., & Sawaguchi, T. (2007). Developmental fractionation of working memory and response inhibition during childhood. *Experimental Psychology*, *54*, 30–37. doi:10.1027/1618-3169.54.1.30

Valiente, C., Swanson, J., & Eisenberg, N. (2012). Linking students' emotions and academic achievement: When and why emotions matter. *Child Development Perspectives*, *6*, 129–135. doi:10.1111/j.1750-8606.2011.00192.x

Van der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J., & Jolles, J. (2008). A large-scale cross-sectional and longitudinal study into the ecological validity of neuropsychological test measures in neurologically intact people. *Archives of Clinical Neuropsychology*, *23*(7-8), 787–800. doi:10.1016/j.acn.2008.09.002

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Jolani, S., & Van Luit, J. E. H. (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

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Prast, E. J., & Van Luit, J. E. H. (2015a). 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., & Van Luit, J. E. H. (2015b). Verbal and visual-spatial working memory and mathematical ability in different domains throughout primary school. *Memory & Cognition*, *43*, 367–378. doi:10.3758/s13421-014-0480-4

Weekers, A., Groenen, I., Kleintjes, F., & Feenstra, H. (2011). *Wetenschappelijke verantwoording papieren toetsen Begrijpend Lezen voor groep 7 en 8 [Scientific justification for the paper and pencil tests for reading comprehension in grades 5 and 6]*. Arnhem, Netherlands: Cito.

Wu, S. S., Chen, L., Battista, C., Smith Watts, A. K., Willcutt, E. G., & Menon, V. (2017). Distinct influences of affective and cognitive factors on children's non-verbal and verbal mathematical abilities. *Cognition*, *166*, 118–129. doi:10.1016/j.cognition.2017.05.016