Ranking Working Memory Ability: Analyzing the Variance of Results through Inhibition

K. Wagner-van der Tol

Master's thesis

Utrecht University

Master's programme in Clinical Child, Family and Education Studies

Student	Koosje (K.) van der Tol
Student number	6521339
First assessor	Eva van de Weijer - Bergsma
Second assessor	Marije Stolte
Date of submission	29 May 2020
Version	Whole thesis v4.03
Word count	Abstract: 244
	Thesis: 4495

Abstract

Working memory (WM) skills are an important predictor of academic achievement. WM may be aided by good response inhibition, which helps a child focus on a task. To ensure high task purity on WM testscores, these tests are traditionally administered in an individual setting. However, it was previously found that WM tests administered in a classroom are a more accurate reflection of a child's real, everyday, capacities. In this paper two WM tasks: the Lion game and the Monkey game, are analysed to see if performing these tasks in the classroom would be a good way to broaden the variety of results, and thus assess differences between participants more accurately.

Data were collected in children aged 7-12 ($N = 99$). To assess WM, the tasks were completed in both an individual and a classroom setting. A Go/No-go task was completed to assess response inhibition. The variety of results was not found to be significantly different between settings. In a Wilcoxon signed rank-test, it was found that the rankings on the WM tasks were significantly different in the different settings. These results indicate that these tasks give a different result when measured in an individual setting, compared to a classroom setting. Response inhibition further did not moderate the relation between setting and ranking differences on the WM tasks. This study found that testing in a classroom situation is useful, if the aim of the assessment is ranking children on their WM ability in real life.

Ranking Working Memory Ability: Analyzing the Variance of Results through Inhibition

Working memory (WM) is the ability to simultaneously store and manipulate information to perform complex cognitive tasks (Friso-van den Bos & Van de Weijer-Bergsma, 2020). WM is one of the core executive functions (EF): the processes needed when a person has to concentrate and pay attention, instead of acting automatically (Diamond, 2013). WM is considered to be an important predictor of academic achievement and behavioral functioning (Alloway & Alloway, 2010; Cortés Pascual, Moyano Muñoz, & Quílez Robres, 2019; Follmer, 2018). Children with better WM skills have an improved ability to integrate new information with information they have previously received (Van de Weijer-Bergsma, Kroesbergen, Jolani, & Van Luit, 2016). They are also better at choosing strategies and carrying them out, because they can more easily break down the steps of a task (Alloway, 2006). Additionally, children with weak WM skills are less proactive, have lower self-monitoring skills and have difficulty organising and concentrating (Alloway, Gathercole, Kirkwood, & Elliott, 2009). People who, as a child, have better WM skills, will generally develop higher reading achievement, mathematical achievement and cognitive flexibility

(Blackwell, Cepeda, & Munakata, 2009; Bull, Andrews Espy, & Wiebe 2008; Nevo & Breznitz, 2011). Consequently, it is important to adequately measure WM skills, because the better development can be predicted, the better children with weaker WM skills can be supported in developing their academic and behavioral skills.

Baddeley and Hitch (1974) were the first to emphasize the functional importance of WM, as opposed to simply storing short-term memories. They referred to WM as a system comprising multiple components. The central executive - an attentional controller that manipulates information held by the short-term memory - with two subsystems that aid it: the phonological loop for acoustic and verbal information, and the visuospatial sketchpad for visual and spatial information (Baddeley & Hitch, 1974). The central executive determines whether the phonological loop and/or the visuospatial sketchpad are used to process this information (Baddeley, 2001). The phonological loop has two components: a phonological store and an articulatory rehearsal system (Baddeley & Hitch, 1974). Traces within the store are assumed to decay over a period of two seconds unless refreshed by rehearsal (Baddeley & Hitch, 1974). Baddeley (2001) states that the visuospatial sketchpad temporarily maintains and manipulates visuospatial information, thus playing an important role in spatial orientation and the solution of visuospatial problems. The visuospatial sketchpad can form a connection between visual and spatial information because it binds together the visual information with similar information from motor, tactile, or haptic senses (Baddeley, 2001). In describing the central executive, Baddeley (2001) emphasizes that some processes, like the influence of other skills, are not yet explained.

One of those other skills influencing WM is inhibition, another core EF (Barkley, 1997; Bull & Scerif, 2001; Diamond, 2013). When inhibitory control fails, the mind wanders and irrelevant information enters the WM workspace (Diamond, 2013). Diamond (2013) defines inhibition as the ability to control attention, thoughts, emotions, and behavior to override an internal or external lure and instead do what is needed. Many studies (e.g. Brydges et al., 2013; Diamond, 2013; MacLeod et al., 2003; Nigg, 2017) make a distinction between cognitive inhibition, whereby interfering information must be suppressed (Brydges et al., 2013), and response inhibition, whereby a prepotent response needs to be inhibited when a context cue is changed (Barkley, 1997; Nigg, 2000). The current study targets response inhibition, which may help a child focus on a task, instead of stopping its work after, for example, being distracted by another child making a noise (Klatte, Bergström, & Lachmann, 2013).

Experimental and correlational research have different purposes (Hedge, Powell, & Sumner, 2018). In experimental research, scientists manipulate a situation and compare the effects on participants (Neuman, 2016). Here, a reliable effect is an effect that most participants in a study show, that produces consistent effect sizes and that almost always replicates (Hedge, Powell, & Sumner, 2018). Correlational research is a non-experimental study that examines correlations in data and indirectly shows cause-effect relations (Neuman, 2016, p. 144). There, a reliable effect is an effect that consistently ranks individuals, to distinguish between them (Hedge, Powell, & Sumner, 2018). When testing WM, practitioners use an existing task to accurately measure individual differences in WM, to distinguish the WM ability of this child from that of others. The child needs to be ranked – they are performing correlational research. However, when a new WM task is developed, developers want the new task to be precise, to be consistent and to replicate the results of existing tasks – they are performing experimental research. Taking into account the definitions of accurate and precise (Heacock, 2009), the contrast is clear: For accurate results that are correct and without any mistakes, such as needed for predictive purposes, high variance of results is needed; for precise results that are exactly the same every time, such as needed for experimental purposes, low variance of results is needed (Helmenstine, 2019). Traditionally, developers have aimed for low variance, because that makes it easier to compare and reproduce results. Contradictory, practitioners need high variance, to make a clear diagnosis, prediction and distinction (Hedge, Powell, & Sumner, 2018; Rogosa, 1988). Summarizing, depending on the specific purpose of an assessment, an instrument sometimes requires high variability to be reliable, and sometimes low variability.

This reliability paradox results in different testing aims for experimental and diagnostical measures of WM (Hedge, Powell, & Sumner, 2018). Experimental researchers strive for task purity to precisely compare participants (Rabbitt, 2004). They assess a child's best performance with neuropsychological tests (Chaytor & Schmitter-Edgecombe, 2003). To ensure high task purity - make as little as possible demand on skills and functions not under study (Rabbitt, 2004) - they administer these tests in an individual setting where there are no outside influences (Barkley, 1991; Friso-van den Bos & Van de Weijer-Bergsma, 2020; Kanerva et al., 2019). Naturally, when limiting outside influence, it is less necessary for children to exhibit inhibitory control. In contrast, in diagnostical research it is useful to consider inhibitory control when assessing WM. Friso-van den Bos and Van de Weijer Bergsma (2020), and Kanerva et al. (2019) suggested that although children may perform better on a WM test in an individual setting, WM tests administered in a classroom setting are a more accurate reflection of a child's real, everyday, capacities. Friso-van den Bos and Van de Weijer Bergsma found that verbal WM functioning in a classroom setting was a more accurate predictor of academic performance than verbal WM functioning in an individual setting. Kanerva et al. found that the correlation between WM performance and cognitive achievement is higher in groups who perform tasks in noisy (classroom) settings. The Lion game, a visuospatial WM task (Van de Weijer-Bergsma, Kroesbergen, Prast, & Van Luit, 2015), and the Monkey game, a verbal WM task (Van de Weijer-Bergsma, et al., 2016), are especially designed to be administered in the classroom. Possibly, performing these tasks in the classroom gives a result that is just as precise as traditional testing, and more accurate because it takes everyday functioning into account, making it a good way to distinct between children's WM skills.

Summarizing, it is important to reliably assess WM, in order to predict and develop academic skills to prepare for later life (Alloway & Alloway, 2010; Cortés Pascual, Moyano Muñoz, & Quílez Robres, 2019; Follmer, 2018). Correlational research has a different purpose, and thus needs a different form of reliability, than experimental research (Hedge, Powell, and Sumner, 2018). To examine correlations and rank differences, a task must be accurate, with high variance within the results (Hedge, Powell, & Sumner, 2018). For the predictive purposes of correlational research, it is useful to test WM ability in a real life situation (Friso-van den Bos & Van de Weijer-Bergsma, 2020; Kanerva et al., 2019). Testing in a real life situation means that other abilities, like the ability to inhibit response, can have an influence on WM skills (Barkley, 1997; Bull & Scerif, 2001; Diamond, 2013). This influence may provide useful information (Friso-van den Bos & Van de Weijer-Bergsma, 2020; Kanerva et al., 2019).

The main aim of this study is to explore the differences between results on WM tasks in the individual and the classroom setting, to see whether administering a WM task in the classroom can help make a better distinction between children's WM skills in real life, because the classroom setting is a closer reflection of real life. Previous study has established both preciseness and accuracy of the Lion and the Monkey game (Friso-van den Bos & Van de Weijer-Bergsma, 2020; Van de Weijer-Bergsma et al., 2016, 2015). The variance within the results of these tasks has not been researched in previous studies. Neither have these studies paid attention to the ranking of children, and the influence setting has on this. Friso-van den Bos and Van de Weijer-Bergsma (2020) have studied whether inhibition moderates the relation between setting and WM, but have not linked this result to the ranking of children. Therefore, the research questions of this paper are:

1) Will performing WM tasks in a classroom setting result in higher variance, compared to an individual setting? I hypothesize that variance in WM testscores is higher in the classroom setting, than in the individual setting (Hedge, Powell, & Sumner, 2018).
2) Will performing WM tasks in a classroom setting influence the ranking of children on the performance of these tasks? I hypothesize that the rankings in the individual setting will differ from the rankings in the classroom setting (Friso-van den Bos & Van de Weijer-Bergsma, 2020).
3) Does inhibition influence the performance differences between settings? I hypothesize that children whose ranking decreases more, from the individual to the classroom setting, have lower inhibition skills (Friso-van den Bos & Van de Weijer-Bergsma, 2020; Kanerva et al., 2019).

Method

Participants

For this study, data from the study by Friso-van den Bosch and Van de Weijer-Bergsma (2020) were provided. In that study, a total of 108 children participated. Nine children were excluded from the analyses: Two children were absent from school during testing and, due to technical errors, data for seven children were missing on some of the necessary variables. Therefore, data from 99 children were analysed. Children were drafted at five Dutch primary schools by use of a convenient sample. Therefore, outcomes cannot be generalized to a larger population (Field, 2018) but will be used to test the likeliness of the hypotheses. Passive consent, whereby parents informed the principal when they did not want their child to participate, was received for all children. This study was approved by the ethics committee of the Faculty of Social and Behavioral Sciene, Utrecht University.

Procedure

All children were tested on WM in two settings: once in a classroom setting, in the presence of the teacher and classmates who were working independently, and once in an individual setting in a quiet room inside the school. To preclude testing order effects, it was randomized which setting was used first. All children first performed the Lion game and then the Monkey game. Children performed these tasks on a computer, instructions were given through audio recordings, using headphones. After completing the WM tasks in the individual setting, the child took part in an inhibition task. Time between sessions was one to two weeks.

Instruments

Working memory. Data for WM were collected using two existing tools for the assessment of WM in children aged six to twelve years: the Lion game and the Monkey game.

The Lion game is an online visuospatial WM task in which children are asked to recall the location of the last lion of a specific color in a 4x4 matrix. A lion of a particular color can appear and reappear at different locations, so a child has to remember and update the information during the test (Van de Weijer-Bergsma et al., 2015). The number of colored lions and its location to be remembered increases gradually, from one to five, during the task, which spans 20 items. The score of the child was the average proportion of correctly recalled locations. Previous research reveals a stable rank-ordering across test sessions, and a high test-retest reliability. High internal consistency was found and correlations with other WM tasks are adequate (Van de Weijer-Bergsma et al., 2015).

The Monkey game is an online verbal backwards recall WM task in which children listen to an audio recording containing a string of spoken, one-syllable words, and are asked to recall the words in reverse order (Van de Weijer-Bergsma et al., 2016). After four practice items, 20 test items across five levels follow. The number of words to recall increases gradually, from two to six words. Scoring is similar to that of the Lion game. Internal consistency of this task is high, correlations with other WM tasks are moderately strong. Test-retest reliability of this task was not assessed, but deemed to be adequate because of high stability over 1-year and 2-year periods (Van de Weijer-Bergsma et al., 2016).

Inhibition. Response inhibition was measured using a computerized Go/No-Go task, based on De Weerdt, Desoete and Roeyers (2013). In this task, children are asked to respond only to a specific stimulus as quickly as possible (by pressing a button). The proportion of no-go trials is .25. Children have to inhibit the prepotent response when the wrong stimulus appears. The task comprises six practice items, followed by three rounds of 40 items. The stimuli in each round are different. Each trial has a fixed intertrial interval of 2250 ms. To measure response inhibition, the number of times a child responds to a no-go trial is counted. Kuntsi, Andreou, Ma, Börger, and Van der Meere (2005) found moderate to high test-retest correlations on a similar task.

Analyses

Prior to analysis, data were checked on assumptions, including normality, linearity and homoscedasticity. In describing the results, Cohen's (1988) guidelines for effect sizes will be adhered to.

To answer the first research question, considering the variety of results, the difference in variance was assessed with Levene's test of homogeneity of variance (Gravetter &

Wallnau, 2009). Levene's test for homegeneity of variance is not available as a stand-alone test. Therefore, an independent t-test was conducted to perform this analysis. To accomplish that, the data on the classroom and individual setting were merged into one variable, for both tasks. A greater variance in the classroom results would indicate a higher scattering, and thus prove our hypothesis.

To answer the second research question, considering the rankings of performance in the WM tasks between the classroom and individual setting, the Wilcoxon signed-rank test was conducted for both the Lion game and the Monkey game. This test compares two sets of scores that are related in some way (Field, 2018). In this case, it compares the two sets of scores from both games that come from the same participants, in the two settings in which our WM tasks were carried out. It is based on ranking the differences between scores in the two settings. If the medians of differences equal 0, then the rankings of performance in the WM tasks of the classroom and individual setting are the same.

Next, to answer the research question whether the changes in ranking are moderated by inhibition, the commission score on the GoNoGo test was added as a covariate to the repeated measures ANCOVA (RMA) of the ranking of WM. The interaction effect between setting and inhibition is an indicator of the extent to which inhibition relates to a difference in scores.

Results

Within the 99 participants (52 boys) aged 7-12 (M = 9.49, SD = 0.79), one univariate outlier with a Z-score of more than $|z| = 3.29$ was found in the classroom Monkey game results. Furthermore, only for this participant the Mahalanobis distance exceeded the critical χ^2 value of 20.52 for $df = 5$ at $\alpha = .001$. Analysis with and without this outlier showed no difference in whether or not findings were significant. Reporting will continue with outliers included. Descriptive statistics are reported in Table 1.

Histograms of the Lion game show a negatively skewed distribution. On top of that, according to skewness values, only scores on the individual Monkey game are distributed normally (see Table 1). The Central Limit Theorem indicates the sample is large enough to assume normality for all dependent variables (Field, 2018, p. 63).

Table 1

Descriptive statistics of Lion and Monkey game and Inhibition task

	Range	*M*	*SD*	Skewness	Kurtosis
Lion game: individual	.51-.96	.80	.10	-.70*	.45
Lion game: classroom	.43-.99	.77	.11	-.77*	.47
Monkey game: individual	.44-.85	.61	.09	.21	-.54
Monkey game: classroom[1]	.12-.80	.59	.11	-.82*	1.65^
Inhibition: commission errors	0-15	5.67	3.47	.74^	.17

Note. ^significant positive, *significant negative; $SE_{skewness}$ = .24, $SE_{kurtosis}$ = .48
WM scores are proportions correct of recalled items. Inhibition: Commission errors is the number of times a child pressed, when it should have inhibited ; *n* = 99.

Pearson correlations between variables can be found in Table 2. Correlation analyses showed that both tasks, in either setting, significantly and positively correlated amongst each other. However, neither WM task correlated with the inhibition measure, in either setting.

Table 2

Pearson correlations between tasks of WM in two settings, and Inhibition

	1.	2.	3.	4.
1. Lion game: individual	-			
2. Lion game: classroom	.62**	-		
3. Monkey game: individual	.41**	.35**	-	
4. Monkey game: classroom	.29**	.33**	.51**	-
5. Inhibition: commission errors	-.06	-.02	-.20	-.12

***p* < .01

Levene's test of equality of variances was performed by converting the results of the individual setting and the classroom setting into a stacked variable, adding setting as an independent variable. This analysis showed that variances were not significantly different for the Lion Game: $F(1,196) = 1.24$, $p = .27$, or the Monkey game: $F(1,196) = 3.58$, $p = .06$. This means that performing WM tests in a classroom setting did not broaden the variety of results for either task, as illustrated by the graphs in Figure 1 and 2.

[1] The results for skewness and kurtosis on this variable are greatly influenced by the outlier. Skewness and kurtosis values are average with the outlier removed.

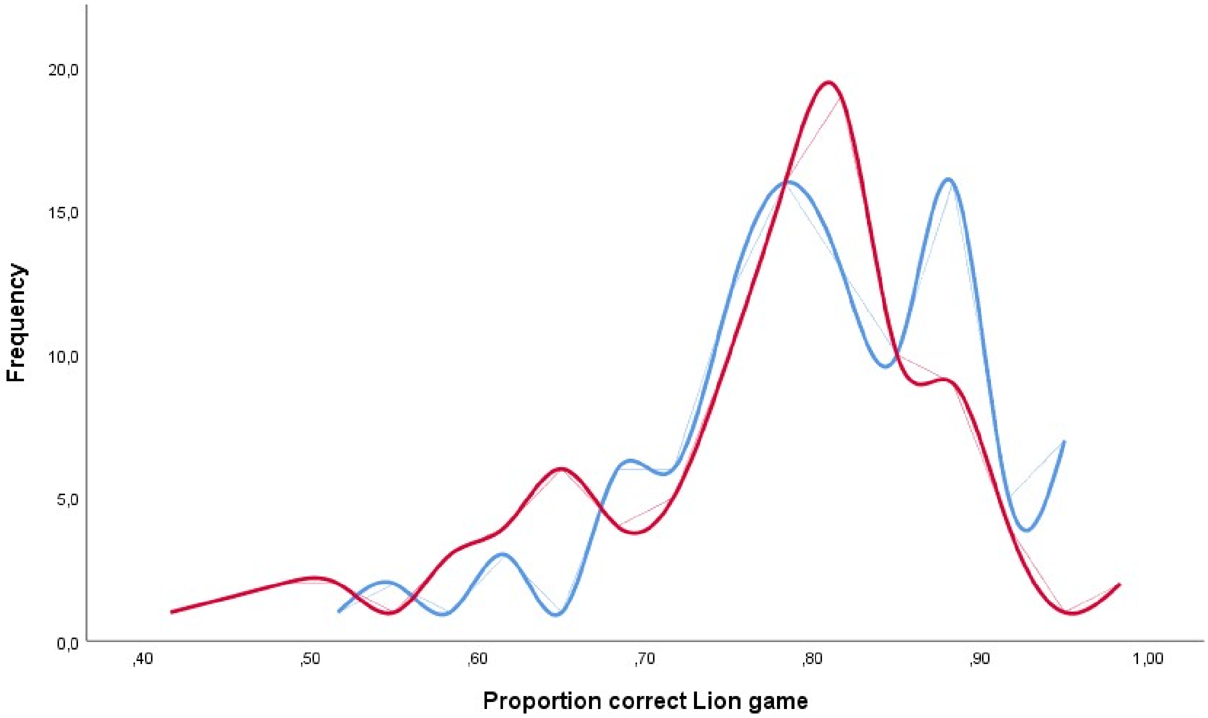

Figure 1. Frequency distribution of Proportion correct in the Lion game. The blue line depicts scores in the individual setting. The red line depicts scores in the classroom setting.

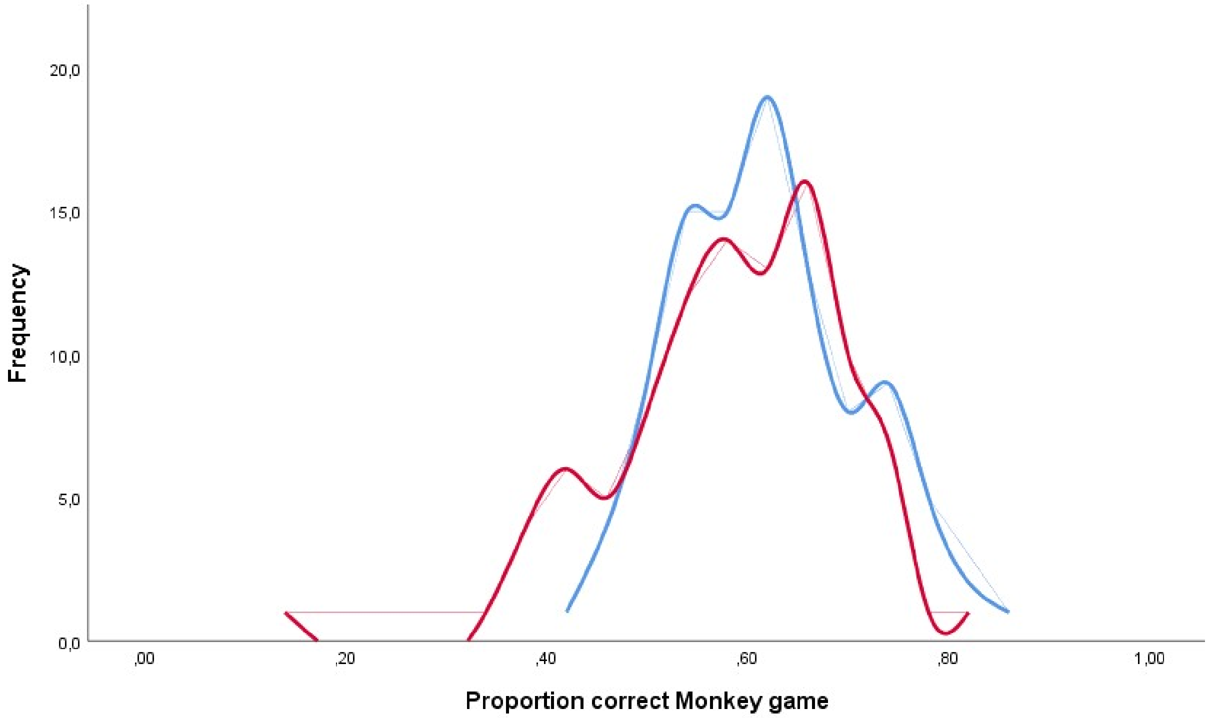

Figure 2. Frequency distribution of Proportion correct in the Monkey game. The blue line depicts scores in the individual setting. The red line depicts scores in the classroom setting.

The Wilcoxon signed-rank test of the Lion game resulted in a significant difference with a medium effect size, between the ranking in the individual setting, *Mdn* = .81, and the ranking in the classroom setting, *Mdn* = .79, $T = 1.64$ $z = -2.91$, $p < .01$, $r = -.29$. The Wilcoxon signed-rank test of the Monkey game resulted in a significant difference with a medium effect size, between the ranking in the individual setting, *Mdn* = .61, and the ranking in the classroom setting, *Mdn* = .59, $T = 1.74$, $z = -2.56$, $p = .01$, $r = -.26$. These results showed that the median of differences did not equal 0, and that in both games the rankings of the performance in the classroom and individual setting are not the same. This means that for both tasks, the setting influenced the ranking of children on the performance of the task.

Prior to RMA analyses, effects of testing order were investigated using an RMA for the ranking of both Lion and Monkey game, with classroom and individual setting as a within-subjects factor and testing order (individual first or classroom first) as a covariate. The analyses showed that testing order significantly affected ranking on the Lion game, $F(1,97) = 6.57$, $p = .01$, partial $\eta^2 = .06$, and on the Monkey game, $F(1,97) = 11.35$, $p < .01$, partial $\eta^2 < .10$. It was therefore decided to include testing order as a covariate in all RMA targeting the difference in ranking between individual and classroom testing.

The RMA of the ranking of the Lion game with testing order included as a covariate showed a medium, significant main effect of setting, $F(1,97) = 12.57$, $p < .001$, partial $\eta^2 = .11$, indicating that there are more children who perform better in the individual setting than in the classroom setting, than there are children who perform better in the classroom setting than in the individual setting. When both testing order and inhibition were included as a covariate, this result was maintained: $F(1,96) = 7.46$, $p = .01$, partial $\eta^2 = .07$. There was no significant effect of inhibition, $F(1,96) = 0.39$, $p = .54$, partial $\eta^2 < .01$, indicating that inhibition did not predict performance differences on this task. Likewise, there was no interaction effect between setting and inhibition, $F(1,96) = 0.51$, $p = .48$, partial $\eta^2 < .01$. This means that inhibition did not have an effect on performance differences between the individual and classroom setting, for this task.

The RMA of the ranking of the Monkey game with testing order included as a covariate showed a medium, significant main effect of setting, $F(1,97) = 7.56$, $p < .01$, partial $\eta^2 = .07$. When both testing order and inhibition were included as a covariate the effect size changed to small: $F(1,96) = 4.13$, $p = .04$, partial $\eta^2 = .04$. There was a small, significant main effect of inhibition, $F(1,96) = 5.52$, $p = .02$, partial $\eta^2 = .05$, indicating that higher inhibition scores did predict higher scores on this task. There was no interaction effect between setting and inhibition, $F(1,96) = 0.53$, $p = .47$, partial $\eta^2 < .01$. This means that inhibition did not have an effect on performance differences between the individual and classroom setting, for this task.

Discussion

In this study, the importance of reliably assessing WM was emphasized. Dependent on the goal of assessment, measures can be reliable when they produce either low or high variance (Hedge, Powell, & Sumner, 2018). The Lion and Monkey game are especially designed to be administered in the classroom, with the aim of producing high variance and high reliability in predicting future academic achievement (Van de Weijer-Bergsma et al., 2015; 2016). The aim of the current study was to explore the differences of administering these WM tasks in the classroom versus in an individual setting. It was examined whether performing tasks in a classroom would result in a better distinction between WM skills than in an individual setting. The influence of setting on the ranking of individuals was examined, and whether this was related to inhibition. It was theorised that because the environment is less sterile in a classroom, WM tests administered in a classroom setting would be a more accurate way of reflecting a child's everyday capacities, and predicting academic performance (Friso-van den Bos & Van de Weijer-Bergsma, 2020; Kanerva et al., 2019).

The first hypothesis was that administering the test in a classroom setting would lead to greater variance than administering it in the individual setting. Results showed that contrary to our hypothesis, performing the test in a classroom setting did not broaden the variance for either task. Second, this study compared the rankings of the two WM tasks in two different settings. It was hypothesized that the ranking in the individual setting would differ from the ranking in the classroom setting. This hypothesis was confirmed for both tasks. Last, inhibition was examined as a possible explanation of this difference in ranking. It was hypothesized that inhibition was the reason some children perform better or worse in classroom settings, predicting higher ranking WM scores in the classroom for children who are good at ignoring interference (Barkley, 1997; Bull & Scerif, 2001; Diamond, 2013). However, the results show that succesful response inhibition does not decrease the influence of setting, thus rejecting our hypothesis.

In conclusion, this study shows the usefulness of testing in a classroom situation, if the aim of the assessment is ranking children on their WM ability in real life. This study did not find that administering WM-tasks in the classroom makes ranking more accurate, because it does not broaden the variety of results. It found that setting influenced the ranking of children on performance of the WM tasks, meaning that some children fare better in a classroom setting, and some in an individual setting. In this study, inhibition does not moderate this influence, further research is necessary to determine what does. With the goal of assessing real life WM ability in mind, task impurity may not be an issue if the causes of the impurities are comparable to real life situations.

A few methodological explanations can be provided for these results, along with recommendations for future research. A limititation of Levene’s test is that in small samples, it will only pick up on big differences (Field, 2018, p. 258). For both games, the differences in variance may be too small to pick up on, indicating a need for a larger sample in future research. For the Lion game, the results are negatively skewed in both settings, indicating that a high number of participants scored in the higher regions. This game may be too easy for some participants, resulting in them reaching the tests performance limit, thus limiting our ability to draw conclusions about variance from it. For future research, increasing the difficulty of the Lion game by adding an extra level of six colored lions and its location to be remembered may be a valuable addition to this task. Also, cognitive inhibition (not paying attention to an interference) may be a better measure to explain the differences in performance and rank than response inhibition (not acting on an interference) (Brydges, Anderson, Reid, & Fox, 2013). Also, recently the Go/No-go task has been the center of criticism (Lee & Lee, 2019; Wessel, 2018). Lee and Lee (2019) indicate that the task is relatively easy, thereby providing little information on the gradience of inhibitory failure. Wessel (2017) found that the design of Go/No-go tasks differs greatly across studies. They concluded that a Go/No-go task only elicits a prepotent response in fast-paced designs (trial duration < 1500 ms) with rare no-go trials (proportion < .20). The Go/No-go task used in this study doesn’t meet either criterium. This possibly creates a problem with the precision of this task. For further study, it is advised to adapt the task so it meets the criteria, and test performance limit is improved.

For a more theoretical explanation, it seems that sometimes, some children are unfased by interference, because they are completely absorbed in their activity. On top of the methodological advice in the previous paragraph, a recommendation is made to explore the concept of flow, as presented by Nakamura and Csikszentmihalyi (2014). They describe flow as a state in which a person is with total involvement and motivation. For more research on the concept of flow in relation to WM, consult Brooks and Shell (2006).

References

Alloway, T. P. (2006). How does working memory work in the classroom? *Educational Research and Reviews, 1*, 134-139. Available online at http://www.academicjournals.org/ERR

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., Kirkwood, H., & Elliot, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development, 80*, 606-621. doi:10.1111/j.1467-8624.2009.01282.x

Baddeley A. (2001). Is working memory still working? *American Psychologist, 56*, 849-864. doi:10.1037/0003-066X.56.11.851

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47-90). New York, USA: Academic Press.

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

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*, 65-94. doi:10.1037/0033-2909.121.1.65

Blackwell, K. A., Cepeda, N. J., & Munakata, Y. (2009). When simple things are meaningful: Working memory strength predicts children's cognitive flexibility. *Journal of Experimental Child Psychology, 103*, 241-249. doi:10.1016/j.jecp.2009.01.002

Brooks, D. W., & Shell, D. F. (2006). Working memory, motivation, and teacher-initiated learning. *Journal of Science Education and Technology*, *15*, 17-30. doi:10.1007/s10956-006-0353-0

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*, 1-8. doi:10.1371/journal.pone.0069826

Bull, R., Andrews Espy, K., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*, 205-228. doi:10.1080/87565640801982312

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental neuropsychology*, *19*, 273-293. doi:10.1207/S15326942DN1903_3

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13*, 181-197. doi:1040-7308/03/1200-0181/0

Cohen, J. (1988*). Statistical power analysis for the behavioral sciences.* (2nd ed.). Hillsdale, USA: Lawrence Erlbaum Associates.

Cortés Pascual, A., Moyano Muñoz, N., & Quílez Robres, A. (2019). The relationship between executive functions and academic performance in primary education: Review and meta-analysis. *Frontiers in Psychology, 10*. doi:10.3389/fpsyg.2019.0158

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

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168. doi: 10.1146/annurev-psych-113011-143750

Field, A. (2018). *Discovering statistics using IBM-SPSS statistics*. Los Angeles, USA: Sage.

Friso-van den Bos, I., & Van de Weijer-Bergsma, E. (2020). Classroom versus individual working memory assessment: Predicting academic achievement and the role of attention and response inhibition. *Memory, 28*, 70-82. doi:10.1080/09658211.2019.1682170

Follmer, D. J. (2018). Executive function and reading comprehension: A meta-analytic review, *Educational Psychologist, 53*, 42-60. doi:10.1080/00461520.2017.1309295

Gravetter, F. J., & Wallnau, L. B. (2009). *Statistics for behavioral sciences, Custom edition*. England, Hampshire: Cengage Learning.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*, 1166-1186.

Heacock, P. (2009). Cambridge academic content dictionary. Cambridge: Cambridge University Press.

Helmenstine, A. M. (2019). *What is the difference between accuracy and precision?* Retrieved from: https://www.thoughtco.com/difference-between-accuracy-and-precision-609328 on 31-12-2019.

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*, 1224-1237. doi:10.1002/acp.3569

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.201300578

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

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

Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In: *Flow and the Foundations of Positive Psychology* (pp. 239-263). Springer, Dordrecht. doi:10.1007/978-94-017-9088-8_16

Neuman, W. L. (2014). *Understanding research*. England, London: Pearson.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. *Psychology of learning and motivation*, *43*, 163-215.

Nevo, E., & Breznitz, Z. (2011). Assessment of working memory components at 6 years of age as predictors of reading achievements a year later. *Journal of Experimental Child Psychology, 109*, 73-90. doi:10.1016/j.jecp.2010.09.010

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological bulletin*, *126*, 220. doi: 10.1037//0033-2909.126.2.220

Nigg, J. T. (2017). Annual research review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry 58*, 361-383. doi:10.1111/jcpp.12675

Rabbitt, P. (2004). Introduction: Methodologies and models in the study of executive function. In: P. Rabbitt. *Methodology of frontal and executive function* (pp. 9-45). USA, New York: Routledge.

Rogosa, D. (1988). Myths about longitudinal research. In: Schaie, K. W., Campbell, R.T., Meredith, W., & Rawlings, S. C. (Eds.), *Methodological issues in ageing research* (pp. 171-210). USA, New York: Springer.

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. (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

Wessel, J. R. (2018). Prepotent motor activity and inhibitory control demands in different variants of the go/no-go paradigm. *Psychophysiology*, *55*(3), e12871. doi:10.1111/psyp.12871