

Dirk, Judith; Schmiedek, Florian

Fluctuations in elementary school children's working memory performance in the school context

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:

formally and content revised edition of the original source in:

The journal of educational psychology 108 (2016) 5, S. 722-739



Bitte verwenden Sie beim Zitieren folgende URN /

Please use the following URN for citation:

urn:nbn:de:0111-pedocs-169471 - <http://nbn-resolving.org/urn:nbn:de:0111-pedocs-169471>

DOI: 10.1037/edu0000076 - <http://dx.doi.org/10.1037/edu0000076>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

©American Psychological Association, 2016. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal *The Journal of educational psychology*. Please do not copy or cite without author's permission. The final article is available, upon publication, at: [10.1037/edu0000076](https://doi.org/10.1037/edu0000076)

WORKING MEMORY FLUCTUATIONS IN THE SCHOOL CONTEXT

Fluctuations in Elementary School Children's Working Memory Performance in the School Context

Judith Dirk^{1,2} and Florian Schmiedek^{1,2,3}

¹ German Institute for International Educational Research (DIPF), Frankfurt a. M., Germany

² IDeA (Individual Development and Adaptive Education of Risk) Center, Frankfurt a. M.,
Germany

³ Goethe University, Frankfurt a. M., Germany

2014-11-07

Author Note

Judith Dirk, Center for Research on Education and Human Development, German Institute for International Educational Research and IDeA (Individual Development and Adaptive Education of Risk) Center, Frankfurt a. M., Germany; Florian Schmiedek, Center for Research on Education and Human Development, German Institute for International Educational Research and IDeA (Individual Development and Adaptive Education of Risk) Center and Department of Psychology and Sports Sciences, Goethe University, Frankfurt a. M., Germany.

Correspondence concerning this article should be addressed to Judith Dirk, Center for Research on Education and Human Development, German Institute for International Educational Research (DIPF), Frankfurt a. M., Germany.

Email: dirk@dipf.de

WORKING MEMORY FLUCTUATIONS IN THE SCHOOL CONTEXT

This research was part of the FLUX project at the Center for Individual Development and Adaptive Education of Children at Risk (IDeA) in Frankfurt, Germany, funded by the Hessian Initiative for the Development of Scientific and Economic Excellence (LOEWE).

We thank Heiko Rölke and the Technology Based Assessment Group at the German Institute for International Educational Research for developing and providing the software to assess children's performance and experiences via smartphones. We owe special thanks to Verena Diel, Tanja Könen, Jan Kühnhausen, Anja Leonhardt, Philipp Wieseemann, and a team of highly committed student assistants for their important roles in conducting the FLUX project.

Fluctuations in Elementary School Children's Working Memory Performance in the School

Context

Resubmission: 2015-07-14

Abstract

Children experience good and bad days in their performance. Although this phenomenon is well known to teachers, parents, and students it has not been investigated empirically. We examined whether children's working memory (WM) performance varies systematically from day to day, and to which extent fluctuations at faster timescales (i.e., occasions, moments) contribute to daily WM fluctuations in the school context. In an ambulatory assessment study, 110 third and fourth grade students (eight to eleven years old) completed WM tasks on smartphones three times a day in school and at home for four weeks. Results showed substantial within-person fluctuations in children's daily WM performance. Across task conditions, day-to-day, occasion-to-occasion, and moment-to-moment variability accounted for roughly the same extent of observed day-to-day variability with large individual differences in the amount of reliable fluctuations at the different timescales. Third graders were more variable than fourth graders at the faster timescales, more variable WM performance at all timescales was related to lower school achievement, more day-to-day variability was associated with lower fluid intelligence. These findings build the foundation for research on the antecedents and consequences of children's fluctuating cognitive resources. Theories about cognitive development and learning should consider performance fluctuations across and within days to understand the processes underlying long-term changes. Educational practice may be informed by the substantial WM fluctuations at all timescales and adopt interventions that increase children's attentional focus and self-regulation.

Keywords: Working Memory, Daily variability, Children, School context, Ambulatory assessment

Fluctuations in Elementary School Children's Working Memory Performance in the School Context

Good and bad days in children's performance at school are phenomena well known to parents, teachers, and school-aged children themselves. These are days on which their performance is better or worse than their average or typical level of performance. It seems widely accepted that children's cognitive performance at school varies and that, in consequence, they have more or less difficulties in being attentive and performing well on cognitive tasks, tests, or exams. The empirical basis for systematic fluctuations in children's cognitive performance from day to day, however, is scarce. Fluctuations in cognitive performance may occur at different timescales (i.e., across days, across daily occasions separated by hours, and across moments separated by minutes or seconds) and contribute to observed day-to-day variability (Schmiedek, Lövdén, & Lindenberger, 2013). Fluctuations or within-person variability in cognitive resources such as working memory (WM) may be a central determinant of children's school achievement since WM has been demonstrated as an important between-person predictor of learning, intelligence, and achievement (Swanson & Alloway, 2012). This study therefore aims at testing whether children's cognitive performance as measured by WM tasks varies systematically in the school context, at quantifying the degree to which WM performance fluctuates at different timescale levels, and at exploring the relation of WM fluctuations to school achievement.

Day-to-Day Variability in Children's Cognitive Performance

Exceptionally little is known about children's day-to-day performance variability. Fluctuations in cognitive performance have been studied to some extent in the adult lifespan with a focus on differences between younger and older adults in measures of reaction times (Anstey, Dear, Christensen, & Jorm, 2005; Der & Deary, 2006; Fozard, Vercruyssen, Reynolds, Hancock, & Quilter, 1994; MacDonald, Hultsch, & Dixon, 2003). Mostly,

fluctuations in experimental settings and at shorter time scales such as from trial to trial within a block of trials or from moment to moment within a testing session have been considered. To a lesser extent, studies have also addressed fluctuations in cognitive performance across daily and weekly testing sessions (Allaire & Marsiske, 2005; Li, Aggen, Nesselrode, & Baltes, 2001; Rabbitt, Osman, Moore, & Stollery, 2001; Schmiedek et al., 2013; Sliwinski, Smyth, Hofer, & Stawski, 2006). The emerging picture reveals that across timescales cognitive performance fluctuates considerably in adults. Moreover, some studies suggest that performance fluctuations at different timescales represent a stable characteristic of individuals with some persons being generally (i.e., on different timescale levels) more variable than others (Rabbitt et al., 2001).

Most studies on children's performance fluctuations have been conducted in the tradition of the microgenetic approach (Siegler, 2006). In this tradition, studies have addressed how change in children's performance in tasks such as mathematical problem solving comes about. Children's performance fluctuations across repeated assessments of a task over days and weeks were found to be considerable and might reflect testing new strategies and behaviors (e.g., Rittle-Johnson & Siegler, 1999). These performance fluctuations seem to be particularly pronounced in individuals with moderate ability in a given domain, as compared to beginners, reflecting strategy exploration and learning processes. Few studies have investigated performance fluctuations at faster timescales, mostly in reaction time tasks in healthy children (Jensen, 1992; Li et al., 2004; Williams, Hultsch, Strauss, Hunter, & Tannock, 2005) and in children with learning disabilities (e.g., Castellanos et al., 2005; Geurts et al., 2008). These studies indicate that children show considerable performance fluctuations across trials or task blocks with younger and disabled children being more variable than older children and young adults. Taken together, we conclude that there is some first empirical evidence for variability in children's cognitive performance at different timescales but the findings are far from being conclusive.

Working Memory as a Central Resource for School Achievement

We focused on WM to investigate the importance of cognitive performance fluctuations for daily school achievement. This choice is primarily grounded in the importance of WM for higher cognitive abilities (i.e., fluid intelligence and reasoning, cf. Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002) and academic achievement (cf. Swanson & Alloway, 2012). Moreover, WM has been shown to vary on a daily basis in adolescents and adults before (Riediger, Wrzus, Schmiedek, Wagner, & Lindenberger, 2011; Schmiedek et al., 2013). We conceive of WM as the ability to maintain and process information simultaneously in a controlled manner (Baddeley & Hitch, 1994). The central mechanisms underlying WM include the building, maintaining, and updating of structural representations via dynamic bindings (Wilhelm, Hildebrandt, & Oberauer, 2013). These bindings temporarily relate informational input (e.g., numbers in a subtraction task) to places in a mental coordinate system. Binding new information that is outside the focus of attention necessitates the switching of attention (Oberauer, Süß, Wilhelm, & Sander, 2008). These mechanisms of binding and attention switching form the basis for solving diverse problems in the school and everyday life context and thus qualify WM as a fundamental cognitive resource. Beyond its theoretical significance, empirical evidence has long demonstrated that WM is particularly important for the acquisition of new capacities in different school subjects (e.g., Hitch, Towse, & Hutton, 2001). WM performance is related to performance in mathematics (Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013; Swanson, 2011) and reading (Loosli, Buschkuehl, Perrig, & Jaeggi, 2012). More generally, relationships have been demonstrated between WM capacity and learning of new competences (Anderson, 1982), language comprehension (Daneman & Merikle, 1996), and general academic attainment (Alloway et al., 2005). Further, WM is related to children's fluid and general intelligence (Giofrè, Mammarella, & Cornoldi, 2013; Hornung, Brunner, Reuter, & Martin, 2011).

Variability in Working Memory Performance

Despite the wealth of research, WM has mostly been studied as a between-person differences construct. Rarely have studies investigated whether WM performance fluctuates within persons from day to day or at faster timescales, and particularly studies on children's WM fluctuations are scarce. In the following, we review studies on WM variability and stability to derive hypotheses about children's WM fluctuations in the school context.

To date, few studies have reported WM fluctuations in younger and older adults (Brose, Schmiedek, Lövdén, Molenaar, & Lindenberger, 2010; Brose, Schmiedek et al., 2013; Brose, Lövdén, & Schmiedek, 2014; Lecerf, Ghisletta, & Jouffray, 2004; Riediger et al., 2011; Riediger et al., 2014; Robertson, Myerson, & Hale, 2006; Schmiedek et al., 2013; Sliwinski et al., 2006), and adolescents (Gasimova et al., 2014; Riediger et al., 2011, 2014). Studies on WM fluctuations with children are rare (Könen, Dirk, & Schmiedek, 2015) although there is increasing interest in the study of performance fluctuations in school-aged children (Pnevmatikos & Trikkalotis, 2013). Overall, the findings are difficult to compare given the use of different WM paradigms (e.g., classical span tasks, updating tasks, and n-back), the focus on different time scales (i.e., weeks, days, and minutes within a task block), and the different methodological approaches applied to study WM fluctuations. WM fluctuations have been studied by obtaining indices of within-person variability (e.g., the intra-individual SD), quantifying the amount of performance fluctuations (Lecerf et al., 2004; Robertson et al., 2006), identifying within-person couplings in multilevel models (e.g. Brose et al., 2012; Sliwinski et al., 2006), and by applying complex variance decomposition (Schmiedek et al., 2013) and dynamical systems analyses (Gasimova et al., 2014). The majority of studies have addresses the coupling of WM variability with socio-emotional variables such as affect (Brose et al., 2012; 2014), stress (Sliwinski et al., 2006), and motivation (Brose et al., 2010; 2012; Riediger et al., 2011), and mostly focused on adult age differences from day to day. Riediger and colleagues (2011) have extended this research to

adolescents and found individual's daily WM performance to be systematically related to their mood states.

The emerging picture reveals that there is substantial within-person variability in adults' and adolescents' WM performance from day to day that can be considered systematic given the significant within-person couplings with affect, stress, motivation, and psychological as well as physiological arousal. Schmiedek and colleagues (2013) addressed the question of reliability of daily WM fluctuations directly and found that true day-to-day performance fluctuations exist in younger and older adults' WM performance but are largely determined by fluctuations at faster timescales. That WM fluctuations are systematic has also been demonstrated by Lecerf and colleagues (2004), who studied young adults' WM fluctuations in four different visuo-spatial WM span tasks. The authors found shared variance among different indices of WM variability and thus argued that WM fluctuations cannot be considered as random error.

The only study we know that addressed daily WM fluctuations in school-aged children found WM performance to be higher in children aged eight to eleven years on days when they reported having slept good and not substantially more or less than usual (Könen et al., 2015). However, there is growing interest in children's performance fluctuations in the school context. For example, Pnevmatikos and Trikkaliotis (2013) studied inhibitory control in children aged eight to twelve and found their cognitive performance to vary substantially in simulated classroom settings after children experienced negative emotions. In this study, performance variability at the timescale of minutes was comparable in size to cross-sectional age differences of two years. As the authors argue, this means that children's performance can vary to a degree that makes students in sixth grade appear as students in fourth grade concerning their cognitive performance.

The finding of substantial within-person fluctuations in WM at different timescales and in individuals of all age groups might be surprising given the high stability of WM

measures (cf. Conway et al., 2005). For example, in adults, test-retest correlations of .67 to .87 have been reported for classical span tasks over months (Klein & Fiss, 1999), weeks (Friedman & Miyake, 2004), and minutes (Turley-Ames & Whitfield, 2003). In children aged nine to eleven years, Hitch and colleagues (2001) reported WM performance in a reading span task to be rather stable over a year (.71), whereas there was less stability over a year in an operation span task (.56). Similarly, in adolescents, Englund, Decker, Woodlief, and DiStefano (2014) reported on average high stability (.83) for a newly developed WM assessment battery over two weeks, but for single WM tasks stability ranged from .49 to .88.

The putative discrepancy between high test-retest stability in WM performance on the one hand and the findings of substantial and systematic fluctuations in WM performance on different timescales can be explained by referring to the concepts underlying classical reliability measures. Within the framework of Classical Test Theory (Nunnally & Bernstein, 1994), considering correlations between repeated measures any within-person variability is treated as error. However, although WM fluctuations are likely transient in nature (i.e., they do not follow a systematic pattern; for an exception see Gasimova et al., 2014), they do not necessarily reflect random error. As Nesselroade and Featherman (1997) pointed out, low test-retest correlations may arise from poor reliability of the measures, substantial within-person variability, or both.

Taken together, substantial and systematic daily fluctuations have been reported for adults' and adolescents' WM performance, and there is first evidence that these findings hold also for school-aged children. Despite this first evidence, it is not clear whether children show reliable fluctuations in their WM performance across different timescales. However, empirical evidence for reliable performance fluctuations at different timescales is indispensable in order to investigate mechanisms underlying such within-person fluctuations, and to understand the processes that relate children's WM performance to their school achievement. This goes along with calls by cognitive and educational researchers for studying cognitive within-person

processes (Molenaar, 2004; Schmitz, 2006) in natural settings as compared to the laboratory (Neisser & Winograd, 1988). Only recently, researchers have started addressing this call and applied a process-analytic within-person perspective to study learning and achievement (e.g., Nagengast, Trautwein, Kelava, & Lüdtke, 2013; Schmitz & Wiese, 2006; Tsai, Kunter, Lüdtke, Trautwein, & Ryan, 2008) and in the field of ambulatory assessment in which researchers focus on studying behavior in real-life settings using new technologies such as smartphones and tablets (Ebner-Priemer, Kubiak, & Pawlik, 2009).

In sum, the first empirical evidence of systematic daily WM fluctuations speaks for the importance of further studying children's day-to-day WM performance. Given first evidence for reliable within-person variation in adults' WM performance across and within days and the importance of WM for children's school achievement, there is a need to study children's cognitive performance in the school context and to decompose the observed daily performance fluctuations at the different timescales.

Decomposing Day-to-Day Working Memory Fluctuations

To investigate whether systematic day-to-day fluctuations exist in children's daily WM performance in the school context, three essential steps need to be undertaken. First, in order to investigate whether within-person fluctuations are not only task-specific but generalize to WM in general, it is indispensable to measure performance in several tasks and to examine to what degree performance fluctuations are present across tasks. Second, to study true day-to-day performance fluctuations, one needs to take variation at faster timescales into account (cf. Schmiedek et al., 2013). The level of cognitive performance on a given day is typically measured by the average performance on a certain number of trials or blocks of trials that might be distributed across several occasions within a day. The observed day-to-day variability, then, is a combination of a true day-to-day variance component (i.e., variance of mean performance across days around the statistically expected mean value of performance) and contributions of variance components at the level of occasions (i.e., variance of mean

performance across days due to a sample of a limited number of occasions throughout the day from a distribution of occasion-to-occasion variability), and at the level of moments (i.e., sample variance of mean performance across days due to random draws of trials or task blocks from a distribution of trial-to-trial or block-to-block variability). Since lower-level variance components are reduced, but not eliminated, by means of aggregation (cf. Rabbitt et al., 2001; Schmiedek et al., 2013), it is necessary to decompose variation at different timescales to obtain an estimate of systematic fluctuations from day to day. This decomposition can be achieved by applying the following relation of observed and true day-to-day variances to the data

$$\sigma^2_{days(observed)} = \sigma^2_{days(true)} + (\sigma^2_{occasions} / n_{occasions}) + (\sigma^2_{blocks} / (n_{blocks} * n_{occasions})) \quad (1)$$

with $n_{occasions}$ being the total number of occasions and n_{blocks} the total number of blocks per occasion, the latter reflecting moment-to-moment variability. As a consequence, the same amount of observed day-to-day variability (i.e., variations of daily performance averaged across occasions and moments) can be due to different combinations of the three variance components. To distinguish between them, repeated assessments of performance across several daily occasions and several moments within each occasion are necessary. The estimated moment-to-moment variance component contains trial-to-trial variability as well as systematic performance fluctuations across blocks. Given some evidence from studies with reaction time paradigms that trial-to-trial variability appears to be increased in younger children as compared to older children and young adults (e.g., Williams et al., 2005), one may expect that estimates of moment-to-moment variability will be comparatively high in children as old as eight to eleven years as well. However, the extent to which findings from previous studies, mainly investigating reaction time paradigms, generalize to WM tasks is unclear.

Moreover, moment-to-moment variability may be influenced by different factors than trial-to-trial variability, all of which might also differ between children. In sum, there is a need to disentangle performance fluctuations at different, hierarchically nested, timescales, as correlates may differ by temporal resolution and between individuals.

Third, given the relevance of WM performance for academic achievement, it is indispensable to assess WM in its natural context. As ambulatory assessment research has demonstrated, assessing psychological phenomena in naturalistic settings via mobile devices increases ecological validity (i.e., the degree to which a study “accurately represents the conditions under which an effect occurs in the real world”, Reis, 2012, p. 6) while preserving relatively high experimental control (Hoppmann & Riediger, 2009). Therefore, this study assessed children’s WM performance via smartphones in school and after school where it is needed for achievement and learning. Thereby, we attempted to accurately represent the typical setting in which children learn and to increase the validity of findings pertaining to the role of WM resources for school achievement.

Summary of the Current Study

This study investigated (1) whether children’s WM performance varies systematically in the school context, (2) to which extent true performance fluctuations at the day-to-day, occasion-to-occasion, and moment-to-moment level contribute to the overall daily variability in children’s WM performance in the school context, and (3) whether WM performance fluctuations at different timescales relate to school achievement and fluid intelligence. Based on previous research demonstrating systematic relationships of WM fluctuations to affect, motivation, and stress in adults and adolescents (e.g., Brose et al., 2012; Riediger et al., 2011; Sliwinski et al., 2006), and first evidence for substantial and systematic fluctuations in children’s cognitive performance (Könen et al., 2015; Pnevmatikos & Trikkalotis, 2013), we expected children’s WM fluctuations to also vary systematically as indicated by a latent factor of WM fluctuations across different tasks within persons. Relying on work with adult samples

(cf. Rabbitt et al., 2001; Schmiedek et al., 2013) and first evidence for systematic cognitive performance fluctuations in school-aged children at timescales of days and minutes (Könen et al., 2015; Pnevmatikos & Trikkaliotis, 2013), we hypothesized that observed day-to-day fluctuations in children's WM in the school context can be decomposed into reliable day-to-day, occasion-to-occasion, and moment-to-moment fluctuations. Finally, drawing on previous research on the importance of WM for achievement in mathematics and reading (Friso-van den Bos et al., 2013; Hitch et al., 2001; Loosli et al., 2012) and its relationship to fluid intelligence (Giofrè et al., 2013), we explored the relation between WM fluctuations and school achievement and intelligence. We refrained from specifying precise hypotheses about this relation given that WM fluctuations have rarely been studied in children.

Method

This study was part of the FLUX project ("Assessment of Cognitive Performance FLUctuations in the School ConteXt") which aims at quantifying daily fluctuations in elementary school children's cognitive performance and identifying their antecedents and consequences in the school context. The study followed a multivariate, replicated, single-subject, repeated measures design (Nesselroade & Jones, 1991), and included an intensive longitudinal study phase with four daily assessments embedded in an intensive pre- and posttest protocol. Within the study, among others, cognitive performance, motivation, affect, sleep, and physical activity were assessed on a daily basis via smartphones and accelerometers. This paper reports mainly on the longitudinal study phase and focuses on WM performance which was assessed three times daily.

Participants

Participants were 110 students (45 girls) in Germany. Fifty of them were enrolled in third and 60 in fourth grade. All children attended the same elementary school in three third grade and four fourth grade classrooms. Their age ranged from eight to eleven years ($M = 9.88$, $SD = 0.61$). For 101 children, information regarding their social background could be

obtained from their parents. The vast majority of children were born in Germany (98%) and German was the native language of 77% of the children, which is common for a German city. We were assured by the class teachers that all children attended classes regularly, were all fluent in German, and could understand the instructions without problems.

Students' participation was voluntary and could be canceled anytime without giving reasons. Informed consent for participation was obtained from both the students and their parents; 71% of the target students participated. They were recruited from seven classes in one public elementary school in an average urban neighborhood. Only four children interrupted the study ahead of time indicating overall good compliance.

Procedure

In this study, we considered the three daily sessions of the intensive longitudinal study phase in which cognitive performance was assessed as well as the background measures obtained at pretest. The pretest and intensive training of the daily assessment battery were administered in six lessons (4.5 hours total, including about 3 hours for pretests) distributed over two weeks in the second term of the school year. All pretest assessments took place in the classroom in groups of up to 20 students. The pretest protocol included, among others, paper and pencil measures of school achievement and fluid intelligence, and a baseline session of a daily assessment session on smartphones, including measures of cognitive performance. Not all children attended all pretest assessments (e.g., due to illness), but we assured that no child missed the intensive training of the daily assessment battery. In the training session, a qualified research assistant extensively instructed students on how to operate the smartphone and the study application and demonstrated each task and question of the daily protocol. Each student received a smartphone for the duration of that lesson, practiced the tasks, and could ask questions. At the end of the session, after a short break, students completed one daily session as a baseline assessment. The smartphones (Dell Streak 5, with Android 2.2 operation system) were equipped with an application specifically

programmed for this project and were given to the students for the duration of the study. The application was programmed in a way that children could not exit it. Thereby we assured that other functions usually available on smartphones (e.g., access to the internet, GPS) were inaccessible.

The intensive longitudinal study phase lasted for 31 consecutive days and started one week after the pretesting. Smartphones rang at the beginning (8:50 am, Occasion 1) and the end of school (11:25 am, Occasion 2) as well as in the afternoon (around 3:00 pm, Occasion 3). School sessions took place during class. Sessions were available up to 60 min and lasted about 10 to 15 min. While school sessions were scheduled to fixed times for all children, afternoon and evening sessions could be scheduled individually within a time-window of +/- 2 hours. All sessions were carried out daily, including weekend days. Teachers and parents kept minutes of children's participation. As to be expected with intensive longitudinal protocols, there was missing data. In this study, missing data resulted, for example, from illness, exams, other obligations during testing times as well as from technical problems such as an empty battery, or smartphones left at home during school hours. For the WM tasks, on average, across tasks conditions and grades 65% of the maximum possible data were available (for details, see task description below). The children received money or a gift certificate for their participation. The local ethical review board approved the study.

Daily Measures

Working memory tasks. We presented two WM updating tasks with numerical and spatial content and two memory load conditions (Load 2 and Load 3) each (see Figure 1). Tasks were adapted versions of WM updating tasks used with children, adolescents, and adults before (Göthe, Esser, Gendt, & Kliegl, 2012; Riediger et al., 2011; Schmiedek et al., 2013). They were pretested in a study with 75 elementary school children (ten occasions). They were specifically designed for children and embedded in a child-appropriate story. The final version of the tasks was pretested with twelve third graders (15 occasions). Children

collected points with their performance and received a short feedback at the end of each session.

Numerical Working Memory. Children had to memorize and update two or three one-digit numbers. In each of two or three horizontally placed cells, one initial digit (0-9) appeared simultaneously for 3000 ms. In the Load 2 condition, after an ISI of 500 ms, a sequence of three updating operations was presented in the cells. In the Load 3 condition, a sequence of four updating operations appeared. The updating operations were additions and subtractions from -2 to +2. The total was never negative or above nine. The updating operations had to be applied to the memorized digits, and the results also had to be memorized. No cell was updated twice in a row. The presentation time for updating operations was 2750 ms, the ISI was 250ms. At the end of each block, the two or three end results had to be entered within a maximum of 20000 ms. In each of the three daily occasions, four blocks of both the Load 2 and the Load 3 condition were included, resulting in a total of 20 responses per occasion (i.e., 8 responses for Load 2, and 12 responses for Load 3). For analyses at the occasion level, accuracy scores were calculated by averaging across all responses of the four blocks, resulting in one performance score for each of the two load conditions per measurement occasion. For analyses at the block level, the mean accuracy of responses per block was obtained. The mean at the block and the occasion level were computed only if there was at least one answer per block to assure continuous task processing. This resulted in an exclusion of 3% of data available for the numerical WM updating task on average across load conditions and grades.

Spatial Working Memory. Children had to memorize and update positions of differently colored and shaped cartoon creatures presented in a 4 by 4 grid. Two or three cartoon creatures appeared simultaneously at distinct positions in the grid for 3000 ms. In the Load 2 condition, after an ISI of 500 ms, three updating operations were presented sequentially. In the Load 3 condition, a sequence of four updating operations was presented.

The updating operations were spatial shifts to adjacent positions indicated by arrows in colors corresponding to the cartoon creatures' color schemes. Arrows were positioned in the center of the grid. The updating operations had to be applied to the memorized positions of the corresponding cartoon creatures, and the results also had to be memorized. No creature's position was updated twice in a row. Intermediate and end positions were never doubly assigned. The presentation time for updating operations was 2500 ms, the ISI was 500 ms. At the end of each block, the two or three end results had to be entered within a maximum of 30000 ms. The total number of responses per occasion and the performance measures used in the analyses were identical to those in the numerical WM task. On average, across load conditions and grades, 2% of the data available for the spatial WM task were excluded.

Baseline Measures

Fluid intelligence. In order to relate our findings to previously reported results on the relationship between WM and fluid intelligence and to assess baseline performance, fluid intelligence was measured. It was assessed with the revised German version of the Culture Fair Intelligence Test (CFT-20-R; Weiss, 2006). Out of the entire sample, 107 children took the test. The sample Cronbach's alpha coefficient on the latent fluid intelligence score was .72.

School achievement. In order to relate our findings to previously reported results on the relationship between WM and school achievement, we administered two standardized tests of mathematics and reading. Due to time limits, two subtests of a German mathematical achievement test for third and fourth graders, respectively, were used to assess mathematical skills (DEMAT 3+; Roick, Göllitz, & Hasselhorn, 2004; DEMAT 4; Göllitz, Roick, & Hasselhorn, 2006). These two tests include computation problems (subtest on arithmetic) and word problems (subtest on written math problems). Out of the entire sample, 106 children took the mathematics test. The sample Cronbach's alpha coefficients on the manifest mathematics scores was .81, for both DEMAT3+ and DEMAT. Similarly, a standardized

German test of reading for elementary school students, including three subtests on word, sentence, and text comprehension, was administered to assess reading skills (Lenhard & Schneider, 2005). Out of the entire sample 96 children took the reading test. The sample Cronbach's alpha coefficient on the reading score was .96.

Baseline working memory. To assess baseline WM performance, one session of the same WM tasks as in the daily protocol were administered in the pretest. This resulted in a measure of WM performance based on four blocks of eight or 12 responses, respectively, for each load condition (2 vs. 3) and task content (numerical vs. spatial). The accuracy scores in the four task conditions were positively correlated with each other (see Table 1) and formed a latent WM construct that correlated positively with fluid intelligence ($r = .81, p < .05$), mathematics ($r = .88, p < .05$), and reading ($r = .66, p < .05$) performance at the construct level (for information about factor loadings and model fits see Appendix A, Table A1). This strong positive relationship of between-person differences in WM and fluid intelligence is in line with previous research (Süß et al., 2002) and confirms that the WM updating tasks applied in this project are appropriate measures of WM in children aged eight to eleven. The medium to high positive correlations between WM and measures of school achievement replicate previous results (e.g., for reading and WM: Swanson & Jerman, 2007; for mathematics and WM: Friso-van den Bos et al., 2013) and further confirm that the WM tasks applied in this study are comparable to other WM measures previously applied in educational research. The sample Cronbach's alpha coefficient on the WM score was .81, indicating good reliability.

Background questionnaire. Socio-demographic information was obtained from parents via a paper-based questionnaire that was distributed to and recollected from the students in school. Given the high number of migrants and foreigners in the region of Germany in which the study was conducted, this questionnaire was available upon request in

languages other than German (e.g., Turkish, Polish, and Russian) in order to increase completion rates. A total of 101 out of 110 parents completed the questionnaire.

Data Analysis

The repeated measures design resulted in daily WM data that were hierarchically structured with repeated measures (Level 1) nested within persons (Level 2). In order to establish the presence and relevance of daily WM fluctuations at different timescales, we followed a two-step approach. In a first step, we conducted two-level confirmatory factor analyses to determine whether children's WM fluctuations are systematic from occasion to occasion (i.e., not explainable in terms of fluctuations at faster timescales and/or measurement error). For that, we used robust maximum likelihood estimation (MLR; Mplus 7, Muthén & Muthén, 1998-2012) and dealt with missing data in a full information approach (FIML; Enders, 2010). In a second step, we fitted multilevel models separately to the WM data from each individual and each task condition to estimate individual variance components reflecting day-to-day, occasion-to-occasion, and moment-to-moment variance in WM. This was accomplished using maximum likelihood estimation with SAS PROC MIXED (SAS 9.3, SAS Institute Inc., 2002-2005). These analyses show, separately for each task condition, how the variability in performance can be partitioned into a day-to-day component, an occasion-to-occasion, and a moment-to-moment component. The day-to-day component captures systematic variations of performance across days, indicating the degree to which observed (i.e., total) day-to-day variability is due to performance being systematically higher or lower on different days. The same logic applies to the occasion-to-occasion and the moment-to-moment component (see Appendix B for a detailed description of the multilevel model). To avoid unreliable variance estimates, analyses in the second step were only conducted for children with more than 20 days of data for a given WM task condition resulting in variance component estimates for 84 children. In addition, one child demonstrated extreme variability from moment to moment in the spatial WM task Load 3 while showing average variability

from moment to moment on all other tasks. Based on detailed descriptive analyses, this case was excluded from further analyses resulting in a sample of 83 children for all analyses involving variance components. To explore individual differences in the variance components, follow-up analyses were conducted fitting multilevel models to this subsample (see Appendix B). In the remainder of this paper, if not stated differently, analyses were conducted with the entire sample of 110 children.

Results

Descriptive Analyses

A total of 100223 blocks of the WM tasks were completed across the four task conditions out of which 97928 were considered in the following analyses based on the criterion of continuous task processing. Based on this criterion, children completed an average of 245.5 blocks ($SD = 65.3$) of the numerical task with Load 2, 244.1 blocks ($SD = 66.9$) of the numerical task with Load 3, 253.5 blocks ($SD = 65.5$) of the spatial task with Load 2, and 252.6 blocks ($SD = 66.7$) of the spatial task with Load 3.

Table 2 shows descriptive statistics for all WM measures, separately for each task condition and occasion. Mean accuracies in WM ranged from .47 to .77, indicating that the influence of ceiling effects was low and guessing of responses was unlikely¹. Across task conditions, performance was somewhat better in the morning than at noon and in the afternoon ($M_{Occ1} = .68$, $M_{Occ2} = .61$, $M_{Occ3} = .65$; $F_{Occ1-Occ2}(1, 1189) = 149.93$, $p < .05$, $r = .33$; $F_{Occ1-Occ3}(1, 1189) = 107.36$, $p < .05$, $r = .29$; $F_{Occ2-Occ3}(1, 1189) = 0.09$, ns), and children on average performed better in the numerical compared to the spatial task ($M_{numerical} = .66$, $M_{spatial} = .63$; $F(1, 1189) = 45.76$, $p < .05$, $r = .19$), and in the Load 2 compared to the Load 3 condition ($M_{Load2} = .72$, $M_{Load3} = .57$; $F(1, 1189) = 1502.74$, $p < .05$, $r = .75$) reflecting effects of task difficulty. Fourth graders showed a higher WM performance than third graders ($M_4 = .73$, $M_3 = .53$; $F(1, 1189) = 225.56$, $p < .05$, $r = .40$).

The intraclass correlation (ICC; i.e., the portion of between-person variance over total variance) ranged from .32 to .54, indicating that overall variance was dominated by within-person fluctuations. Across task conditions and occasions, WM measures showed a substantial average intraindividual SD. The magnitude of the within-person occasion-to-occasion variability was found to average between 66% and 79% of the between-person variability.

Psychometric Properties of WM Fluctuations

Reliability. Several authors have recently argued that validity and reliability should be considered separately for within- and between-person measures (Cranford et al., 2006; Geldhof, Preacher, & Zhytur, 2014; Shrout & Lane, 2012). Accordingly, we considered the psychometric properties of the daily WM measures and assessed reliability in two ways, according to the two main questions of this study.

First, following our first question, we addressed the systematicity of WM fluctuations at the latent construct level by testing a two-level confirmatory one-factor model for the four WM task conditions (see Figure 2 and next section for details). Based on this model, we assessed reliability separately for the within- and between-person level following an approach suggested by Wilhelm and Schoebi (2007). Relating the proportion of latent variation to total variation on each level, we found internal consistencies of WM of .78/.97 on the within- and between- person level, respectively. The reliability of WM measures was also relatively robust across measurement occasions (Occasion 1: .79/.96, Occasion 2: .79/.98, Occasion 3: .77/.97).

Second, we considered the WM fluctuations across moments of each WM task since our second question addressed WM fluctuations in each of the four task conditions. Following suggestions by Cranford and Colleagues (2006), we conducted multilevel models decomposing the variance in each task condition into variance related to blocks, persons, time, and all two-way interactions of these factors, and estimated between-person reliability

and reliability of change (i.e., within-person reliability; see Cranford et al., 2006, pp. 924-925, formulas 2 and 5, respectively). Thereby, we obtained within-person reliability of .66/.70 for the numerical task conditions (Load 2/Load 3) and .67/.58 for the spatial task conditions (Load 2/Load 3). At the between-person level, reliability was 1.00 for the numerical and .99 for the spatial task conditions. These excellent reliabilities can be attributed to the fact that 91 repeated assessments together with four indicators for each task reduced the error term considerably. Taken together, we conclude that at the latent construct and at the task level, the WM measures administered in this study were reliable in assessing within-person fluctuations as well as between-person differences in WM performance.

Validity. The daily accuracy scores in the four task conditions were positively correlated with each other within persons (see Table 1). This indicates convergent validity across task conditions and lends evidence to the validity of daily WM fluctuations as assessed with the four updating tasks presented in this study.

WM Fluctuations in the School Context

To assess WM fluctuations in the school context, we tested for systematic within-person and between-person factors across the four WM task conditions (see Figure 2). The factors were well-defined with significant factor loadings within and between, implying systematic common variance on both levels². The within-person factor demonstrated that on occasions when children showed higher performance in one task condition, they also showed higher performance in the other task condition. Found systematic between-person differences were also apparent indicating that children who generally showed better performance in one task condition were also better in another task condition across study occasions. Thus, the tasks presented in the present study allow assessing both systematic within-person and between-person differences in children's WM performance in the school context.

WM Fluctuations at Different Timescales

Next, we decomposed children's WM performance at the different timescales considered in this study (i.e., days, occasion, and moments). Results are based on the data of 83 children for whom sufficient data was available to estimate variance components at the different timescales. The findings are summarized in Figure 3, in which the total size of the bars corresponds to the average amount of observed day-to-day variability (i.e., the variance of average performance across days). This variance is decomposed into a variance component of systematic day-to-day fluctuations (black), the contribution of occasion-to-occasion (light grey) and of moment-to-moment variability (dark grey) to observed day-to-day variability.

The average contribution of true day-to-day variability to observed day-to-day variability was highly reliable across task conditions. Across WM task conditions, day-to-day, occasion-to-occasion, and moment-to-moment variability accounted for roughly the same extent of observed day-to-day variability. Overall, this means that fluctuations from day to day, from occasion to occasion, and from moment to moment contribute to children's good and bad days of performance in the school context.

Exploring Individual Differences in WM Fluctuations and Their Relation to School Achievement and Fluid Intelligence

The preceding findings focused on average WM variability only. In contrast, the existing literature on adult cognitive performance fluctuations suggests that there might also be substantial individual differences in children's performance variability (e.g., Rabbitt et al., 2001). We thus explored (a) whether individual differences in the amount of children's WM variability can be observed, (b) to what degree children's WM variability is a stable individual differences construct concerning all task conditions and timescales, and (c) whether individual differences in WM variability relate to school achievement and fluid intelligence. Again, results are based on the data of 83 children for whom sufficient data was available to estimate variance components at the different timescales. Findings on individual differences in the amount of children's performance variability are summarized in Figure 4, in which each bar

refers to one child, and the total size of the bars corresponds to the individual amounts of observed day-to-day variability (i.e., the variance of average performance across days). This variance is decomposed into an individual variance component of systematic day-to-day (black), occasion-to-occasion (light grey), and moment-to-moment fluctuations (dark grey) all together contributing to observed day-to-day variability.

Figure 4 shows large individual differences in children's variability at the different timescales with some children's observed day-to-day variability being dominated by performance fluctuations across moments and occasions without a contribution of systematic day-to-day variability. To test whether these observed individual differences were stable across timescales, we conducted correlation analyses between the average manifest day-to-day, occasion-to-occasion, and moment-to-moment variance components of the four WM task conditions. All but the day-to-day and the moment-to-moment variance components showed statistically significant moderate positive correlations (day-to-day with occasion-to-occasion: $r = .33, p < .05$; day-to-day with moment-to-moment: $r = .07, ns$; occasion-to-occasion with moment-to-moment: $r = .32, p < .05$). This indicates that relatively small amounts of variance are shared between the different timescales. As a consequence, children who are more variable from day to day (i.e., whose performance deviates more strongly between today and tomorrow) are not necessarily the same children who differ in their WM performance within a day (e.g., who show another level of WM performance at the beginning as compared to the end of school). Similarly, children who are more variable from day to day and within a day are not necessarily the same children who demonstrate WM performance fluctuations from moment to moment (i.e., who show varying levels of WM performance within a couple of minutes in which they solve four blocks of a WM task).

Finally, we further explored individual differences in performance variability at the different timescales. First, we considered grade differences in WM fluctuations. With the exception of the spatial Load 3 condition, children in fourth grade were on average

significantly less variable in WM performance from occasion to occasion and from moment to moment than children in third grade (see Table 3)³. Day-to-day and occasion-to-occasion variance components were reliably different from zero, however, for both grades and all task conditions. These results were also valid once differences in performance levels between the seven classes were taken into account. Second, we investigated whether individual differences in the amount of WM fluctuations relate to school achievement in mathematics and reading, and to fluid intelligence. Children who were more variable in their WM performance, independent of timescale, also demonstrated significantly lower achievement in mathematics and reading (r ranging from $-.22$ to $-.35$, all $ps < .05$). Children who showed more variable WM performance from one day to the next, also obtained significantly lower scores in a test of fluid intelligence ($r = -.32$, $p < .05$, see Table 4).

Discussion

This study investigated whether children's cognitive performance varied systematically from day to day, from occasion to occasion, and from moment to moment in the school context. WM measures administered in this study were reliable and valid measures for detecting both within-person fluctuations as well as between-person differences in cognitive performance in the school context. Based on these WM measures, we identified substantial within-person fluctuations in elementary school children's daily WM performance that averaged about two thirds of between-person differences. These fluctuations could be decomposed into reliable day-to-day, occasion-to-occasion, and moment-to-moment fluctuations. Overall, children's true day-to-day fluctuations contributed to their overall day-to-day variability in performance to a similar degree as their performance fluctuations at faster timescale levels. However, there were large individual differences in the degree to which observed day-to-day variability was composed of fluctuations at the different timescale levels. Some children showed no systematic day-to-day variability in their WM performance. Finally, on average, we found children in fourth grade to be significantly less variable in their

WM performance from occasion to occasion and from moment to moment than children in third grade. Moreover, across timescales, more variable WM performance was related to lower performance in mathematics and reading, and more variable WM performance from one day to the next was associated with lower scores in a test of fluid intelligence. In sum, what appears to be a bad day in school seems to be a combination of bad moments, bad occasions, and, at least for some children, bad days.

Systematic Variation in Children's WM Performance in the School Context

The present study is the first to investigate how WM performance varies within children at different timescales in the school context. Over four weeks, in three daily occasions in school and in the afternoon, fluctuations in children's cognitive performance as measured by four WM task conditions accounted for between 48% and 65% of the overall variance as indicated by the ICC. Whereas variation between children indicates that cognitive performance differs from one child to another, variation within children indicates that cognitive performance fluctuates within children over time. Moreover, the amount of performance fluctuations was very similar across the four WM task conditions, although it was slightly higher in the spatial task with WM load 2. The fact that performance fluctuations correlated substantially within children across the four tasks excludes the possibility that these performance fluctuations are mainly due to measurement error. Trying to quantify the size of these performance fluctuations, we also considered the ratio of variation between children to variation within children as described by the between-person and within-person SDs (cf. Nesselrode & Salthouse, 2004). Following this approach, within-person fluctuations in children's WM performance amounted to between 66% and 79% of between-person differences. These findings are in line with previous results from studies on reaction time variability which reported that the magnitude of within-person fluctuations is about half that of between-person differences (Li, Aggen, et al., 2001; Nesselrode & Salthouse, 2004).

This first descriptive evidence for substantial fluctuations within children's cognitive performance over time was confirmed by a two-level confirmatory factor analysis which demonstrated that the observed performance fluctuations are systematic. On days when children's performance was higher on one task, it tended to also be higher on the other tasks. Thus, like adult's and adolescents' cognitive performance (Riediger et al., 2011; Schmiedek et al., 2013) children's WM performance is also not stable but shows substantial and systematic variation. To date, few studies have considered children's performance fluctuations (Könen et al., 2015; Pnevmatikos & Trikkaliotis, 2013; Siegler, 2006) and the majority of studies have been conducted in laboratory settings. The present findings confirm that WM performance as a fundamental cognitive capacity underlying the maintaining and updating of information as needed in various school subjects shows fluctuations in the school context. Moreover, we could demonstrate that these performance fluctuations can be measured in the school context using ambulatory assessment by making use of new technological developments (i.e., smartphones). These results generally confirm teachers' and parents' observations of good and bad days in children's performance and form the basis for further investigation of the factors potentially underlying these fluctuations in children's cognitive performance (e.g., Könen et al., 2015).

Decomposition of Children's WM Performance Fluctuations at Different Timescales

Taking a more detailed look at children's cognitive performance fluctuations, we decomposed children's WM fluctuations in day-to-day, occasion-to-occasion, and moment-to-moment variance components. Day-to-day fluctuations contributed to overall day-to-day variability in children's WM performance to a similar degree as fluctuations from occasion to occasion and from moment to moment over several blocks of a cognitive task. Thus, children's bad days concerning their achievement in school may indeed result from factors influencing the entire day (e.g., a bad night's sleep), situational effects on a given occasion of that day (e.g., enduring noise), and momentary influences effecting the performance within

one task lasting for a couple of minutes (e.g., neurocognitive processes potentially causing attentional lapses). This result confirms our hypothesis that like adults' performance also children's WM performance shows substantial variation at different timescales. Generally, the finding of reliable WM performance fluctuations from day to day, occasion to occasion, and moment to moment confirms previous findings by Schmiedek and colleagues (2013) who found younger and older adults' WM performance to vary substantially at different timescales in a laboratory study. Children's performance fluctuations at the different timescales contributed to roughly the same extent to day-to-day variability whereas for younger and older adults fluctuations at faster timescales were particularly important. There might be developmental differences in WM variability that account for the somewhat different contributions of performance fluctuations at different timescales to overall daily performance between children and adults. Few lifespan studies have demonstrated that performance fluctuations in reaction time paradigms are more pronounced in children and older adults as compared to younger adults (Li et al., 2004; Williams et al., 2005), potentially reflecting developmentally changing neuromodulatory processes such as dopaminergic transmission (Li, Lindenberger, & Sikström, 2001).

Following up on these results, we found important individual differences in the degree to which day-to-day variability in performance is composed of fluctuations at the day-to-day, occasion-to-occasion, and moment-to-moment level. For many children, a bad day of cognitive performance in school results from a mixture of systematic variation from moment to moment, from occasion to occasion, and from day to day. However, for some children, what appears to be a bad day of cognitive performance in the school context results from a series of bad moments or bad occasions without any systematic day-to-day variation. These individual differences in the contribution of systematic variability at different timescales highlight the fact that variation at all timescales needs to be further studied in order to identify the potentially different mechanisms underlying performance fluctuations for different

children and at different timescales. For example, for many children, a good night's sleep is strongly related to next day's WM performance (Könen et al., 2015). However, sleep quality might be equally important for cognitive performance in school as momentary affective states for some children or motivation to achieve good results on a given test occasion for other children. For example, Pnevmatikos and Trikkaliotis (2013) found elementary school children's inhibitory control to vary substantially within the time frame of about ten minutes following induced negative emotions. This indicates that momentary affective states may be one antecedent of children's moment-to-moment performance fluctuations in the school context. At the within-day level (e.g., school occasions in the morning hours vs. after school occasions in the afternoon) situational factors such as noise and disturbance in the classroom might be a potential antecedent of performance fluctuations (e.g., Klatter, Hellbrück, Seidel, & Leistner, 2010), but also dispositional factors such as children's time of day preference surely impact on within-day fluctuations in children's cognitive performance and thereby influence school achievement (e.g., Goldstein, Hahn, Hasher, Wiprzycka, & Zelazo, 2007). Whereas children in middle childhood can show best cognitive performance in the morning hours, this preference changes in late middle childhood to the afternoon and might thus be an individual difference characteristic likely influencing within-day performance fluctuations, particularly during the transition to puberty. At the day-to-day level, besides nightly sleep quality, experienced stress (e.g., Sliwinski et al., 2006), negative events, and intrusive thoughts might be factors influencing children's daily WM performance and thereby their school achievement. Overall, children might differ in the mechanisms underlying their performance fluctuations and different mechanisms might play a role for fluctuations at different timescales.

The finding that performance variability was to some degree consistent across timescales is in line with previous studies (Rabbitt et al., 2001) and theoretical accounts of variability (Martin & Hofer, 2004). We found day-to-day and occasion-to-occasion as well as

occasion-to-occasion and moment-to-moment variability to correlate positively with each other and to share between 10% and 11% of variance at the level of manifest aggregates across the four task conditions. On the one hand, this means that children who were more variable on one timescale were also more variable on other timescale levels potentially characterizing performance variability as a stable individual-differences construct (Rabbitt et al., 2001). This was particularly the case for occasion-to-occasion and moment-to-moment variability, and day-to-day and occasion-to-occasion variability, and less so for day-to-day and moment-to-moment variability. On the other hand, the amounts of variance that were shared between the different timescales were relatively small. A possible explanation for this finding might lie in different mechanisms accounting for variability at the different timescales. These mechanisms and the covariates previously studied in adult and adolescent samples, such as motivation (Brose et al., 2012), affect (Riediger et al., 2011), and stress (Sliwinski et al., 2006) need further consideration in the study of children's cognitive variability.

Individual Differences in Children's WM Performance Fluctuations and Their Relation to School Achievement and Fluid Intelligence

The finding that children in fourth grade were on average significantly less variable in their WM performance from occasion to occasion and from block to block than children in third grade might be explained by their longer schooling. The observed grade differences may reflect differences between children in their fluid intelligence and their school achievement (see below) that are at least in part related to their schooling (cf. Ceci, 1991). Further, children in fourth grade might show some more maturity in dealing with disturbances at the occasion-to-occasion level and thus they are less bothered and somewhat more constant in their WM performance from occasion to occasion and from moment to moment. Potentially, this finding also reflects more mature self-regulatory processes since WM has been discussed as a central process in the regulation of behavior (Barrett, Tugade, & Engle, 2004; Hofmann, Schmeichel,

& Baddeley, 2012). Recent empirical evidence shows that maintaining active representations of desired outcomes and updating information under changing conditions is essential for children's self-regulation in many social, emotional, and cognitive processes in school and beyond. For example, McQuade, Murray-Close, Shoulberg, and Hoza (2013) found a positive relationship between WM and social adjustment and discussed their finding in the light of children's conflict resolution abilities. With regards to well-being, a focus of cognitive resources on negative thoughts (i.e., rumination) has been found to have detrimental effects on children's well-being (Rood, Roelofs, Bögels, Nolen-Hoeksema, & Schouten, 2009).

Paralleling the observed grade differences in WM performance fluctuations, we also found that children who show lower achievement in mathematics and reading were more variable in their WM performance. Moreover, children with lower fluid intelligence also varied more from one day to the next. This finding suggests, on the one hand, that WM fluctuations might be a vulnerability factor identifying lower achievers and children with difficulties in school-related performance paralleling results reported on performance fluctuations in reaction time tasks (e.g., MacDonald et al., 2003) and for children with learning disabilities (e.g., Geurts, et al., 2008). On the other hand, this finding might be somewhat surprising in light of the research by Siegler and colleagues (cf. Siegler, 2006) who showed that increased performance variability is associated with strategy development and indicates learning and developmental change. The latter might be explained by the nature of our WM tasks that did not encourage strategy development to an extent similar to the complex reasoning tasks applied in the studies by Siegler and colleagues. Moreover, we want to stress that although highly relevant, the finding of WM fluctuations being associated with lower levels of school achievement and fluid intelligence is based on between-person correlations. These between-person findings do not necessarily inform us about cognitive within-person processes (Molenaar, 2004; Schmitz, 2006) that build the causal link between WM capacity and school achievement. Our findings do not imply that within-person fluctuations in WM

resources are negatively coupled with fluctuations in mathematical or reading performance, or fluid intelligence to the same degree for all, or even any, of the children in our sample.

Therefore, we refrain from an interpretation of the found negative relations between WM fluctuations and school achievement in terms of functional mechanisms explaining differences between children in school achievement. Nevertheless, we believe that the finding of substantial across- and within-day fluctuations in elementary school children's WM performance and their average negative relation to school achievement and fluid intelligence offers new insights for cognitive and educational research as well as practical implications for teachers and parents.

Limitations and Implications for Research and Educational Practice

Although our results add to the literature on cognitive performance fluctuations and the role of WM differences for school achievement, they must be considered alongside a number of limitations. First, we focused on within-person processes and administered a multivariate micro-longitudinal study to children in third and fourth grade of elementary school. Since this study is the first to report substantial and reliable WM fluctuations at different timescales in children as young as eight years, the mechanisms underlying these fluctuations at different timescales are not clear. Future studies are needed to investigate factors influencing the consistency of cognitive performance and to understand the mechanisms underlying performance variability. Within the larger project that this study is part of, we will be able to at least investigate some of the relationships of performance fluctuations with other daily varying constructs thereby shedding light on some influential factors, such as sleep, affect, motivation, and physical activity. However, future research will surely benefit from study designs combining ambulatory assessment studies focusing on cognitive performance in contexts where they are needed with laboratory assessments and neuro-cognitive studies in order to understand the mechanisms underlying cognitive performance variability at different timescales. Moreover, the intensive study protocol that

was applied in children's everyday context did not allow studying a larger sample concerning between-person differences and precluded us from further investigating variables that potentially explain differences between children in the amount and relative contribution of variation at different timescales. This is also reflected in the even smaller sample for which we could estimate variance components describing WM fluctuations at the different timescales. Future studies should therefore replicate the present findings and include micro-longitudinal ambulatory assessment bursts with children in larger educational studies that focus on the development of cognitive performance and achievement, thereby allowing to better integrate the within-person and the between-person perspective.

Second, we focused on the individual development of children in the school context. This within-person perspective certainly enriches our understanding of learning processes. However, for a comprehensive understanding of children's performance fluctuations and their underlying mechanisms in the school context the teacher's perspective and a detailed classroom assessment are indispensable. Performance fluctuations, particularly from occasion to occasion, might reflect lesson-specific effects and also effects of time of the school day (e.g., first lesson vs. last lesson). On the one hand, the teacher's cognitive activation likely supports cognitive performance and achievement and might lead to more stable performance. On the other hand, there is first evidence that instructional quality itself, including classroom management, cognitive activation, and supportive climate are not stable across lessons and school days (Praetorius, Pauli, Reusser, Rakoczy, & Klieme, 2014).

Third, we assessed cognitive performance fluctuations in WM tasks because of the empirical and theoretical importance of WM capacity for school achievement in elementary school. Despite its high correlations with school achievement, WM can only serve as a proxy for school achievement and future studies should try to repeatedly assess both WM capacity and school achievement with standardized tasks with good psychometric properties in order to directly relate day-to-day WM fluctuations to daily achievement in school. However, given

that the repeated assessment of school achievement on a daily or weekly basis comes with a number of difficulties (e.g., need for a large number of psychometrically comparable test items) we consider WM tasks to be a very useful proxy for school achievement.

To conclude, children show substantial variation in their daily cognitive performance in the school context. What appears to be a good vs. bad day is due to a combination of experiencing bad moments, bad occasions, and bad days for most children. This investigation represents an addition to the literature on WM and establishes fluctuations in children's WM performance at different timescales as an important source of cognitive performance and school achievement. The findings of the present study highlight the need for theoretical models of learning and development that incorporate variability in cognitive performance (cf. Nesselroade, 1991; Siegler, 2006) and address mechanisms underlying individual differences in performance fluctuations at different timescale levels. Educational research should replicate the present findings and further investigate the within-person relation between children's cognitive resources and their school achievement. Teachers and parents can be assured that performance fluctuations are typical for children in middle childhood. Yet, the observation of substantial performance fluctuations for individual children should naturally lead to an exploration of potential antecedents producing these fluctuations. Although the mechanism underlying children's WM fluctuations are not well understood yet, interventions helping children to focus their attention and to increase their self-regulation in learning activities are likely good candidates for reducing the number of bad days, occasions, and moments, and thereby increasing children's performance in school.

Words: 9890

References

- Allaire, J. C., & Marsiske, M. (2005). Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. *Psychology and Aging, 20*, 390–401. doi:10.1037/0882-7974.20.3.390
- Alloway, T. P., Gathercole, S. E., Adams, A.-M., Willis, C., Eaglen, R., & Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *British Journal of Developmental Psychology, 23*, 417-426. doi:10.1348/026151005X26804
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.
- Anstey, K. J., Dear, K., Christensen, H., & Jorm, A. F. (2005). Biomarkers, health, lifestyle, and demographic variables as correlates of reaction time performance in early, middle, and late adulthood. *Quarterly Journal of Experimental Psychology A, 58*, 5–21. doi:10.1080/02724980443000232
- Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*, 485-493. doi:10.1037/0894-4105.8.4.485
- Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*, 553–573. doi:10.1037/0033-2909.130.4.553
- Brose, A., Lövdén, M., & Schmiedek, F. (2014). Daily fluctuations in positive affect positively co-vary with working memory performance. *Emotion, 14*, 1-6. doi: 10.1037/a0035210.
- Brose, A., Schmiedek, F., Lövdén, M., & Lindenberger, U. (2012). Daily variability in working memory is coupled with negative affect: The role of attention and motivation. *Emotion, 12*, 605–617. doi:10.1037/a0024436
- Brose, A., Schmiedek, F., Lövdén, M., Molenaar, P. C. M., & Lindenberger, U. (2010). Adult age differences in covariation of motivation and working memory performance:

- Contrasting between-person and within-person findings. *Research in Human Development, 7*, 61–78. doi:10.1080/15427600903578177
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*, 703-722. doi: 10.1037/0012-1649.27.5.703
- Castellanos, F. X., Sonuga-Barke, E. J. S., Scheres, A., Di Martino, A., Hyde, C., & Walters, J. R. (2005). Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. *Biological Psychiatry, 57*, 1416-1423. doi: 10.1016/j.biopsych.2004.12.005
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769-786. doi: 10.3758/BF03196772
- Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin, 32*, 917–929. doi:10.1177/0146167206287721
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422–433.
- Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey. *Psychology and Aging, 21*, 62-73. doi: 10.1037/0882-7974.21.1.62
- Ebner-Priemer, U. W., Kubiak, T., & Pawlik, K. (2009). Ambulatory Assessment. *European Psychologist, 14*, 95–97. doi:10.1027/1016-9040.14.2.95
- Englund, J. A., Decker, S. L., Woodlief, D. T., DiStefano, C. (2014). Development and evaluation of an online, multicomponent working memory battery. *Assessment, 21*, 543-561. doi: 10.1177/1073191114524016

- Fozard, J. L., Vercruyssen, M., Reynolds, S. L., Hancock, P. A., & Quilter, R. E. (1994). Age differences and changes in reaction time: The Baltimore Longitudinal Study of Aging. *Journal of Gerontology, 49*, P179-P189. doi: 10.1093/geronj/49.4.P179
- Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory & Language, 51*, 136-158. doi: 10.1016/j.jml.2004.03.008
- Friso-van den Bos, I., van der Ven, S. H., Kroesbergen, E. H., & van Luit, J. E. (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
- Gasimova, F., Robitzsch, A., Wilhelm, O., Boker, S. M., Hu, Y., & Hülür, G. (2014). Dynamical systems analysis applied to working memory data. *Frontiers in Psychology, 5*:687. doi:10.3389/fpsyg.2014.00687
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*, 72–91. doi:10.1037/a0032138
- Geurts, H. M., Grasman, R. P. P. P., Verté, S., Oosterlaan, J., Roeyers, H., van Kammen, S. M., & Sergeant, J. A. (2008). Intra-individual variability in ADHD, autism spectrum disorders and Tourette's syndrome. *Neuropsychologia, 46*, 3030-3041. doi:10.1016/j.neuropsychologia.2008.06.013
- Giofrè, D., Mammarella, I. C., & Cornoldi, C. (2013). The structure of working memory and how it relates to intelligence in children. *Intelligence, 41*, 396-406. doi:10.1016/j.intell.2013.06.006
- Goldstein, D., Hahn, C. S., Hasher, L., Wiprzycka, U. J. & Zelazo, P. D. (2007). Time of day, intellectual performance, and behavioral problems in morning versus evening type adolescents: Is there a synchrony effect? *Personality and Individual Differences, 42*, 431-440. doi: 10.1016/j.paid.2006.07.008

- Gölitz, D., Roick, T., & Hasselhorn, M. (2006). *Deutscher Mathematiktest für vierte Klassen (DEMAT 4)* [German mathematics test for fourth grades (DEMAT 4)]. Göttingen, Germany: Beltz.
- Göthe, K., Esser, G., Gendt, A., & Kliegl, R. (2012). Working memory in children: Tracing age differences and special educational needs to parameters of a formal model. *Developmental Psychology*, 48, 459–476. doi:10.1037/a0025660
- Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, 130, 184-198. doi: 10.1037/0096-3445.130.2.184
- Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences*, 16, 174–180. doi:10.1016/j.tics.2012.01.006
- Hornung, C., Brunner, M., Reuter, R. A., & Martin, R. (2011). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence*, 39, 210–221. doi:10.1016/j.intell.2011.03.002
- Hoppmann, C. & Riediger, M. (2009). Ambulatory assessment in lifespan psychology: An overview of current status and new trends. *European Psychologist*, 14, 98-108. doi: 10.1027/1016-9040.14.2.98
- Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences*, 13, 869-881. doi: 10.1016/0191-8869(92)90004-9
- Klatte, M., Hellbrück, J., Seidel, J., & Leistner, P. (2010). Effects of classroom acoustics on performance and well-being in elementary school children: A field study. *Environment & Behavior*, 42, 659-692. doi: 10.1177/0013916509336813
- Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments, & Computers*, 31, 429-432. doi: 10.3758/BF03200722

- Könen, T., Dirk, J., & Schmiedek, F. (2015). Cognitive benefits of last night's sleep: Daily variations in children's sleep behavior are related to working memory fluctuations. *Journal of Child Psychology and Psychiatry*, *56*, 171-182. doi: 10.1111/jcpp.12296
- Lenhard, W., & Schneider, W. (2005). *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässer* [A reading comprehension test for first to sixth graders]. Göttingen: Hogrefe.
- Lecerf, T., Ghisletta, P., & Jouffray, C. (2004). Intraindividual variability and level of performance in four visuo-spatial working memory tasks. *Swiss Journal of Psychology*, *63*, 261-272. doi: 10.1024/1421-0185.63.4.261
- Li, S.-C., Aggen, S. H., Nesselroade, J. R., & Baltes, P. B. (2001). Short-term fluctuations in elderly people's sensorimotor functioning predict text and spatial memory performance: The MacArthur Successful Aging Studies. *Gerontology*, *47*, 100–116.
- Li, S.-C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences*, *5*, 479–486. doi:10.1016/S1364-6613(00)01769-1
- Loosli, S. V., Buschkuhl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology*, *18*, 62–78. doi:10.1080/09297049.2011.575772
- MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria Longitudinal Study. *Psychology and Aging*, *18*, 510–523. doi:10.1037/0882-7974.18.3.510
- Martin, M., & Hofer, S. M. (2004). Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology*, *50*, 7–11. doi:10.1159/000074382
- McQuade, J. D., Murray-Close, D., Shoulberg, E. K., & Hoza, B. (2013). Working memory and social functioning in children. *Journal of Experimental Child Psychology*, *115*, 422–435. doi:10.1016/j.jecp.2013.03.002

- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2*, 201-218.
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide*. Los Angeles, USA: Muthén & Muthén.
- Nagengast, B., Trautwein, U., Kelava, A., & Lüdtke, . (2013). Synergistic effects of expectancy and value on homework engagement: The case for a within-person perspective. *Multivariate Behavioral Research, 48*, 428-460. doi: 10.1080/00273171.2013.775060
- Neisser, U., & Winograd, E. (1988). *Remembering reconsidered: Ecological and traditional approaches to the study of memory*. Cambridge, UK: Cambridge University Press.
- Nesselroade, J. R. (1991). The warp and the woof of the developmental fabric. In R. M. Downs, L. S. Lybben, & D. S. Palermo (Eds.), *Visions of aesthetics, the environment & development: The legacy of Joachim F. Wohlwill* (pp. 213-240). Hillsdale, NJ: L. Erlbaum Associates.
- Nesselroade, J. R., & Featherman, D. L. (1997). Establishing a reference frame against which to chart age-related changes. In S. A. Hardy (Ed.), *Studying aging and social change: Conceptual and methodological issues* (pp. 191-205). Newbury Park, CA: Sage.
- Nesselroade, J. R., & Jones, C. J. (1991). Multi-modal selection effects in the study of adult development: A perspective on multivariate, replicated, single-subject, repeated measures designs. *Experimental Aging Research, 17*, 21–27. doi: 10.1080/03610739108253883
- Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *Journal of Gerontology, 59B*, P49-P55.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory (3rd ed.)*. New York: McGraw-Hill.

- Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2008). Individual differences in working memory capacity and reasoning ability. In A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. Towse (Eds.), *Variation in Working Memory* (pp. 49–75). Oxford: Oxford University Press.
- Pnevmatikos, D., & Trikkaliotis, I. (2013). Intraindividual differences in executive functions during childhood: The role of emotions. *Journal of Experimental Child Psychology, 115*, 245-261. doi: 10.1016/j.jecp.2013.01.010
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction, 31*, 2–12. doi:10.1016/j.learninstruc.2013.12.002
- Rabbitt, P., Osman, P., Moore, B., & Stollery, B. (2001). There are stable individual differences in performance variability, both from moment to moment and from day to day. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54*, 981–1003.
- Reis, H. T. (2012). Why researchers should think “real-world”: A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 3-21). New York: The Guilford Press.
- Riediger, M., Wrzus, C., Schmiedek, F., Wagner, G. G., & Lindenberger, U. (2011). Is seeking bad mood cognitively demanding? Contra-hedonic orientation and working-memory capacity in everyday life. *Emotion, 11*, 656–665. doi:10.1037/a0022756
- Riediger, M., Wrzus, C., Klipker, K., Müller, V., Schmiedek, F., & Wagner, G. G. (2014). Outside of the laboratory: Associations of working-memory performance with psychological and physiological arousal vary with age. *Psychology and Aging, 29*, 103-114. doi:10.1037/a0035766

- Rittle-Johnson, B., & Siegler, R. S. (1999). Learning to spell: Variability, choice, and change in children's strategy use. *Child Development, 70*, 332-348. doi: 10.1111/1467-8624.00025
- Robertson, S., Myerson, J., Hale, S. (2006). Are there age differences in intraindividual variability in working memory performance? *Journals of Gerontology: Psychological Sciences, 61B*, 18-24. doi: 10.1093/geronb/61.1.P18
- Rood, L., Roelofs, J., Bögels, S. M., Nolen-Hoeksema, S., & Schouten, E. (2009). The influence of emotion-focused rumination and distraction on depressive symptoms in non-clinical youth: A meta-analytic review. *Clinical Psychology Review, 29*, 607–616. doi:10.1016/j.cpr.2009.07.001
- Roick, T., Göllitz, D., & Hasselhorn, M. (2004). *Deutscher Mathematiktest für dritte Klassen (DEMAT 3+)* [German mathematics test for third grades (DEMAT 3+)]. Göttingen, Germany: Beltz.
- SAS Institute Inc (2002-2005). *SAS Online Documentation 9.1.3*. Cary, NC: SAS Institute Inc.
- Schmiedek, F., Lövdén, M., & Lindenberger, U. (2013). Keeping it steady: Older adults perform more consistently on cognitive tasks than younger adults. *Psychological Science, 24*, 1747–1754. doi:10.1177/0956797613479611
- Schmitz, B. (2006). Advantages of studying processes in educational research. *Learning and Instruction, 16*, 433–449. doi:10.1016/j.learninstruc.2006.09.004
- Schmitz, B., & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology, 31*, 64–96. doi:10.1016/j.cedpsych.2005.02.002
- Shrout, P. E. & Lane, S. P. (2012). Psychometrics. In M. R. Mehl, & T. S. Conner (Eds). *Handbook of Research Methods for Studying Daily Life* (pp. 302 - 320). New York: Guilford.

- Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Volume 2: Cognition, perception, and language* (6th ed., pp. 464-510). Hoboken, NJ: Wiley.
- Sliwinski, M. J., Smyth, J. M., Hofer, S. M., & Stawski, R. S. (2006). Intraindividual coupling of daily stress and cognition. *Psychology and Aging, 21*, 545–557. doi:10.1037/0882-7974.21.3.545
- Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability - and a little bit more. *Intelligence, 30*, 261–288. doi:10.1016/S0160-2896(01)00100-3
- Swanson, H. L. (2011). WM, attention, and mathematical problem solving: A longitudinal study of elementary school children. *Journal of Educational Psychology, 103*, 821–837. doi:10.1037/a0025114
- Swanson, H. L., & Alloway, T. P. (2012). Working memory, learning, and academic achievement. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, Vol 1: Theories, constructs, and critical issues* (pp.327–366). Washington: American Psychological Association.
- Swanson, H. L., & Jerman, O. (2007). The influence of working memory on reading growth in subgroups of children with reading disabilities. *Journal of Experimental Child Psychology, 96*, 249-283. doi: 10.1016/j.jecp.2006.12.004
- Tsai, Y.-M., Kunter, M., Lüdtke, O., Trautwein, U., & Ryan, R. M. (2008). What makes lessons interesting? The role of situational and individual factors in three school subjects. *Journal of Educational Psychology, 100*, 460–472. doi:10.1037/0022-0663.100.2.460
- Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory & Language, 49*, 446-468. doi: 10.1016/S0749-596X(03)00095-0

Weiss, H. R. (2006). *Grundintelligenztest Skala 2 - Revision (CFT 20-R)*. Göttingen: Hogrefe.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4*, 433. doi:10.3389/fpsyg.2013.00433

Wilhelm, P., & Schoebi, D. (2007). Assessing mood in daily life. Structural validity, sensitivity to change, and reliability of a short scale to measure three basic dimensions of mood. *European Journal of Psychological Assessment, 23*, 258–267. doi: 10.1027/1015-5759.23.4.258

Williams, B. R., Hultsch, D. F., Strauss, E., Hunter, M. A., & Tannock, R. (2005).

Inconsistency in reaction time across the life span. *Neuropsychology, 19*, 88-96. doi: 10.1037/0894-4105.19.1.88

Footnotes

1 Overall, only few children (at maximum 2% in Grade 3, and 3.3% in Grade 4) showed an average accuracy of above 95% across task conditions and study days, and grade differences in the number of high performing children were negligible.

2 Taking linear long-term trends into account lowered factor loadings at the within-person level as would be expected (numerical load 2/3: .69/ .77, spatial load 2/3: .46/ .42, respectively); the model fit was good [$\chi^2(1) = 19.66, p = .00$; CFI = .99; RMSEA = .05; SRMR within = .01; SRMR between = .00] and comparable to a model without trends (see Figure 2).

3 Note that all results in Table 3 were obtained controlling for item difficulty effects.

Figure Captions

Figure 1. Examples of the numerical (top) and spatial (bottom) WM updating tasks presented. Examples represent Memory Load 2. Both tasks were administered with both Load 2 and Load 3. Load 2 task conditions demand 3 updates, Load 3 task condition demand 4 updates.

Figure 2. Within- and between-person factor of working memory performance. WM = working memory. Factor loadings are standardized. Squares represent observed variables and circles represent latent variables. All factor loadings and correlations were significant at $p < .05$. The fit of the model to the data was good [$\chi^2(3) = 51.81, p = .00$; CFI = .99; RMSEA = .05; SRMR within = .01; SRMR between = .01].

Figure 3. Children's average estimated day-to-day, occasion-to-occasion, and moment-to-moment variance components across WM task conditions. $N = 83$ children. Num = numerical, Spat = Spatial, L2 = Load 2, L3 = Load 3. The total size of the bars corresponds to the sample average variance of observed day-to-day variability (i.e., the variance of average performance across days). This variance is decomposed into a variance component of systematic day-to-day fluctuations (black), the contribution of occasion-to-occasion variability to observed day-to-day variability (light grey), and the contribution of moment-to-moment variability to observed day-to-day variability (dark grey).

Figure 4. Children's individual estimated day-to-day, occasion-to-occasion, and moment-to-moment variance components across WM task conditions. $N = 83$ children. Individual variance components of the four WM task conditions: a) Num L2 = numerical load 2, b) Num L3 = numerical load 3, c) Spat L2 = spatial load 2, and d) Spat L3 = spatial load 3. Each bar represents one individual's observed day-to-day variability. The total size of the bars

corresponds to the individual variance of observed day-to-day variability (i.e., the variance of average performance across days).

WORKING MEMORY FLUCTUATIONS IN THE SCHOOL CONTEXT

Table 1

Correlations Between WM Task Conditions at the Between-Person and Within-Person Level

	Numerical WM load 2	Numerical WM load 3	Spatial WM load 2	Spatial WM load 3
Numerical WM load 2	1	.58	.40	.33
Numerical WM load 3	.98 [.64]	1	.40	.39
Spatial WM load 2	.87 [.50]	.84 [.46]	1	.51
Spatial WM load 3	.80 [.53]	.82 [.48]	.89 [.60]	1

Notes. * $p < .05$, WM = Working Memory. Daily within-person correlations ($N = 6469$) are presented above the diagonal, average daily between-person correlations are presented below the diagonal. Correlation coefficients presented in square brackets below the diagonal are baseline between-person correlations ($N = 104$).

WORKING MEMORY FLUCTUATIONS IN THE SCHOOL CONTEXT

Table 2

Descriptive Statistics of Daily WM Measures

Task Condition	<i>Occ</i>	<i>M (SD)</i>	<i>ICC</i>	Mean <i>ISD (SD)</i>
Numerical WM Load 2				
	morning	.75 (.32)	.45	.23 (.11)
	noon	.68 (.35)	.50	.23 (.10)
	afternoon	.68 (.34)	.38	.26 (.08)
<i>Mean numerical WM Load</i>	all	.70 (.34)	.44	.25 (.08)
2				
Numerical WM Load 3				
	morning	.65 (.34)	.53	.22 (.09)
	noon	.58 (.36)	.54	.23 (.10)
	afternoon	.63 (.34)	.49	.24 (.10)
<i>Mean numerical WM Load</i>	all	.62 (.35)	.52	.23 (.08)
3				
Spatial WM Load 2				
	morning	.77 (.30)	.35	.23 (.11)
	noon	.69 (.35)	.35	.27 (.11)
	afternoon	.73 (.33)	.32	.25 (.11)
<i>Mean spatial WM Load 2</i>	all	.73 (.33)	.35	.26 (.09)
Spatial WM Load 3				
	morning	.56 (.32)	.43	.23 (.07)
	noon	.47 (.33)	.38	.25 (.07)
	afternoon	.53 (.33)	.39	.24 (.07)

WORKING MEMORY FLUCTUATIONS IN THE SCHOOL CONTEXT

<i>Mean spatial WM Load 3</i>	all	.52 (.33)	.40	.24 (.06)
-------------------------------	-----	-----------	-----	-----------

Notes. Statistics presented are based on accuracy of performance in the WM task conditions.

ICC = intraclass correlation (the portion of between-person variance on total variance). *Occ* =

Occasion. Mean *ISD* = mean intraindividual standard deviation.

WORKING MEMORY FLUCTUATIONS IN THE SCHOOL CONTEXT

Table 3

Variance Components and Grade Differences in Variance Components.

	Grade 3			Grade 4		Grade differences			
	VC Days (SE)	VC Occ.s (SE)	VC Blocks (SE)	VC Days (SE)	VC Occ.s (SE)	VC Blocks (SE)	Days χ^2	Occ.s χ^2	Blocks χ^2
Numerical WM load 2	0.007 (0.002)	0.035 (0.002)	0.085 (0.002)	0.005 (0.001)	0.019 (0.001)	0.061 (0.001)	0.5	44.2	199.9
Numerical WM load 3	0.007 (0.002)	0.031 (0.002)	0.069 (0.001)	0.007 (0.001)	0.023 (0.001)	0.057(0.001)	0.0	12.6	65.4
Spatial WM load 2	0.010 (0.002)	0.036 (0.002)	0.097 (0.002)	0.005 (0.001)	0.025 (0.001)	0.083 (0.001)	6.0	15.7	44.0
Spatial WM load 3	0.009 (0.002)	0.025 (0.002)	0.096 (0.002)	0.006 (0.001)	0.025 (0.002)	0.099 (0.001)	2.5	0.0	2.0

WORKING MEMORY FLUCTUATIONS IN THE SCHOOL CONTEXT

Note. Variance components (VC), associated standard errors (SE), and χ^2 -Tests (with one degree of freedom; critical value = 3.84 for $p < .05$) based on likelihood ratios comparing unconstrained (i.e., parameters freely estimated) to constrained models (i.e., parameters constrained to be equal across grades). Occ. = occasion. $N = 83$ children.

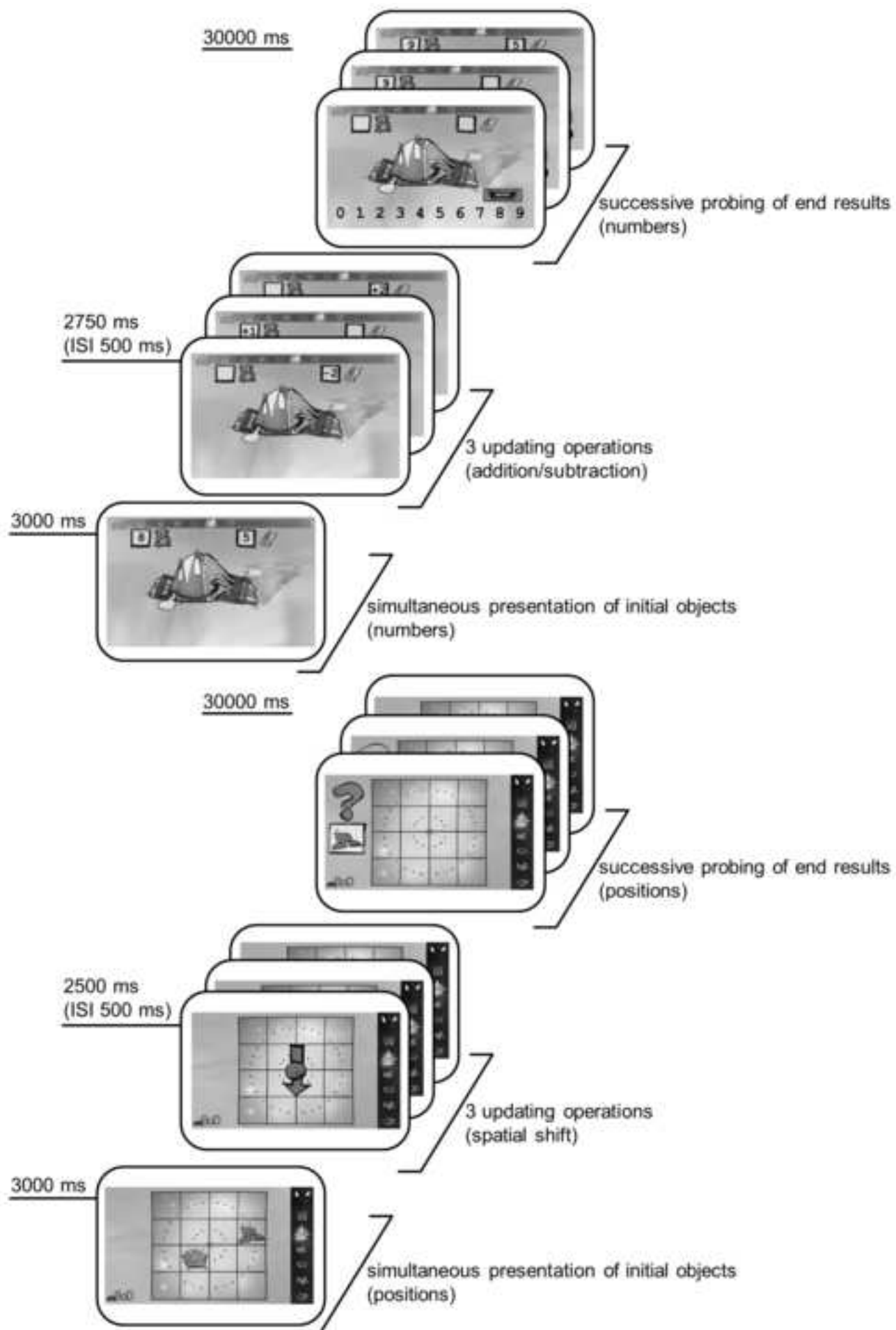
WORKING MEMORY FLUCTUATIONS IN THE SCHOOL CONTEXT

Table 4. *Relationships Between Children's WM Fluctuations, and School Achievement and Fluid Intelligence*

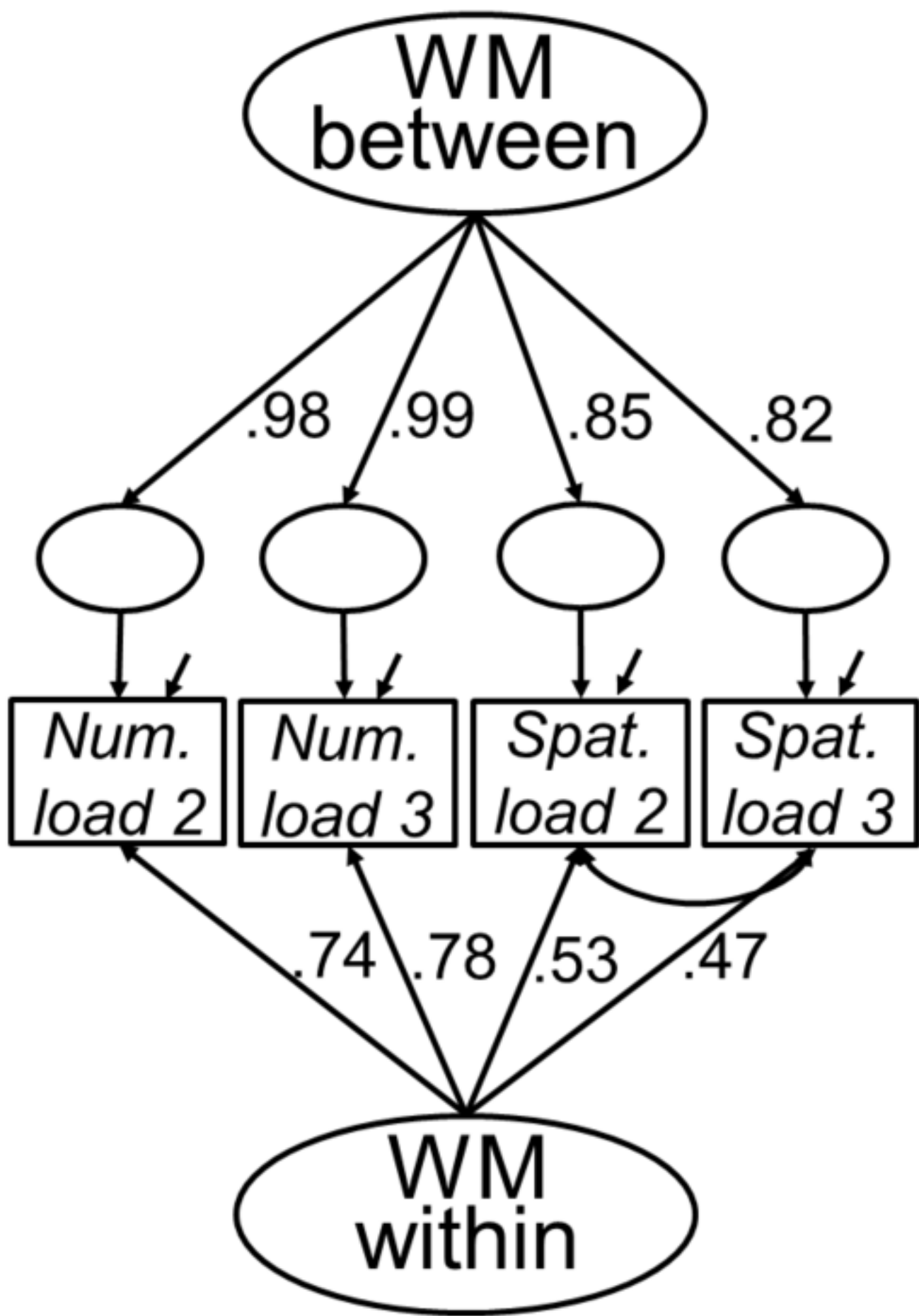
	Mathematics	Reading	Fluid Intelligence
VC Days	-.30	-.35	-.32
VC Occasions	-.26	-.27	-.09
VC Blocks	-.28	-.22	-.15

Note. VC = variance component. Cells contain standardized latent correlation coefficients, obtained in three separate structural equation models with the following model fit: Model 1 (VCs-Mathematics): $\chi^2(3) = 0.34, p = .95, CFI = 1.00, RMSEA = .00 [.00; .00], SRMR = .01$. Model 2 (VCs-Reading): $\chi^2(6) = 7.16, p = .31, CFI = 1.00, RMSEA = .04 [.00; .14], SRMR = .03$. Model 3 (VCs-Fluid Intelligence): $\chi^2(10) = 12.47, p = .26, CFI = .97, RMSEA = .05 [.00; .12], SRMR = .05$. All coefficients in bold face are statistically significant at $p < .05$.

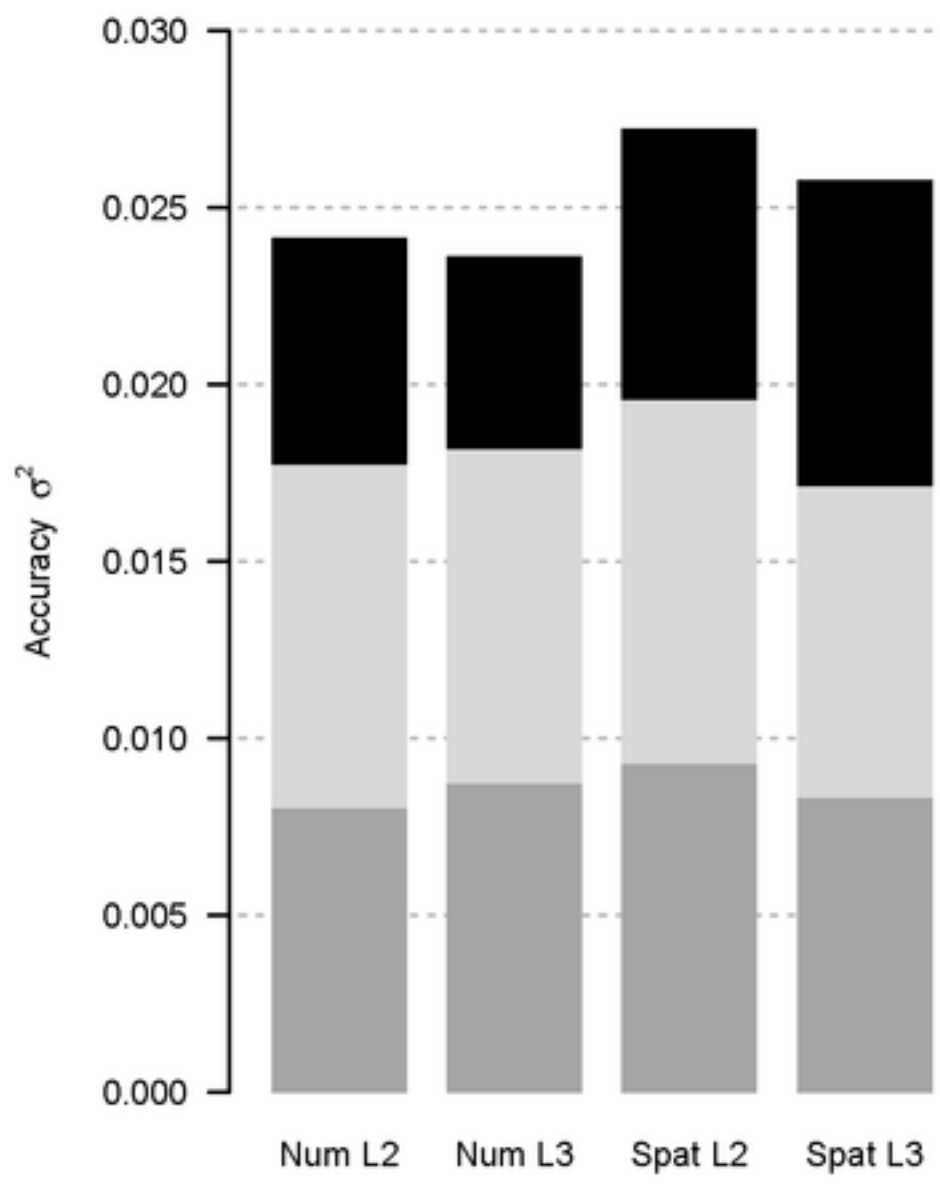
Figure



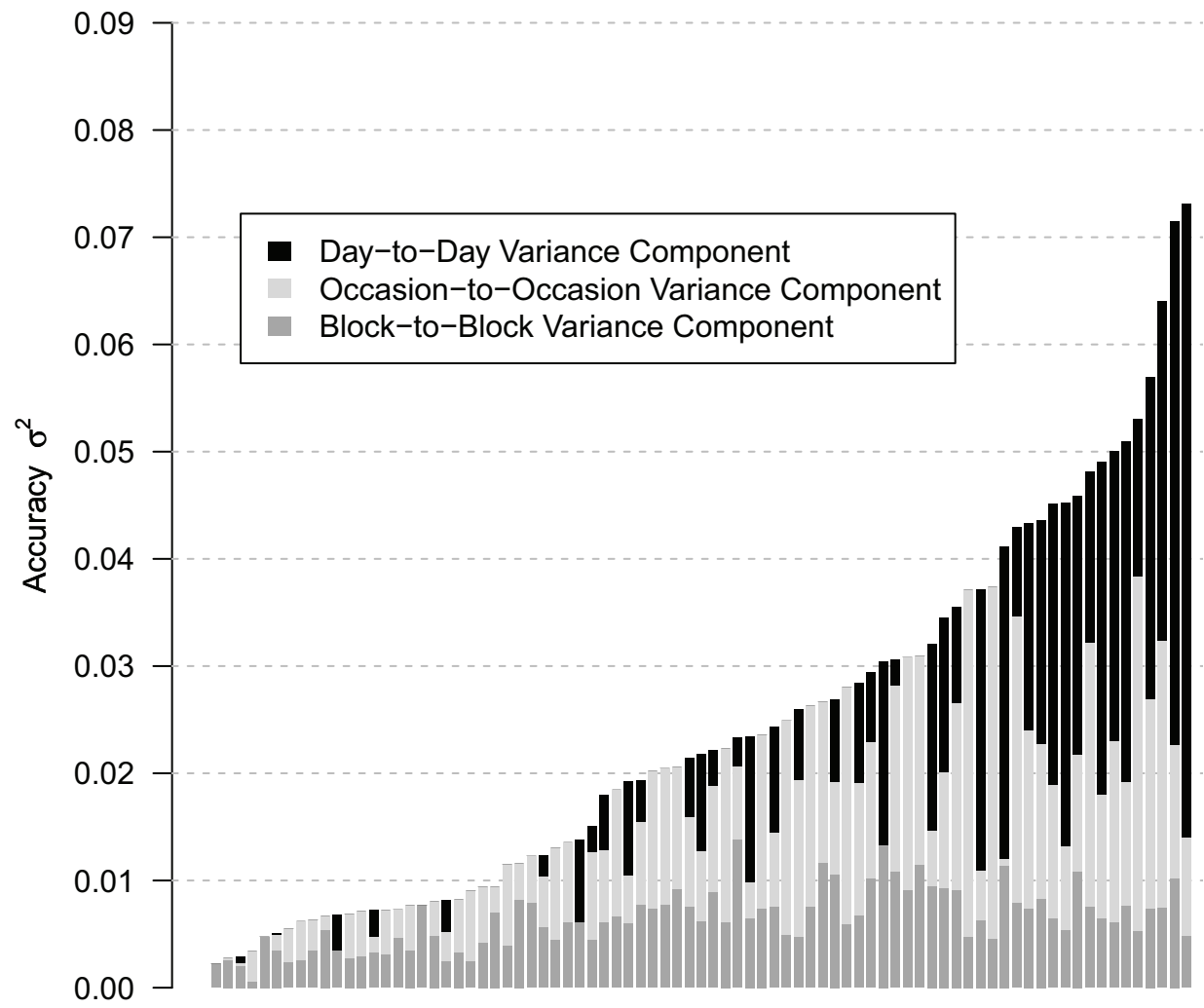
Figure



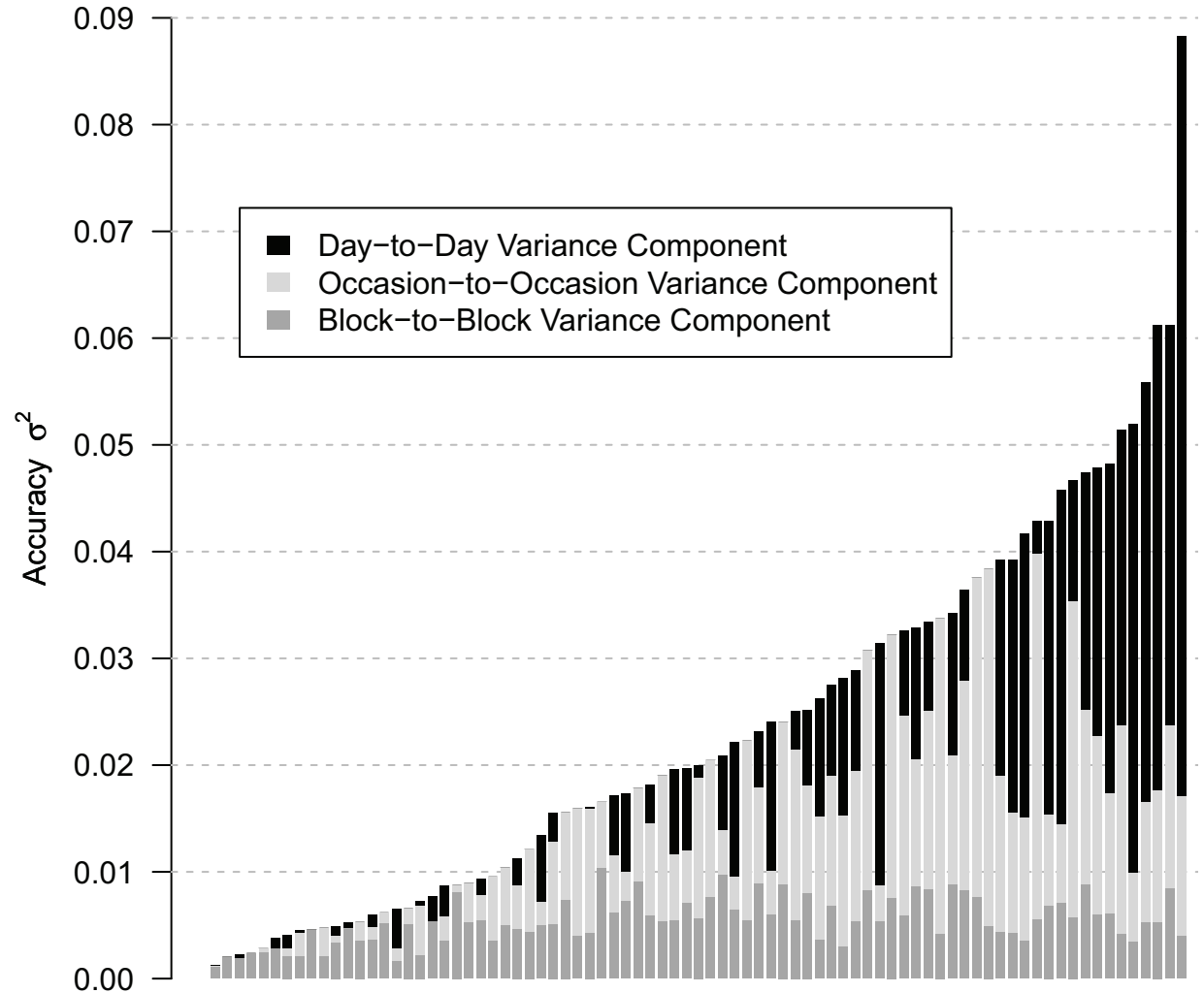
Figure



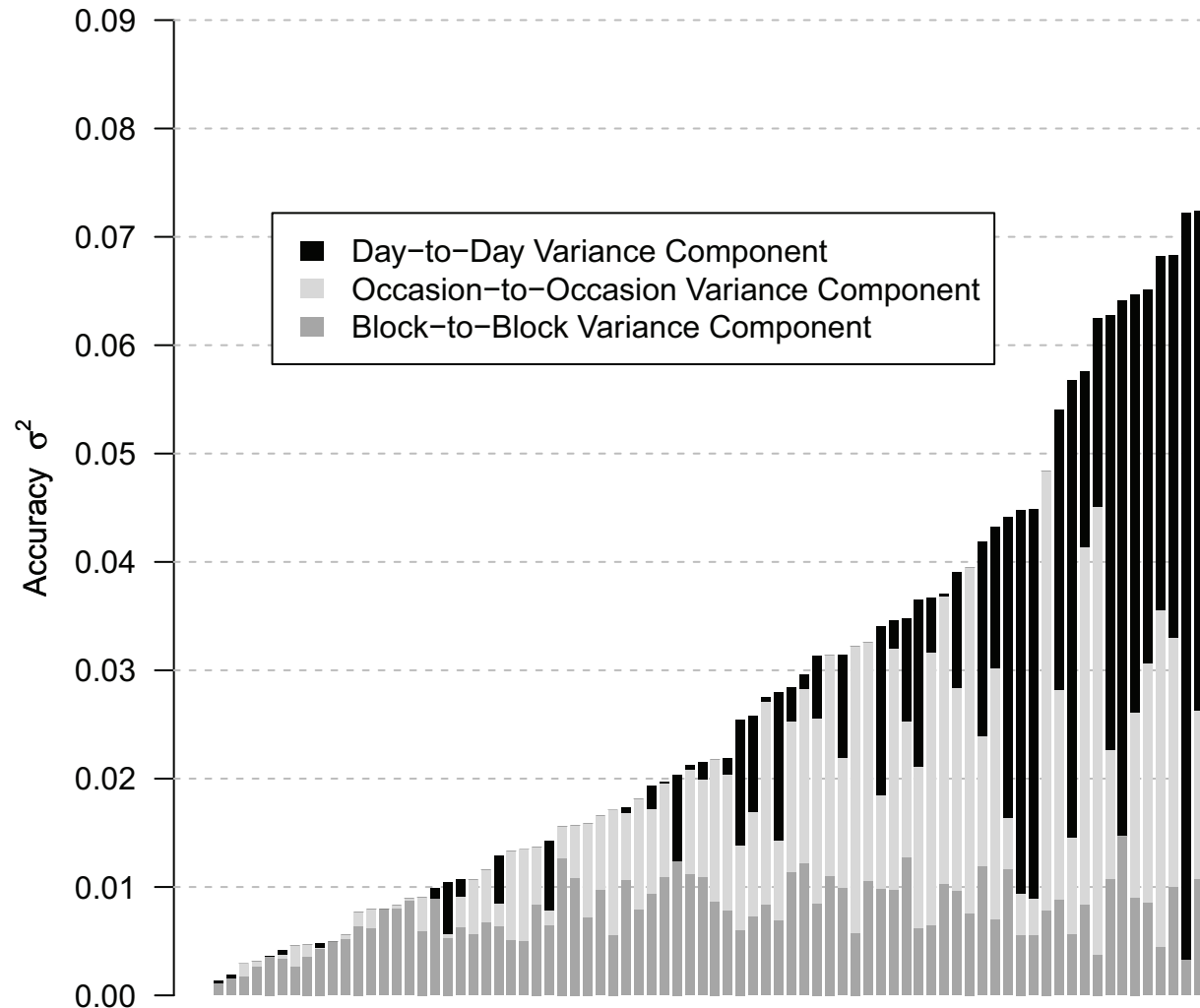
Numerical Load 2



Numerical Load 3



Spatial Load 2



Spatial Load 3