Screening for Working Memory Deficits in the Classroom: The Psychometric Properties of the Working Memory Rating Scale in a Longitudinal School-Based Study
Sébastien Normand and Rosemary Tannock
Journal of Attention Disorders 2014 18: 294 originally published online 24 May 2012
DOI: 10.1177/1087054712445062
The online version of this article can be found at:
http://jad.sagepub.com/content/18/4/294

Published by:
SAGE
http://www.sagepublications.com

Additional services and information for Journal of Attention Disorders can be found at:

Email Alerts: http://jad.sagepub.com/cgi/alerts
Subscriptions: http://jad.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav
Citations: http://jad.sagepub.com/content/18/4/294.refs.html

>> Version of Record - Apr 24, 2014
OnlineFirst Version of Record - May 24, 2012
What is This?
Screening for Working Memory Deficits in the Classroom: The Psychometric Properties of the Working Memory Rating Scale in a Longitudinal School-Based Study

Sébastien Normand¹,² and Rosemary Tannock²,³

Abstract

Objective: The aim of this 18-month longitudinal study was to explore the psychometric properties of the recently developed Working Memory Rating Scale (WMRS) within a general school population of 524 six- to nine-year-old children (259 boys, 265 girls) and with an examination of sex and time differences. Method: Teachers completed the WMRS and children completed objective measures of WM and standardized measures of academic achievement. Results: Confirmatory factor analyses indicated a poor fit for the 20 original WMRS items. Post hoc analyses, however, revealed that the factor structure of an alternative five-item short form was strong for both boys and girls and at the two time points, spanning two consecutive academic years. Internal consistency, criterion-related validity, and convergent validity of this alternative five-item WMRS were also excellent. Conclusion: The short five-item WMRS may eventually provide teachers with a useful and time-effective method to screen for WM deficits at school. (J. of Att. Dis. 2014; 18(4) 294-304)

Keywords
behavioral ratings, longitudinal study, psychometric properties, sex differences, working memory

Working memory (WM) refers to a limited-capacity, multicomponent cognitive system that allows us to temporarily hold and manipulate relevant information in the mind over brief periods of time while performing complex tasks in daily life, such as mental calculation, reading comprehension, and reasoning (e.g., Hoskyn & Swanson, 2003). Although there are several competing models of WM (see Miyake & Shah, 1999, for a review), most differentiate between two processes: storage of information (i.e., the process of keeping information in mind in the absence of the external stimulus) and manipulation of information (i.e., the reorganization of information that is being maintained).

In his highly influential model of WM, Baddeley (2003; Baddeley & Hitch, 1974) presents WM as a multicomponent system coordinated and controlled by a central executive. Auditory-verbal WM is used to manipulate verbal information and learning, whereas visual-spatial WM manipulates visual images and visual-spatial learning. Extensive evidence from factor-analytic (Alloway, Gathercole, & Pickering, 2006), neuropsychological (Repovš & Baddeley, 2006), neuroanatomical (Smith, Jonides, & Koepppe, 1996), and neuroimaging (Fassbender & Schweitzer, 2006) studies support this multicomponent model of WM.

Burgeoning evidence from cognitive and clinical research also links WM impairments to behavioral difficulties (such as severe inattention, lapses-of-attention, and physical restlessness; for example, Kofler, Rapport, Bolden, Sarver, & Raiker, 2010; Rapport et al., 2009) and to academic underachievement in various domains, including written expression, reading and listening comprehension, problem solving, and mathematical reasoning (Alloway, Gathercole, & Elliott, 2010; Berninger et al., 2010; Swanson, 2011). More specifically, objective assessment of auditory-verbal WM system has been linked with the ability to acquire new knowledge and skills, particularly in the development of reading and language (Atkins & Baddeley, 1998; Cowan & Alloway, 2009; Service, 1992). Objective assessment of visual-spatial WM has also been implicated...
as a determinant of children’s academic attainment in literacy, comprehension, arithmetic, and science (Gathercole & Pickering, 2000; St. Clair-Thompson & Gathercole, 2006).

WM is not an established concept in education, and the majority of teachers have not received training in how to recognize WM problems in the classroom or how to best support such students in their learning (Gathercole, Lamont, & Alloway, 2006). Furthermore, teachers are currently unlikely to recognize WM impairments given the lack of efficient, psychometrically sound, and ecologically valid screening tools to identify students with WM impairment in the classroom. To date, there is no consensus as to the assessment of students with WM impairments. Performance-based tasks of WM lack ecological validity due to the manner in which they are typically administered. That is, testing occurs over a short time frame and in environments that are designed to minimize distractions, maximize support, and provide individuals with a high degree of structure (e.g., clear instructions, well-specified goals). This is in sharp contrast with the integrated, multidimensional, relativistic, priority-based decision making that is often demanded in real-world situations (Goldberg & Podell, 2000). Thus, performance tasks may not engage the same set of skills that are required in naturalistic settings. Moreover, current research indicates only a modest relationship at best between rating scales and performance-based measures of WM (e.g., McAuley, Chen, Goos, Schuchar, & Crosbie, 2010; Toplak, Bucciarelli, Jain, & Tannock, 2009). Rating scales, on the other hand, may circumvent this limitation by offering a view of WM in the everyday world rather than in isolation on clinic-based performance tests. Thus, there is a need for ecologically valid methods of measuring WM in the classroom.

The few available published scales assessing WM have important shortcomings. The 86-item Behavior Rating Inventory of Executive Functions (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000) is a behavior checklist that measures various aspects of executive functioning, including inhibition, shifting, emotional control, initiation, planning/organization, organization of material, monitoring, and WM. The BRIEF is therefore long and time-consuming for use as a classroom-wide screening tool, it covers a broad array of cognitive constructs and many items overlap with key features of ADHD, and therefore may confound WM with behavioral symptoms and other cognitive functions. Furthermore, studies examining the BRIEF (Gioia et al., 2000) have reported that parental ratings on the BRIEF WM subscale correlate with children’s frontal gray matter volume, but not with the children’s performance scores on WM tests (Mahone, Martin, Kates, Hay, & Horská, 2009).

The WM Rating Scale (WMRS; Alloway, Gathercole, & Kirkwood, 2008) consists of 20 descriptions assessing WM deficits in the classroom. This questionnaire, published by PsychCorp, focuses solely on WM-related problems in a single scale, it can be rapidly administered and scored, and it does not require any training in psychometric assessment prior to use. Alloway, Gathercole, Kirkwood, and Elliott (2009) provided preliminary data on the reliability and criterion validity of the WMRS in their normative sample of 417 five- to eleven-year-old children from England. Their initial findings suggested excellent internal consistency of the WMRS factor (Cronbach’s α = .978). Exploratory factor analysis suggested that a single factor accounted for 70.72% of the total variance. Then, using confirmatory factor analyses (CFAs) with the same data, the authors suggested that a two-factor model tapping behavioral (measured by the WMRS) and cognitive aspects of WM (measured by the Automated Working Memory Assessment [AWMA], a performance-based WM measure; Alloway, 2007) had a good fit. The authors also found preliminary evidence of criterion-related validity of the WMRS: Scores on the WMRS (indicating WM deficits) were significantly negatively associated with concurrent scores on the AWMA measures and on the Wechsler Intelligence Scale for Children IV (WISC-IV; Wechsler, 2003) WM Index. In other words, WM deficits as rated by teachers were significantly associated with a lower performance on the objective measures of WM (Alloway et al., 2009).

Although the investigation by Alloway et al. (2009) provided preliminary evidence that the 20 original WMRS items are well described by a one-factor structure, several constraints of this study limit the usefulness of their conclusions. First, Alloway et al. (2009) simultaneously tested CFA models of the WMRS and AMWA. Their measurement model contained the relationships between these two correlated factors and their indicators. There is still no published report of CFAs of the original 20 WMRS items taken intrinsically. This gap in the literature is important because several psychometric studies have suggested that many factor structures that are well validated and replicated through exploratory factor analyses do not show good fit in confirmatory models (e.g., Marsh et al., 2010). Second, Alloway et al. (2009) followed their explorations with CFAs using the same sample. Given that a model is likely to fit the data set it was created from better than any other random sample from the same population, researchers need to test the WMRS on different data sets. Building and testing on the same sample biased the assessment of model fit upwards. Hence, CFAs supporting the one-factor structure of the 20 WMRS items have yet to be conducted in an independent sample. Therefore, despite the fact that the WMRS is now commercially available and used to screen for WM deficits in the classroom, the psychometric properties of its original items are still inadequately understood. In addition, it is still unclear whether the WMRS, developed in England, is a valid and reliable WM-screening tool that can be used by North American teachers. Another remaining empirical question is whether the psychometric...
properties of the WMRS are robust for both boys and girls and if they are stable over time. It is also still unclear whether the WMRS scores are related to subsequent academic attainment.

The Present Study

The current study is intended as a conceptual replication and extension of Alloway et al. (2009). It replicates the earlier study by examining the factor structure, internal consistency, and criterion-related validity of the WMRS (Alloway et al., 2009). However, the current analysis is based on a large, independent sample of North American elementary schoolchildren within an 18-month longitudinal design, and with an examination of potential gender differences. This will enhance confidence in the stability of all estimates obtained, including model-fit indices and estimates of individual parameter values. Second, whereas there is extensive research on the association between objective assessments of WM and academic achievement, the relationship between classroom behaviors characteristic of WM deficits and academic learning is less clearly understood. This is particularly important given that WM deficits negatively affect students’ behavioral performance in several classroom activities, including remembering lengthy activities, keeping track of progress in multistep tasks, and completing activities previously started while simultaneously storing and processing new information (e.g., Gathercole, Lamont, et al., 2006). Thus, we explored the prospective relationship of the WMRS to several standardized measures of reading and mathematical fluency.

We predicted that the original one-factor WMRS model would fit well to our North American data for boys and girls in both years. We also anticipated that we would replicate and expand on the initial findings on the internal consistency, criterion-related validity, and convergent validity of the WMRS. More specifically, we expected that the WMRS items would be internally consistent and that the WMRS factor would be significantly negatively correlated with objective measures of WM and with standardized measures of academic attainment for both boys and girls at each time point.

Method

Participants. As detailed in Table 1, participants were 524 children (259 boys, 265 girls), aged 6 to 9 years ($M = 7.60; SD = 0.92$), their parents, and their teachers. The majority of children had English as their primary language (97%) and was primarily Caucasian (86%). Children were recruited from Grades 1 to 3 in seven public elementary schools, which constitute 20% of the 33 schools in a large rural and suburban district school board in Southern Ontario, Canada. The data were collected as part of an 18-month prospective study focused on the behavioral symptoms of attention, objective measures of attention, and academic outcomes in elementary school-aged children.

Participating parents were mostly mothers (90%) and had diverse levels of education: less than a high school degree (2%), high school degree or equivalent (6%), and more than high school (92%). According to parental report, some children were reported to have the following learning and mental-health issues: ADHD (4%), language impairment (3%), learning disability (3%), and behavior problem (2%). Among the 52 teachers who participated, 95% were females and 95% were Caucasian. Classroom size varied from 13 to 26 students with an average of 19.6 students per class ($SD = 2.5$). The number of years of experience as a teacher ranged from 1 year to 33 years with an average of 14.8 years ($SD = 7.9$). In addition, 38% of teachers possessed additional qualifications in special education. As indicated in Table 1, there were no significant differences between males and females in most demographic variables. However, females were slightly more likely than males to have a parent with less than a high school degree.

Procedure. This study was approved by the respective institutions and by the participating school boards. An initial meeting was held with all school principals to describe the study. Then, principals interested in having their school participate contacted the research team to learn more about the study. Information sessions for teachers of Grades 1 to 3 were provided along with study information and consent packages. Consenting teachers then received research packages (containing a cover letter, study information, consent forms, and the study questionnaires) to give to parents. Eligibility criteria for children were (a) education in mainstream classrooms in either English or French (25% were in French immersion), (b) no major sensory or physical impairment that would preclude a child from hearing the instructions or completing the assessment tasks, and (c) written informed consent from the child’s teacher and parent and verbal assent from child.

Teachers completed a behavior-rating scale assessing WM in November of each year. Complete data for teacher ratings were available for 506 participants in Year 1 (i.e., 18% or 3% of the 524 consenting teachers did not complete the ratings) and 491 in Year 2 (i.e., 97% retention). Different sets of teachers rated the same set of children in Years 1 and 2. All participating children completed measures of academic achievement in May of each year. We also administered performance-based neuropsychological measures of WM to a representative subsample of 214 participants in May of each year. Complete data for this subsample were available for 211 participants in Year 1 (42% of the full sample; 104 boys,
107 girls) and 202 participants in Year 2 (40% of the full sample; 96% retention in Year 2; 103 boys, 99 girls).

Measures

The study measures included teacher ratings of WM, objective measures of auditory-verbal and visual-spatial WM, and standardized tests of academic achievement (reading and math).

Teacher behavioral ratings of WM. We used the WMRS (Alloway et al., 2008) developed for teachers to easily identify children with WM deficits. The WMRS contains 20 questions and has one subscale. Teachers were asked to rate each of the statements on a 4-point Likert-type scale (1 = not typical at all, 2 = occasionally, 3 = fairly typical, 4 = very typical). Sample items include “To move on to the next step in an activity, needs frequent prompts by teaching staff”; “loses his or her place in complicated activities”; “requires regular repetition of instructions”; and “forgets how to continue an activity that was previously started, despite teacher explanation.”

Objective measures of WM. Measures consisted of the WISC-IV Digit Span Forward and Backward subtests (Wechsler, 2003) and the Wide Range Assessment of Memory and Learning 2 (WRAML-II) Finger Windows Forward and Backward subtests (Adams & Sheslow, 2004). Both have excellent psychometric properties, including high internal-consistency reliability (Digit Span = .87, Wechsler, 2003; Finger Windows = .99, Adams & Sheslow, 2004).

The WISC-IV Digit Span Forward (WISC-IV DSF) is an objective measure of verbal storage-only component of WM (see Gathercole, Pickering, Ambridge, & Wearing, 2004, for construct validity of WISC-IV DSF). The participants heard a sequence of digits at a rate of one digit per second and were asked to immediately repeat the sequence of digits in the exact order it was presented. The length of the sequence started with two digits and became increasingly more difficult (up to a maximum of nine digits) until the participant obtained the required number of errors for discontinuation. The raw score for WISC-IV DSF is the number of correct trials (maximum raw score of 16 correct trials). In contrast, the WISC-IV Digit Span Backward (WISC-IV DSB) subtest is a measure of the storage and processing components of verbal WM (see Gathercole et al., 2004, for construct validity of WISC-IV DSB). During this task, the participants repeated the sequence of digits in reverse order. The raw score of WISC-IV DSB is the number of correct trials (maximum raw score of 14 correct trials).
The Wide Range Assessment of Memory and Learning 2 (WRAML-II) Finger Windows Forward subtest (WRAML-II FWF; Adams & Sheslow, 2004) is a measure of visuospatial storage-only component of WM. For this subtest, the examiner sequentially touched asymmetrically located holes, or windows, using an 8 × 11-inch plastic card at the rate of one per second. Participants sitting behind the card that was held horizontally perpendicular to the work surface repeated the given sequence of windows by placing their fingers through the same windows in the correct order. One point was awarded for each correct sequence. The total number of correct sequences constituted the total raw score for this subtest. Series length ranged from one window to nine windows. In contrast, in the WRAML-II Finger Windows Backward subtest (WRAML-II FWB; Adams & Sheslow, 2004), the participant repeated the given sequence of windows by placing his/her finger through the same windows in the reverse order; it is a measure of the storage and processing components of visual-spatial WM.

**Standardized tests of academic achievement.** We used both standardized (Woodcock–Johnson III Tests of Achievement [WJ-III]; Woodcock, McGrew, & Mather, 2001) and curriculum-based measures (Dynamic Indicators of Early Literacy Skills; Good, Kaminski, Smith, Laimon, & Dill 2001; M-CBM; Mathematics Curriculum-Based Measurement: AIMSweb) to assess academic skills. The WJ-III has well-established psychometric properties (Mather & Woodcock, 2001). Curriculum-based measures of mathematics and reading that were selected for this study have moderate to strong single-probe and multiple-probe reliability (ranging from .65 to .99), as well as moderate to good concurrent and predictive validity (e.g., Foejen, Jiban, & Deno, 2007; Keller-Margulis, Shapiro, & Hitze, 2008; Wayman, Wallace, Wiley, Ticha, & Espin, 2007).

**Reading skills** were measured using the Oral Reading Fluency subtest from the Dynamic Indicators of Early Literacy Skills (DIBELS, 5th ed.; Good et al., 2001) as a measure of oral reading fluency. During this standardized, individually administered test of reading fluency and accuracy, participants were presented with three grade-level passages to read aloud and instructed to try to read each word as accurately and to read as many words of the passage as they can in 1 min. Words omitted, substituted, and hesitations of more than 3 s are scored as errors. The raw median number of correct words (oral reading fluency) was used as the final scores. Good et al. (2001) reported median alternate-form reliability for oral reading of passages to be .94. Extensive research has demonstrated that oral reading fluency is a good indicator of children’s overall reading skills development (e.g., Yovanoff, Duesbery, Alonzo, & Tindal, 2005).

**Mathematics skills** were first assessed using the Math Computation/Math Facts measure from www.AIMSweb.com. This group-administered test assessed basic addition and subtraction. Credit was given to each individual correct digit appearing in the solution to a math fact, which afforded a more precise analysis of a child’s number skills, by capturing emerging and partial skills as well as mastered skills. The total number of correct digits for addition and subtraction problems (raw scores) was used as the final scores. We also employed the Math Fluency and Calculation subtests from the WJ-III (Woodcock et al., 2001). The Math Fluency test measured the ability to solve simple arithmetic facts quickly. This test has a 3-min limit. Participants were presented with 2 pages of math facts having 8 rows with 10 facts of mixed operations in each row. Raw scores were used.

**Data Analysis**

We used the following stepwise approach to the data. We first used CFA using the AMOS statistical program (version 18; Arbuckle, 2009), to assess how well the 20 original items of the WMRS fit the hypothesized one-factor structure of WM in our North American sample. We used a confirmatory approach because previous research found support for this one-factor structure (Alloway et al., 2009). We evaluated model fit using the root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker–Lewis index (TLI) with acceptable model fit indicated by RMSEA values of .08 or lower along with CFI and TLI values of .95 or higher (Yu & Muthén, 2002). Second, because the a priori theoretical WMRS model had poor fit to the sample data, we used AMOS Modification Indexes to sequentially eliminate weak items from the model in an attempt to improve model fit (Byrne, 2010; Sörbom, 1989). Third, we examined the reliability data of the WMRS final items using Cronbach’s alpha coefficients. We used .80 as the cutoff for acceptable reliability, as recommended by Nunnally and Bernstein (1994). Fourth, we conducted partial correlational analyses to explore prospective relationships between the final WMRS model, objective measures of WM, and standardized measures of academic achievement 6 and 18 months later, with age partialled out.

**Results**

**Factor Structure**

Our results suggest that the original 20-item WMRS model has poor fit for Canadian boys and girls, as indicated by all RMSEA values higher than .10 along with CFI and TLI values of .91 and lower (see Table 2 for specific fit statistics). This conclusion is cross-validated with the follow-up data from Year 2. Furthermore, the higher end of the item-total correlation range of the original 20-item model is above .85 for males and females (see Table 2). This suggests that some of the 20 items were redundant. Follow-up
post hoc factor analysis results suggest that a 5-item WMRS factor model fit the data better than the original 20-item WMRS model (as evidenced by all RMSEA values of .05 or lower along with all CFI and TLI values of .95 or higher; see Table 2). This conclusion is cross-validated with data for both males and females and with the follow-up data from Year 2 (see Table 2). Each of these 5 items had significant, positive loadings on the WMRS factor, with standardized loadings homogeneously ranging from .81 to .90 for males and from .76 to .89 for females in Year 1. Similarly, standardized loadings ranged from .82 to .93 for males and from .74 to .88 for females in Year 2 (all ps < .001; see Table 3 for a list of the five final items, their standardized factor loading estimates and $R^2$ values). Thus, the measurement properties of the 5-item WMRS are equivalent for males and females and stable across a 12-month period, spanning two consecutive academic years, with respect to relating the WM construct implied by the 5-item model. As a whole, we consider that the alternative 5-item WMRS model represent the final best-fitting and most parsimonious model to represent the data.

**Reliability**

As summarized in Table 2, internal-consistency data for the alternative 5-item WMRS was high in Year 1 (range = .92-.93) and in Year 2 (range = .90-.93). Item-total correlations were also high in both years, ranging from .57 to .79 in Year 1 and from .57 to .81 in Year 2 (see Table 2). These results converge with the partial correlations between the WMRS scores across both years: WMRS scores in Year 1 are significantly positively correlated with the WMRS scores in Year 2, with age partialled out ($r = .63$ for males and $r = .77$ for females, all $p$ < .001; see Table 4). This is particularly important given that different sets of teachers rated the same set of children in Years 1 and 2.

**Criterion-Related Validity: WMRS and Objective Measures of WM**

We tested the degree to which the alternative five-item WMRS correlated with objective measures of WM measured 6 and 18 months later, with age partialled out. Both boys’ and girls’ baseline WMRS scores were significantly negatively associated with the objective measures of WM (i.e., WISC-IV DSB and WRAML-II FWB) 6 months later, with age partialled out ($r = .63$ for males and $r = .77$ for females, all $p$ < .001; see Table 4). Similarly, baseline WMRS scores were significantly negatively associated with WISC-IV DSB (boys only) and WRAML-II FWB (boys and girls) at the 18-month follow-up, after controlling for age. Although there was no significant relationship between baseline WMRS scores and WISC-IV DSF for both boys and girls at the 6-month follow-up, this relationship became significantly negative at the 18-month follow-up for both sexes, after controlling for age (see Table 4).

### Table 2. Fit of the Original WMRS CFA Model and a Post Hoc Alternative Model by Sex and Year.

<table>
<thead>
<tr>
<th>Group</th>
<th>$n$</th>
<th>$\chi^2$</th>
<th>df</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA [90% CI]</th>
<th>$\alpha$</th>
<th>Interitem correlation range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original 20-item one-factor WMRS model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Year 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole sample</td>
<td>506</td>
<td>1215.64</td>
<td>170</td>
<td>.912</td>
<td>.902</td>
<td>.110 [.105, .116]</td>
<td>.98</td>
<td>(48.87)**</td>
</tr>
<tr>
<td>Female</td>
<td>258</td>
<td>907.05</td>
<td>170</td>
<td>.875</td>
<td>.860</td>
<td>.130 [.122, .138]</td>
<td>.98</td>
<td>(43.86)**</td>
</tr>
<tr>
<td>Male</td>
<td>248</td>
<td>665.40</td>
<td>170</td>
<td>.914</td>
<td>.904</td>
<td>.109 [.100, .117]</td>
<td>.98</td>
<td>(48.87)**</td>
</tr>
<tr>
<td><strong>Year 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole sample</td>
<td>491</td>
<td>1267.85</td>
<td>170</td>
<td>.905</td>
<td>.893</td>
<td>.115 [.109, .121]</td>
<td>.98</td>
<td>(38.89)**</td>
</tr>
<tr>
<td>Female</td>
<td>244</td>
<td>738.76</td>
<td>170</td>
<td>.889</td>
<td>.875</td>
<td>.117 [.109, .126]</td>
<td>.98</td>
<td>(40.86)**</td>
</tr>
<tr>
<td>Male</td>
<td>247</td>
<td>765.70</td>
<td>170</td>
<td>.899</td>
<td>.887</td>
<td>.119 [.111, .128]</td>
<td>.97</td>
<td>(30.92)**</td>
</tr>
<tr>
<td><strong>Post hoc alternative 5-item one-factor WMRS model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Year 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole sample</td>
<td>506</td>
<td>5.60</td>
<td>5</td>
<td>1.000</td>
<td>.999</td>
<td>.015 [.000, .065]</td>
<td>.93</td>
<td>(66.79)**</td>
</tr>
<tr>
<td>Female</td>
<td>258</td>
<td>6.88</td>
<td>5</td>
<td>.998</td>
<td>.996</td>
<td>.038 [.000, .101]</td>
<td>.92</td>
<td>(57.77)**</td>
</tr>
<tr>
<td>Male</td>
<td>248</td>
<td>3.10</td>
<td>5</td>
<td>1.000</td>
<td>1.004</td>
<td>.000 [.000, .068]</td>
<td>.92</td>
<td>(64.77)**</td>
</tr>
<tr>
<td><strong>Year 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole sample</td>
<td>491</td>
<td>7.79</td>
<td>5</td>
<td>.999</td>
<td>.997</td>
<td>.034 [.000, .077]</td>
<td>.93</td>
<td>(67.77)**</td>
</tr>
<tr>
<td>Female</td>
<td>244</td>
<td>6.15</td>
<td>5</td>
<td>.999</td>
<td>.997</td>
<td>.031 [.000, .098]</td>
<td>.90</td>
<td>(57.71)**</td>
</tr>
<tr>
<td>Male</td>
<td>247</td>
<td>1.02</td>
<td>5</td>
<td>1.000</td>
<td>1.001</td>
<td>.000 [.000, .000]</td>
<td>.93</td>
<td>(68.81)**</td>
</tr>
</tbody>
</table>

Note: WMRS = Working Memory Rating Scale; CFA = confirmatory factor analysis; $df$ = degrees of freedom; $\chi^2$ = chi-square fit statistic; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean squared error of approximation; 90% CI = 90% confidence interval for RMSEA; $\alpha$ = Cronbach’s alpha.

**$**p < .01.
Interestingly, boys’ baseline WMRS scores were significantly negatively associated with WRAML-II FWF (controlling for age) at both the 6- and 18-month follow-up, whereas this relationship was nonsignificant for girls at both time points.

**Convergent Validity: WMRS and Academic Achievement**

We conducted partial correlation analyses to examine the longitudinal relationships among the five-item WMRS, reading, and math achievement measures 6 and 18 months later after controlling for age differences. The five-item WMRS factor was significantly correlated with academic achievement measures of reading and math: teachers’ ratings of boys and girls with high levels of WM difficulties on the WMRS in Year 1 were associated with poorer reading and math fluency scores on standardized academic tests 6 and 18 months later (all rs ranging from −.50 to −.61, all ps < .001 for boys; and all rs ranging from −.28 to −.42, all ps at least < .01 for girls; see Table 4).

**Discussion**

The unique purpose of this research was to examine the factor structure, internal consistency, criterion validity, and convergent validity of the WMRS in a large, independent sample of North American elementary schoolchildren within an 18-month longitudinal design, and with an examination of potential gender differences. Our initial attempt at replicating the 20-item original WMRS factor structure was unsuccessful, contrary to predictions. Instead, screening of WM was best described as a combination of five items, all of which are readily observable by teachers: (a) “abandons activities before completion”; (b) “benefits from continued teacher support during lengthy activities”; (c) “does not follow classroom instructions accurately, for example, carries out some but not all steps in an instruction”; (d) “is making poor progress in literacy and math”; and (e) “depends on neighbor to remind them of the current task.” The factor structure of this 5-item WMRS factor model was superior to the 20-item original WMRS factor model for males and females and stable across a 12-month period, spanning two consecutive academic years, with respect to relating the WM construct implied by the 5-item model. The current study is the first published report of CFAs of the original 20 WMRS items taken intrinsically and in an independent sample. Our findings also complement those of Alloway et al. (2009), who followed their exploratory factor analyses with CFAs to evaluate the relative fit of a two-factor model tapping behavioral (i.e., WMRS) and cognitive aspects (i.e., AMWA) of WM using the same sample.

Additional analyses showed that this short 5-item scale had excellent internal consistency, item-total correlations, and cross-year WMRS factor correlations for boys and girls. The internal consistency was similar to the one obtained by Alloway with the original 20 items (Alloway et al., 2009). Cross-year WMRS factor correlations (i.e., between WMRS scores in Year 1 and WMRS scores in Year 2) were high and this is particularly important given that different sets of teachers rated the same set of children in Years 1 and 2.

We then assessed the criterion validity of the five-item WMRS scale. We found that the 5-item WMRS factor was generally moderately and negatively related to objective measures of WM measured 6 and 18 months later. Controlling for age differences, both boys’ and girls’ baseline WMRS scores were associated with measures of the storage/processing components of WM 6 months later. However, only boys’ baseline WMRS scores were significantly negatively associated with an objective measure of storage/processing components of verbal WM at the 18-month follow-up, after controlling for age. Similarly, only boys’ baseline WMRS scores were significantly negatively associated with the storage-only component of visuospatial WM (with age partialled out) at both the 6- and 18-month follow-up, whereas this relationship was

**Table 3. Standardized Factor Loadings for the Post Hoc Alternative 5-Item WMRS Model by Sex and Year.**

<table>
<thead>
<tr>
<th>Item</th>
<th>Year 1</th>
<th></th>
<th></th>
<th>Year 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All F M</td>
<td></td>
<td></td>
<td>All F M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Abandons activities before completion</td>
<td>.82 .78 .82</td>
<td>.67 .62 .67</td>
<td>.84 .74 .86</td>
<td>.70 .55 .74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Benefits from continued teacher support during lengthy activities</td>
<td>.90 .86 .90</td>
<td>.80 .74 .81</td>
<td>.92 .88 .93</td>
<td>.67 .78 .87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Does not follow classroom instructions accurately, for example,</td>
<td>.84 .86 .81</td>
<td>.70 .74 .65</td>
<td>.83 .81 .82</td>
<td>.69 .65 .66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>carries out some but not all steps in an instruction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Is making poor progress in literacy and math</td>
<td>.80 .76 .81</td>
<td>.64 .58 .65</td>
<td>.82 .79 .83</td>
<td>.67 .62 .69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Depends on neighbor to remind them of the current task</td>
<td>.89 .89 .87</td>
<td>.78 .79 .76</td>
<td>.84 .87 .82</td>
<td>.71 .75 .68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: WMRS = Working Memory Rating Scale; All = whole sample; F = females; M = males. All estimates are significant at p < .001.
Table 4. Partial Correlations Between the Five-Item WMRS, Objective Measures of Working Memory, and Academic Achievement in Both Years With Age Partialed Out.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Y1 WMRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Y1 DSF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Y1 DSB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Y1 FWF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Y1 FWB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Y1 DORF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Y1 M-CBM+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Y1 M-CBM–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Y1 WJ-III MF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Y2 WMRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Y2 DSF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Y2 DSB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Y2 FWF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Y2 FWB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Y2 DORF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Y2 M-CBM+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Y2 M-CBM–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Y2 WJ-III MF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Y1 = Year 1; WMRS = Working Memory Rating Scale; DSF = Digit Span Forward; DSB = Digit Span Backward; FWF = Finger Windows Forward; FWB = Finger Windows Backward; DORF = dynamic indicators of early literacy skills; M-CBM+ = math curriculum–based measurement total number of correct digits-addition problems; M-CBM– = math curriculum–based measurement total number of correct digits-subtraction problems; WJ-III MF = Woodcock–Johnson III Tests of Achievement–Math Fluency; Y2 = Year 2. Males data (n = 104) appear below diagonal whereas females data (n = 108) appear above diagonal.

*p < .05. **p < .01. ***p < .001.
nonsignificant for girls at both time points. Our preliminary findings therefore indicate the need to consider sex differences in behavioral ratings of WM. They are also consistent with the evidence documenting gender differences in the performance of WM in elementary school-aged children perhaps because of a slower maturation of the prefrontal cortex in boys than girls (Vuontela et al., 2003).

Our findings also suggest excellent predictive convergent validity of the five-item WMRS factor model with standardized measures of academic measures. We found that teachers’ ratings of boys and girls with high levels of WM difficulties on the five-item WMRS at the beginning of Year 1 were associated with poorer reading and math fluency scores on standardized academic tests 6 and 18 months later, with age partialed out. These preliminary results are consistent with the extensive research on the association between objective assessments of WM and academic achievement (Cowan & Alloway, 2009; Gathercole & Pickering, 2000) and support other studies showing that WM deficits negatively affect students’ behavioral performance in several classroom activities, including remembering lengthy activities, keeping track of progress in multistep tasks, and completing activities previously started while simultaneously storing and processing new information (Gathercole, Lamont, et al., 2006).

Many other questions remain to be answered. First, our findings are limited to a specific age range (6-9) and so may not be generalizable to other age groups. It will therefore be important to replicate the study findings to confirm the findings with regard to the factor structure and internal consistency in both clinical and community samples and using longer prospective studies. For example, it will be essential to test and refine the WMRS for older populations of middle and high school students given that WM is critical to many classroom activities (e.g., remembering which materials are needed for the next class, following directions, taking notes, writing essays, listening to lectures; Dehn, 2008). Second, it will also be crucial to examine the incremental validity of the short WMRS versus other objective and behavioral measures of WM. It is still unclear whether the ratings of WM provide useful information over and above objective assessments of WM. Third, given that children with ADHD are known to exhibit deficits in multiple components of WM (Martinussen, Hayden, Hogg-Johnson, & Tannock, 2005), it will also be important in future studies to explore the factor structure of the short WMRS with that of symptoms of ADHD and other correlates of ADHD, including cognitive functions, motivational processes, personality trait profiles, and daily impairment. Furthermore, the relationship between WMRS and other types of psychopathology remains unclear (e.g., internalizing disorders). Finally, future studies may include measures of intellectual abilities. However, evidence suggests that the relationship between WM and academic achievement cannot be accounted for by differences in general intellectual abilities (Cain, Oakhill, & Bryant, 2004; Gathercole, Alloway, Willis, & Adams, 2006).

In summary, WM is not a well-understood concept in education and is typically not included in general teacher training (Gathercole, Lamont, et al., 2006). Teachers are also currently unlikely to recognize WM impairments given the lack of efficient, psychometrically sound, and ecologically valid screening tools to identify students with WM impairment in the classroom. The few available published scales assessing WM have important shortcomings. The unique purpose of this research was to examine the factor structure, internal consistency, criterion validity, and convergent validity of the WMRS in a large, independent sample of North American elementary schoolchildren within an 18-month longitudinal design, and with an examination of potential gender differences. Results showed that the short five-item WMRS had a clearly identified factor structure and produced reliable and valid scores for boys and girls across both time points. Test validation is an ongoing, dynamic process, not an end state (American Education Research Association, American Psychological Association, & National Council on Measurement in Education, 1999); however, the results suggest that the short version of the WMRS is promising. Further research is needed on the measure before it can be recommended for wide use in research or clinical settings. The five-item WMRS may eventually provide teachers with a useful and time-effective method to screen for WM deficits at school.

Authors’ Note

Portions of this paper were presented at the 2011 International Society for Research on Child and Adolescent Psychopathology Biennial Conference, Chicago.

Acknowledgments

We express appreciation to the families and schools from the Peterborough Northumberland Clarington Catholic School Board and the Kawartha Pine Ridge District School Board that participated in our study. The dedicated assistance of Min-Na Hockenberry, Marisa Catapang, Peter Chaban, Sarah Anne Gray, and Danielle Pigan is also gratefully acknowledged.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Dr. Tannock has been a consultant for Eli Lilly, McNeil, Purdue, and Shire, for which she has received honoraria that are donated to The Hospital for Sick Children’s Foundation to support ADHD research. Dr. Normand reports no financial relationships with commercial interests.
Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was financially supported by a grant from the Social Sciences and Humanities Research Council of Canada (SSHRC #410-2008-1052) and the Canada Research Chair Program (RT).

References


**Author Biographies**

Sébastien Normand, PhD, is a professor of psychology at the University of Quebec in Outaouais, Gatineau, Canada. His clinical and research interests include social, family, and school functioning of children with ADHD, evidence-based assessment and psychosocial treatments of ADHD, and the sociocultural aspects of ADHD.

Rosemary Tannock, PhD, holds a Canada Research Chair in Special Education and Adaptive Technology at the Ontario Institute for Studies in Education, University of Toronto; is a senior scientist at the Hospital for Sick Children; and is professor of psychiatry at the University of Toronto, Canada. She is also a member of the DSM-5 Work Group on ADHD and cross-appointed to the Neurodevelopmental Disorders Group to lead the work on learning disabilities. Her research investigates the causes and treatment of ADHD, with a specific focus on its cognitive manifestations and overlap with learning disabilities.