

Measuring working memory capacity in children using adaptive tasks: Example validation of an adaptive complex span

Corentin Gonthier¹ · Alexandre Aubry² · Béatrice Bourdin²

Published online: 22 June 2017
© Psychonomic Society, Inc. 2017

Abstract Working memory tasks designed for children usually present trials in order of ascending difficulty, with testing discontinued when the child fails a particular level. Unfortunately, this procedure comes with a number of issues, such as decreased engagement from high-ability children, vulnerability of the scores to temporary mind-wandering, and large between-subjects variations in number of trials, testing time, and proactive interference. To circumvent these problems, the goal of the present study was to demonstrate the feasibility of assessing working memory using an adaptive testing procedure. The principle of adaptive testing is to dynamically adjust the level of difficulty as the task progresses to match the participant's ability. We used this method to develop an adaptive complex span task (the ACCES) comprising verbal and visuo-spatial subtests. The task presents a fixed number of trials to all participants, allows for partial credit scoring, and can be used with children regardless of ability level. The ACCES demonstrated satisfying psychometric properties in a sample of 268 children aged 8–13 years, confirming the feasibility of using adaptive tasks to measure working memory capacity in children. A free-to-use implementation of the ACCES is provided.

Keywords Working memory capacity · Complex span task · Computerized adaptive testing · French validation · Child cognition

There are many reasons to be interested in working memory in children. The number of items that can be held in memory while performing concurrent processing increases with age, following a developmental trajectory distinct from that of short-term memory (Gathercole, 1998, 1999). Working memory capacity (WMC) seems to factor in the development of complex abilities, including language (Cain, 2006; Linck, Osthus, Koeth, & Bunting, 2013; Seigneuric & Ehrlich 2005) and mathematics (Swanson, 2011). WMC is also a good predictor of scholastic achievement (Gathercole, Brown, & Pickering, 2003; Gathercole & Pickering, 2000; Gathercole, Pickering, Knight, & Stegmann, 2004; Hitch, Towse & Hutton, 2001), perhaps even better than intelligence (Alloway & Alloway, 2010). Working memory impairments are found in a number of developmental disorders such as dyslexia (Jeffries & Everatt, 2004), specific language impairment (Montgomery, Magimairaj, & Finney, 2010) and attention deficit/hyperactivity disorder (ADHD; Alloway, 2011; Holmes et al., 2014), whereas intellectually precocious children tend to demonstrate higher WMC (Hoard, Geary, Byrd-Craven, & Nugent, 2008; Kornmann, Zettler, Kammerer, Gerjets, & Trautwein, 2015; Swanson, 2006).

On the other hand, developing appropriate measures of WMC in children comes with a challenge. Because WMC increases through the course of development and because it demonstrates significant variability from one group to another, suitable tasks have to cover a wide range of ability levels to allow for longitudinal studies and cross-group comparisons. In other words, tests designed to assess working memory in children need to include items appropriate for both very low

✉ Corentin Gonthier
corentin.gonthier@univ-rennes2.fr

¹ Université Rennes 2, CRPCC EA 1285, Rennes, France

² Université de Picardie Jules Verne, CRP-CPO EA 7273, Amiens, France

and high WMC, without being overly difficult or frustrating for children in the lower ability range. This issue is usually addressed by presenting trials in ascending order of difficulty (i.e., with progressively more items to memorize) and discontinuing testing when the child fails a particular level. A low-ability child will only complete the first trials with a small set size, whereas a high-ability child will progress through easy, medium and difficult trials. This procedure has been implemented in the vast majority of working memory tests designed for use with children, including the digit span subtests in Wechsler's intelligence scale for children (WISC; Wechsler, 2014), the AWMA (Alloway, 2007), the WMTB-C (Pickering & Gathercole, 2001) and the TOMAL (Reynolds & Voress, 2007). It has also been used by a large number of researchers developing their own working memory tasks (e.g., Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009; Bayliss, Jarrold, Baddeley, Gunn, & Leigh, 2005; Case, Kurland, & Goldberg, 1982).

Main shortcomings of the ascending procedure

Despite its ubiquity, the procedure of presenting trials in ascending order of difficulty with a discontinue criterion raises numerous issues. The first major issue is related to the number of trials performed by a testee, which can vary unpredictably from one child to the next. For example, children performing the counting span subtest of the AWMA (Alloway, 2007) may complete between three and 42 trials depending on their performance; the picture is similar for other common working memory tests. This variability makes it difficult to plan a testing time and to use a task for group testing. It also means that children will be exposed to a variable number of stimuli: in the same subtest of the AWMA, children may have to memorize between three and 168 items. Thus, different children will have to cope with different levels of fatigue, but also with different levels of proactive interference. This variation in proactive interference is especially worrying: the amount of proactive interference in a working memory test influences its predictive power (Lustig, May, & Hafer, 2001), and resistance to proactive interference varies as a function of age (Loosli, Rahm, Unterrainer, Weiller, & Kaller, 2014) and as a function of WMC (Kane & Engle, 2000; Unsworth & Engle, 2007). In other words, WMC scores obtained with the ascending procedure are partly confounded with proactive interference and individual differences in resistance to proactive interference, but the exact impact of proactive interference varies unpredictably between children as a function of their age and ability level. Importantly, these issues are exacerbated when a task battery comprises multiple subtests: in the case of the AWMA, the same structure repeats over 12 subtests, further compounding differences in testing time, fatigue, and

proactive interference. These issues may also transfer to any tasks completed after the WMC assessment.

Presenting trials in order of ascending difficulty also raises several problems. This approach confounds the level of difficulty with the buildup of proactive interference, potentially worsening the aforementioned issue (Conway et al., 2005). The ascending order of difficulty can get frustrating for high-ability children, who have to work their way through a large number of trials that are much too easy before reaching appropriate levels (e.g., Vandierendonck, De Vooght, & Van der Goten 1998a). This may be especially problematic for gifted children with very high WMC and could lead to significantly underestimating their memory spans. Again, this is even truer when the test includes multiple subtests where the child always has to start over at the easiest level. This procedure also makes the set size of the next trial inherently predictable, allowing participants to use strategies based on anticipating the number of items to memorize. Working memory tests designed for adults often present trials in random order to avoid the use of such strategies (Conway et al., 2005). Another subtle side effect of this procedure is that it can convey a feeling of failure: a subtest is always discontinued when the child fails multiple trials in a row, which can give children the impression that their performance is inadequate. This feeling, too, may build up over multiple subtests.

Lastly, the psychometric properties of WMC scores obtained with this procedure give cause for concern (see Woods et al., 2011). The fact that participants complete different numbers of trials may inflate the error variance of WMC scores and reduce the sensitivity of the test (Woods et al., 2011). Discriminating power is also limited by the fact that participants only complete a few trials appropriate for their span level (Weiss, 1974): most trials are too easy and bring little information for children with high WMC, whereas the test is cut short quickly for children with low WMC. For example, a WMC assessment using the WISC may include as few as four trials (Wechsler, 2014). Furthermore, discontinuing testing when the child fails a particular level usually implies the use of all-or-nothing scoring (i.e., scoring trials as failed unless all items are correctly recalled in the appropriate serial position; Conway et al., 2005), but this scoring method presents limited sensitivity to individual differences in WMC (Friedman & Miyake, 2005; Unsworth & Engle, 2007), even in children (St Clair-Thompson & Sykes, 2010).

Moreover, when a trial is considered failed provided a single item is not recalled correctly, mind-wandering or suboptimal effort on even one trial can lead to discontinuing the test and drastically altering the estimated WMC of a child. This may be equally problematic for children with limited attention, and for high-ability children who become disinterested in the test during the easy first few trials. Mind-wandering is known to play an important role in working memory tasks (Mrazek et al., 2012), and attentional lapses on easy trials do

not seem related to individual differences in WMC (Adam, Mance, Fukuda, & Vogel, 2015), which suggests that they could seriously bias estimates obtained with an ascending procedure.

There are few satisfying alternatives to the classic strategy of presenting trials in ascending order of difficulty with a discontinue criterion. One could imagine having children always perform all trials (e.g., Woods et al., 2011), as is usually the case in tasks designed for adults (e.g., Redick et al., 2012). However, this would confront low-ability children with recurring failure on many difficult trials, and many trials would still be of inappropriate difficulty for any given participant. Another possibility, and the focus of the present study, is to use adaptive testing.

Developing an adaptive working memory test

Adaptive testing is a tried technique (Lord, 1968; Weiss, 1974) with a straightforward rationale. Its core principle is to dynamically adjust the level of difficulty of a task throughout the testing session, as a function of participant performance. The ability level of a participant can be re-estimated after each trial, and the difficulty of the following trial can then be adjusted accordingly. In practice, the optimal level of difficulty for a trial is the closest to the examinee's actual level, because trials that are too easy or too difficult provide little information (Lord, 1968). Thus, children with high performance should be confronted with progressively more difficult trials and children with low performance with easier trials. This procedure is common in fields such as formative assessment, and adaptive tests often demonstrate psychometric properties that are comparable or even better than their non-adaptive counterparts (Weiss, 1982). Adaptive tasks are also routinely employed for working memory training (e.g., Holmes, Gathercole, & Dunning, 2009; Klingberg, Forssberg, & Westerberg, 2002). However, adaptive testing has to our knowledge never been used to develop a validated working memory test.

The adaptive procedure presents several potential advantages for working memory assessment in children. Because difficulty is adjusted to match the participant's skills, the same adaptive working memory task could be used throughout development as well as with clinical samples, regardless of the child's actual ability. Importantly, all participants could also be confronted with the same number of trials matching their WMC, without ever discontinuing the task. Equalizing the number of trials performed by all participants would largely smooth differences in frustration, fatigue, and proactive interference between children with different ability levels (see Weiss, 1974). Measuring WMC with the same number of trials for all participants could also help provide more precise estimates, especially for extreme ability levels (Weiss, 1974). Lastly, one practical advantage is that the very nature of

memory span tests makes adaptive procedures particularly easy to implement. Adaptive tasks usually require item response modeling and large participant samples to calibrate the difficulty of items, but in a memory span task the difficulty of a trial can simply be indexed as the number of items to memorize.¹

Based on this reasoning, the objective of the present study was twofold: (1) demonstrating that adaptive testing can be used to obtain a sensitive measure of WMC in children throughout development, and (2) validating an adaptive working memory test for use with French-speaking children.² To achieve these two objectives, we constructed the Adaptive Composite Complex Span (ACCES). The ACCES was adapted from the Composite Complex Span (CCS; Gonthier et al., 2016), a French-validated working memory test designed for adults and devoid of an adaptive component. The ACCES includes three computerized subtests: a reading span, a symmetry span, and an operation span (see Unsworth, Heitz, Schrock, & Engle, 2005). The use of three shortened subtests involving numerical, verbal and visuo-spatial materials makes it possible to extract a single domain-general measure of WMC within a reasonable testing time (see Foster et al., 2015; Gonthier et al., 2016; Oswald, McAbee, Redick, & Hambrick, 2015). All three subtests are based on the complex span procedure: participants have to alternate between memorizing items and solving simple processing tasks. Complex spans constitute the main paradigm for the study of individual differences in WMC in adult participants (Conway et al., 2005; Redick et al., 2012); they have also been used in many studies with children (for a review, see Tillman, 2011), and demonstrate excellent psychometric properties in both adults (Redick et al., 2012) and children (e.g., Hitch et al., 2001). The three subtests of the ACCES were directly adapted from the automated versions developed by Engle and colleagues, which have been used with over 6,000 participants (see Redick et al., 2012; Unsworth et al., 2005). All materials for the ACCES – including the stimuli, an open source implementation of the task script under PsychoPy (Peirce, 2007), and all validation data – can be accessed via the Open Science Framework platform at <https://osf.io/bk7pm/>.

¹ One study has investigated the use of item response theory in the context of various working memory tasks for children (Vock & Holling, 2008). The authors found that a one-parameter logistic model (1-PL, i.e., a model with a single parameter for trial difficulty) summarized performance patterns on almost all trials. Trial difficulty was also found to increase monotonically as a function of the number of items to memorize, although trials with the same set size did vary somewhat in difficulty. Thus, the set size of a trial yields a good approximation of its actual difficulty.

² There are currently few options to assess a child's WMC in the French language. Clinicians seem to rely mostly on the backward digit span included in the WISC (Wechsler, 2014), but its use as a working memory task is controversial at best, even in children (e.g., Conway et al., 2005; Hutton & Towse, 2001). Researchers sometimes use a French version of the reading span (Delaloye, Ludwig, Borella, Chicherio, & De Ribaupierre, 2008), but this task differs markedly from English-speaking versions (see Gonthier, Thomassin, & Roulin, 2016).

Method

The Adaptive Composite Complex Span (ACCES)

The ACCES was composed of three computerized subtests with identical structure (based on the versions used by Unsworth et al., 2005). In each trial, participants had to alternate between solving simple processing problems and memorizing unrelated stimuli presented after each problem. At the end of a trial, participants had unlimited time to recall all to-be-remembered stimuli in the correct order by clicking the appropriate options in a response grid. The reading span subtest required participants to alternate between deciding whether sentences are correct and memorizing digits; the symmetry span subtest required deciding whether geometric shapes are symmetrical and memorizing spatial locations within a grid; the operation span subtest required deciding whether mathematical operations are correct and memorizing letters. The three subtests are illustrated in Fig. 1. The number of problem-stimulus pairs within a trial varied between two and eight; all participants completed six trials for each subtest.

Procedure

Each subtest began with a training phase including three consecutive practice sessions (Unsworth et al., 2005). Participants received feedback after each trial

during the training phase; to ensure correct understanding of the instructions, each practice session was repeated until the participant reached 70% accuracy. In the first practice session, participants practiced memorizing stimuli and recalling them in the same order (without a concurrent processing demand), by completing two trials of set sizes two and three. In the second session, they practiced the processing task (e.g., deciding whether sentences were correct) by completing ten problems. Response times on this session were recorded and served to define the maximal time allowed for each problem in the actual task, which prevented participants from freely rehearsing to-be-remembered stimuli (see Unsworth et al., 2005). The time limit for answering a problem was computed as the participant's median response time plus 2.5 times the median absolute deviation (with a minimum of 2 s and a maximum of 8 s). In the final practice session, participants were trained to perform the actual task, including both memory and processing demands, by completing two trials of set sizes two and three. They then proceeded to the actual task, with the instructions emphasizing that processing and memory demands were equally important. Participants completed six trials in the actual task. After a short break, they proceeded to the practice session of the next subtest, and so on. Completing the whole ACCES task required approximately 30 min.

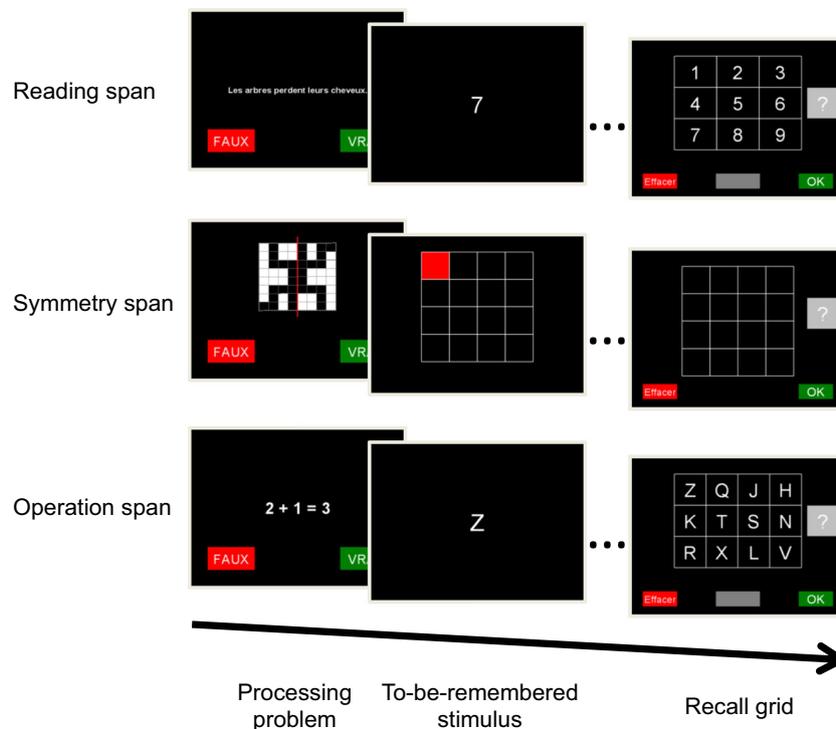


Fig. 1 Illustration of the reading span, symmetry span, and operation span subtests

Stimuli

For the reading span, to-be-remembered stimuli were single digits (between 1 and 9) presented for 800 ms each and followed by an 800-ms inter-stimulus interval. Sentences were adapted from a French-speaking version of the reading span designed for use with children (Delaloye et al., 2008). Incorrect sentences were constructed to be factually wrong, but syntactically correct (e.g., *Les girafes sont très petites* [Giraffes are very small]). Sentences were kept as short as possible (5.3 words on average) to limit the effect of reading skills on WMC performance. For the symmetry span, to-be-remembered stimuli were spatial locations in a 4×4 matrix. Each spatial location was displayed to the participant by coloring one square of the matrix in red for 1,600 ms, with an 800-ms inter-stimulus interval. The images used for the symmetry judgment task were adapted from an adult version of the symmetry span (Unsworth et al., 2005). To make the task easier for children, asymmetrical images were slightly altered to make the violations of symmetry more obvious; the vertical axis of symmetry was also materialized by drawing a vertical red line in the middle of each image (see Fig. 1). For the operation span, to-be-remembered stimuli were consonants chosen to be visually and phonologically distinctive in French (*H, J, K, L, N, Q, R, S, T, V, X, and Z*), each displayed for 800 ms with an 800-ms inter-stimulus interval. All operations were additions of the form $4 + 3 = 7$; the first term of the addition was a two-digit number 50% of the time. Incorrect operations retained the same form, but the result was off by one or more units in either direction (e.g., $4 + 3 = 6$ or $4 + 3 = 9$). Numbers in the operations never exceeded 40.

In all subtests, trials were constructed so that the same stimulus never appeared twice and stimuli did not form meaningful sequences. Processing problems were chosen to be relatively easy, even for the younger children: indeed, prior research has shown that the processing task in a complex span does not need to be especially difficult as long as it diverts the participant's attention from rehearsing to-be-remembered stimuli (Conlin, Gathercole, & Adams, 2005; Lépine, Barrouillet, & Camos, 2005; Turner & Engle, 1989). To ensure that the difficulty was appropriate, sentences and operations were pre-tested in a class of fourth-graders (*CFMI*, with children aged 9–10 years; $N = 17$). Only problems answered correctly by more than 85% of children were included in the task.

Adaptive procedure

The first trial of each subtest was identical for all participants³ and was selected to be of moderate difficulty. Starting with a moderately difficult item is typical in adaptive testing, because

Table 1 Change in the level of difficulty of the next trial as a function of the set size and number of stimuli correctly recalled on the current trial

| Current set size | Number of stimuli correctly recalled | | | | | | | | |
|------------------|--------------------------------------|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 8 | – | – | – | – | – | = | = | = | = |
| 7 | – | – | – | – | = | = | = | + | |
| 6 | – | – | – | – | = | = | + | | |
| 5 | – | – | – | = | = | + | | | |
| 4 | – | – | – | = | + | | | | |
| 3 | – | – | = | + | | | | | |
| 2 | = | = | + | | | | | | |

The symbols –, =, and + indicate that the set size of the next trial will be decreased by one, kept constant, or increased by one, respectively

this is the optimal solution to adjust the difficulty of the test to the ability level of all participants in as few trials as possible (see Weiss, 1974). The starting set size was fixed to 4 for the reading span and operation span, and 3 for the more difficult symmetry span. These values were determined based on data collected with adults (in the adult CCS, perfect recall is attained by more than 50% of adult participants on set size 5 in the reading span and operation span and set size 4 in the symmetry span; Gonthier et al., 2016).

For all subsequent trials, difficulty was adjusted on a trial-by-trial basis: depending on the proportion of items correctly recalled on a trial, the set size of the following trial was either increased by one (up to a limit of 8), decreased by one (down to a limit of 2), or kept constant. Difficulty was increased when a participant recalled all to-be-remembered stimuli in the correct order (as in the traditional ascending procedure), decreased when a participant recalled less than 60% of to-be-remembered stimuli, and kept constant otherwise. These cut-offs are summarized in Table 1 for all possible set sizes. To standardize the task as much as possible for all participants, stimuli within a trial were always the same, and the set size of a given trial was only altered by changing the number of items to display: for example, participants completing the second trial in the operation span subtest had to recall either three letters (*V, R, Q*), four letters (*V, R, Q, S*), or five letters (*V, R, Q, S, J*), depending on their performance on the first trial. The resulting trial structure formed a pyramid with constant step size (Weiss, 1974); an illustration of this structure is displayed in Fig. 2. This design is typical in adaptive testing: the pyramidal structure is simple, covers a wide range of ability levels with relatively few items, and ensures that all participants complete the same number of trials; the constant step size

³ In adaptive testing, the difficulty of the first trial is sometimes adjusted as a function of performance during practice trials. This procedure would not be appropriate here: because WMC scores are computed over all six trials in a subtest, starting with different set sizes would give practice trials an influence over final WMC estimates.

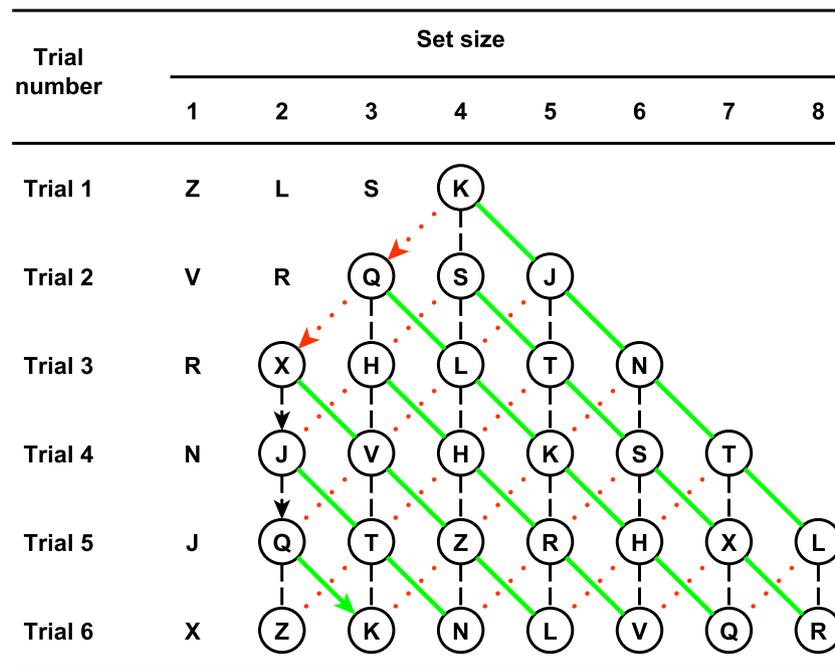


Fig. 2 Stimuli and trial structure of the operation span. All participants start at Trial 1 with set size 4 (four letters to memorize: Z, L, S, K). As a function of performance, the set size of subsequent trials is either decreased (dotted red lines), kept constant (dashed black lines), or

increased (solid green lines). The arrows represent the example trajectory of a participant completing six trials of set sizes 4, 3, 2, 2, 2, and 3

limits the influence of chance successes and failures on final estimates.

Scoring

A recall score was computed in each subtest using the partial-credit load scoring method (Conway et al., 2005), which consists of tallying the total number of stimuli correctly recalled in the correct order. As in the adult version of the CCS, recall scores computed on each subtest were not designed to be used as separate estimates of verbal and visuo-spatial WMC, because the low number of trials in each subtest does not make for sufficient psychometric qualities (Gonthier et al., 2016). The three recall scores were instead standardized and averaged to provide a single domain-general WMC score. Partial credit scoring was chosen because it is appropriately sensitive to individual differences in WMC (Friedman & Miyake, 2005; St Clair-Thompson & Sykes, 2010; Unsworth & Engle, 2007). Unit scoring (tallying the proportion of items correctly recalled per trial, rather than the number of items correctly recalled; Conway et al., 2005), the more common alternative to load scoring, cannot be used in the context of an adaptive task since it would disregard the level of difficulty of trials actually completed by the participant. Other scoring methods can be used in the context of an adaptive task, such as the set size reached on the final trial or the average set size of trials completed by the participant (Lord, 1968; Weiss, 1974); these methods were not retained for the final version

of the task because they demonstrated slightly poorer psychometric qualities than partial-credit load scoring in the validation sample.⁴

Validation sample

Psychometric properties of the ACCES were tested in a sample of 268 children aged 8–13 years. These children were recruited in French schools from the third grade to the seventh grade [*CE2* to *5ème*]. Demographic characteristics of the sample are summarized in Table 2. Informed consent was requested from the legal guardians of all children in a class. All children for whom consent was obtained completed the protocol; there were no exclusion criteria. Children performed the tasks in groups of 12–25 in a school computer room. They completed the three subtests of the ACCES in a first testing session. At the end of the task, children were required to rate the subjective difficulty of each subtest on a scale from 1 (*very easy*) to 5 (*very difficult*). To test for convergent validity, all children then completed Raven's Standard Progressive Matrices (SPM; Raven, Raven, & Court, 2000) in a second session set approximately 2 weeks after the first. Convergent

⁴ For reference, the following scoring methods were examined: average set size of trials in a subtest, highest set size reached in a subtest, and final set size reached in a subtest. In the validation sample, composite WMC scores calculated using these methods had test-retest reliabilities of .64, .63, and .63, respectively; correlations with Raven's progressive matrices were .27, .27, and .31.

Table 2 Characteristics of the sample as a function of grade

| Grade | French designation | Number of participants | Mean age (years) | Gender ratio |
|-----------|--------------------|------------------------|------------------|--------------|
| 3rd grade | CE2 | 36 | 8.68 | 58% male |
| 4th grade | CM1 | 33 | 9.83 | 39% male |
| 5th grade | CM2 | 56 | 10.76 | 48% male |
| 6th grade | 6ème | 74 | 11.76 | 53% male |
| 7th grade | 5ème | 69 | 12.89 | 46% male |

validity was not tested with another working memory task due to the lack of suitable WMC assessments for French-speaking children. To allow for an examination of test-retest stability, the 143 children from the sixth grade and seventh grade also completed the ACCES a second time during the second session.

Results

Descriptive statistics and discriminating power

The median completion time for the ACCES was 31 min (median absolute deviation = 3 min). Descriptive statistics are displayed in Table 3 for recall scores. Recall scores on each subtest increased with age (as indexed by grade; all $p < .001$), but they approximately followed a normal distribution at all ages, as indicated by skewness and kurtosis coefficients close to zero; this was also true for composite WMC scores. In other words, no floor effect or ceiling effect was apparent in the data for any of the age groups. Descriptive statistics for performance on processing problems are displayed in Table 4. Accuracy increased with age and response times decreased with age (all $p < .001$), but the difficulty of the processing problems was appropriate even for the youngest children who answered correctly about 85% of the problems on average (Unsworth et al., 2005).

The percentage of trials performed at each set size was also examined to confirm that the adaptive procedure appropriately regulated the difficulty of the test. This step was necessary to ensure that the cutoff criteria used to adjust the difficulty level were appropriate: for example, an excessively stringent criterion to increase set size would manifest as a very small proportion of trials completed above the starting set size. Results are displayed in Fig. 3. The distribution of levels of difficulty was approximately normal for all age groups: in other words, participants mostly performed trials of intermediate difficulty (set sizes of 4–5 for the reading span and operation span and 3–4 for the symmetry span), but a fair percentage of trials was completed at more extreme set sizes. The distribution had a tendency to shift towards more difficult trials for older children. These results suggest that the adaptive procedure functioned correctly.

Descriptive statistics for subjective difficulty of the ACCES are displayed in Table 5. Subjective difficulty was relatively constant as a function of age: there was no effect of age for the reading span and operation span subtests (both $p > .20$). There was an effect of age on subjective difficulty for the symmetry span subtest ($p = .029$), but it was driven by the youngest (third grade) children rating the task as *easier* than the other groups; the other age groups did not differ in their subjective difficulty ratings of the symmetry span ($p = .304$). On average, children of all age groups rated the difficulty of the task between 2 (*relatively easy*) and 3 (*medium difficulty*) out of 5.

Reliability and validity

Two-way correlations between individual scores across subtests are summarized in Table 6. Overall, the three subtests demonstrated satisfying cross-correlations, in the same range as the values observed with the adult version of the task (Gonthier et al., 2016). Performance on all three subtests also strongly correlated with the composite WMC score, as expected. Although conventional factor analyses could not be run at the trial level (because participants completed different versions of each trial), this pattern of correlations suggests that the domain-general composite WMC score adequately summarized performance on the ACCES. All subsequent analyses were performed on composite WMC scores, as the short individual subtests were not designed to provide reliable scores that could be used separately.

Like factor analyses at the trial level, conventional indices of internal consistency such as Cronbach's alpha could not be computed due to participants completing different versions of each trial. Test-retest reliability of the composite WMC score was acceptable, $r(141) = .70$, indicating reliability of the ACCES as a whole. On average, children did not perform higher on the second session than on the first session, $t(142) = 0.23$, $p = .821$, revealing no practice effect. Composite WMC scores demonstrated the expected correlation with Raven's SPM, $r(266) = .34$, $p < .001$, very close to the estimated population value of the correlation between WMC and fluid intelligence ($r = .36$; Ackerman, Beier, & Boyle, 2005), thus confirming concurrent validity. The increase of composite WMC scores with age (as indexed by

Table 3 Descriptive statistics for recall scores on the ACCES as a function of grade

| Measure | 3rd grade | | | 4th grade | | | 5th grade | | | 6th grade | | | 7th grade | | |
|----------------------|--------------|-------|----------|--------------|-------|----------|--------------|-------|----------|--------------|-------|----------|--------------|-------|----------|
| | Mean (SD) | Skew | Kurtosis |
| Reading span score | 13.89 (5.16) | -0.31 | 0.29 | 15.85 (5.07) | 0.07 | 0.02 | 19.46 (5.61) | 0.44 | -0.21 | 18.79 (6.41) | -0.28 | -0.17 | 19.26 (6.16) | -0.24 | -0.28 |
| Symmetry span score | 12.31 (4.52) | -0.81 | 0.49 | 14.00 (5.06) | -0.04 | -0.98 | 14.45 (3.59) | 0.12 | 1.68 | 15.66 (4.75) | -0.44 | 0.22 | 15.87 (4.35) | -0.35 | 0.55 |
| Operation span score | 17.17 (5.01) | -0.04 | -1.11 | 16.42 (6.30) | -0.39 | -0.17 | 20.46 (5.10) | -0.46 | -0.12 | 19.99 (5.75) | -0.63 | 0.76 | 20.43 (6.22) | -0.49 | -0.49 |
| Composite WMC score | -0.54 (0.62) | -0.03 | -0.79 | -0.35 (0.83) | -0.42 | 0.16 | 0.08 (0.70) | -0.04 | 0.05 | 0.13 (0.81) | -0.36 | 0.36 | 0.19 (0.77) | -0.31 | 0.34 |

Span score = total number of stimuli recalled in a subtest; Composite WMC score = average of the three span scores after standardization. Possible values for span scores ranged from 0 to 38. Because the composite WMC score was computed as the average of standardized scores, the value 0 corresponds to the grand mean of the sample

grade) was significant, $F(4, 263) = 8.13, p < .001, \eta^2_p = .11$; WMC appeared to increase during early years and reached a plateau for children around 11 years old (see Table 3), compatible with prior studies (e.g., Gathercole, Pickering, Ambridge, & Wearing, 2004; Thaler et al., 2013).

Discussion

The ACCES demonstrated satisfying psychometric properties. The task retained sensitivity at all ages, as indicated by the normal distributions of scores and trial set sizes; there was no floor effect for younger children or ceiling effect for older children. Participants also subjectively rated the difficulty of the task as appropriate at all ages. ACCES scores were relatively stable in time, and the validity of the task was confirmed by its significant correlation with Raven's SPM and by the

expected developmental trajectory with a ceiling around 11–12 years (Gathercole et al., 2004; Thaler et al., 2013). In sum, results from the validation sample indicate that the ACCES can be used to obtain an adequate estimate of domain-general WMC for children of different ages and levels of ability. Thus, the adaptive procedure can be used to appropriately measure WMC in children.

Its adaptive nature gave the ACCES multiple advantages over traditional working memory assessments based on the ascending procedure. In the validation sample, the same task could be used with all children from 8 to 13 years old with comparable discriminating power; in addition to subjective difficulty ratings, anecdotal reports from children involved in the validation effort confirmed that the task was not deemed particularly frustrating or boring. Despite the task including multiple subtests, this was achieved with a constant number of 18 trials for all children, meaning that they were all exposed

Table 4 Descriptive statistics for processing problems on the ACCES as a function of grade

| Measure | 3rd grade | | 4th grade | | 5th grade | | 6th grade | | 7th grade | |
|---------------------|-----------|------|-----------|------|-----------|------|-----------|------|-----------|------|
| | Mean | SD |
| Reading span | | | | | | | | | | |
| Processing accuracy | 0.84 | 0.13 | 0.82 | 0.17 | 0.91 | 0.08 | 0.92 | 0.07 | 0.92 | 0.07 |
| Processing time | 4.63 | 0.68 | 4.32 | 0.99 | 3.51 | 0.81 | 3.37 | 0.74 | 2.96 | 0.69 |
| Symmetry span | | | | | | | | | | |
| Processing accuracy | 0.89 | 0.15 | 0.83 | 0.18 | 0.87 | 0.15 | 0.92 | 0.12 | 0.93 | 0.09 |
| Processing time | 2.96 | 0.77 | 2.66 | 0.75 | 2.54 | 0.66 | 2.34 | 0.64 | 2.41 | 0.84 |
| Operation span | | | | | | | | | | |
| Processing accuracy | 0.87 | 0.16 | 0.84 | 0.18 | 0.93 | 0.09 | 0.92 | 0.08 | 0.93 | 0.07 |
| Processing time | 3.31 | 0.71 | 2.86 | 0.73 | 2.57 | 0.63 | 2.59 | 0.62 | 2.39 | 0.66 |

Processing accuracy = percentage of processing problems correctly answered; Processing time = median response time on processing problems

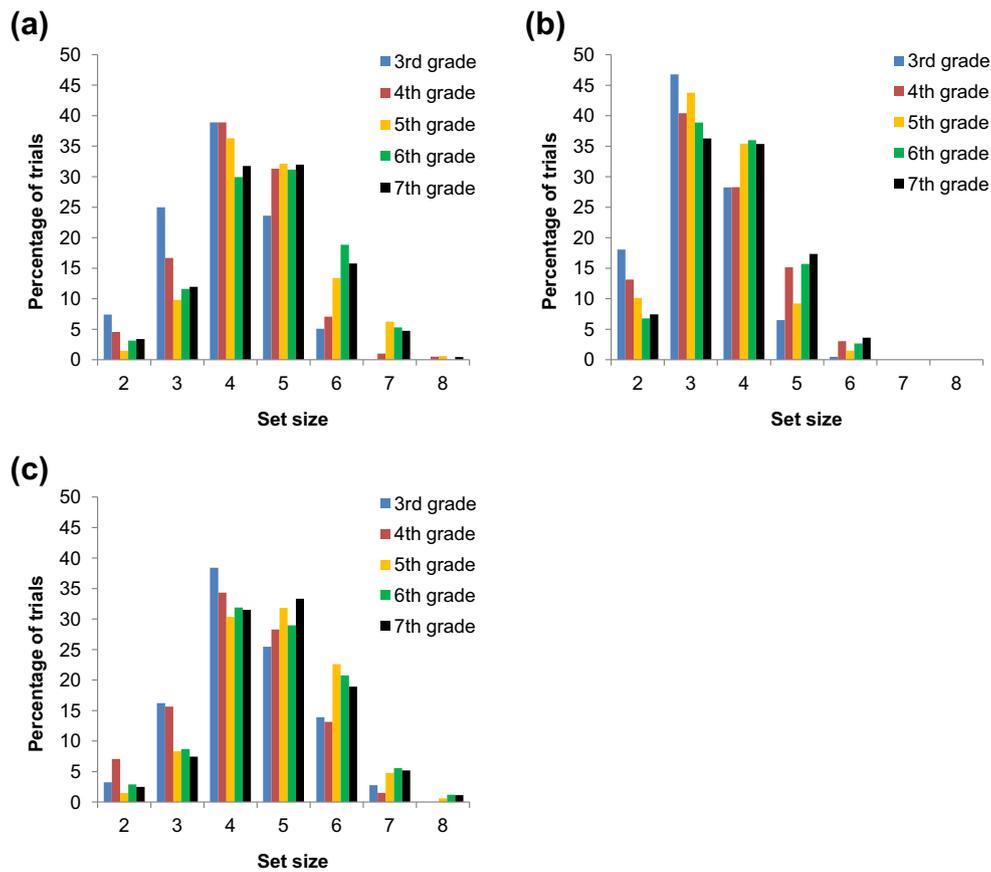


Fig. 3 Total percentage of trials performed at each set size as a function of grade for the reading span (a), the symmetry span (b), and the operation span (c)

to comparable levels of fatigue and proactive interference. This advantage is best illustrated by the fact that the lowest-performing participant in our sample did not obtain perfect accuracy even on set size 2 on any subtest, whereas the highest-performing participant reached set sizes 8, 6, and 8 on the reading span, symmetry span, and operation span, respectively. Obtaining the same information with a traditional ascending procedure (starting at set size 2 with three trials per level) would have required the first child to complete nine trials and memorize 18 items, and the second child to complete 57 trials and memorize 270 items. More practically, the constant

number of trials made the duration of the task relatively constant, allowing for easy group testing, even in groups of up to 25 children. Because the task was never discontinued, scores were estimated based on the same number of trials for all children, even in the lower ability range; this also limited the influence of temporary mind-wandering on final scores.

In our opinion, these combined advantages are significant enough to warrant further use of adaptive testing in the context of working memory assessment. In other words, the ACCES represents a first proof-of-concept for the use of the adaptive procedure in working memory measurement. Although the

Table 5 Descriptive statistics for subjective difficulty of the ACCES as a function of grade

| Measure | 3rd grade | | 4th grade | | 5th grade | | 6th grade | | 7th grade | |
|---------------------------|-----------|------|-----------|------|-----------|------|-----------|------|-----------|------|
| | Mean | SD |
| Reading span difficulty | 2.06 | 0.97 | 2.15 | 1.06 | 2.07 | 0.93 | 1.91 | 0.69 | 1.87 | 0.75 |
| Symmetry span difficulty | 2.20 | 0.83 | 2.85 | 1.12 | 2.73 | 1.04 | 2.62 | 0.86 | 2.51 | 0.80 |
| Operation span difficulty | 2.34 | 1.16 | 2.85 | 1.35 | 2.43 | 1.19 | 2.54 | 1.12 | 2.59 | 1.03 |
| Average difficulty | 2.20 | 0.61 | 2.62 | 0.97 | 2.41 | 0.78 | 2.36 | 0.67 | 2.32 | 0.60 |

Subjective difficulty was assessed on a scale from 1 to 5. Average difficulty was computed by averaging the difficulty ratings of each child over the three subtests

Table 6 Two-way correlations between scores on the ACCES and the SPM

| Measure | Reading span | Symmetry span | Operation span | Composite score | Raven's SPM |
|-----------------|--------------|---------------|----------------|-----------------|-------------|
| Reading span | - | .40 | .51 | .80 | .31 |
| Symmetry span | .40 | - | .46 | .78 | .29 |
| Operation span | .51 | .46 | - | .82 | .22 |
| Composite score | .80 | .78 | .82 | - | .34 |
| Raven's SPM | .31 | .29 | .22 | .34 | - |

$N=282$. All correlations are significant at the $p < .001$ level

task was developed and validated in French, it would be straightforward to extend to other languages and to other complex span tasks. In fact, the adaptive procedure can be generalized to paradigms other than complex spans: the pyramidal trial structure of the ACCES (Fig. 2) can be directly implemented in other memory tasks involving the memorization of a discrete number of items in each trial. As a consequence, a variety of memory tasks could be developed or modified to accommodate adaptive testing. The list includes the majority of forward, backward, transposed and complex span tests currently in use.

The present results suggest several opportunities for improvement in the development of adaptive working memory tasks. One example is the set of cutoff criteria used to modulate difficulty levels in the ACCES (see Table 1), which were defined relatively arbitrarily. The difficulty level was only increased when participants correctly recalled all stimuli (to preserve comparability with traditional ascending tasks), but it would be possible to define a less stringent criterion. This would increase the number of trials performed at high set sizes and presumably make the task more sensitive in the high ability range. The adaptive procedure itself was implemented in the form of a pyramid with constant set size, but other procedures exist and may help improve reliability of WMC estimates (although many of these procedures require different numbers of trials for different participants; see Weiss, 1974). Another possibility would be to explore the use of other psychometric analyses: the difficulty of using traditional factor analyses and internal consistency indices in the context of an adaptive task limits the possibilities for task validation.

At least two possible extensions for the ACCES task itself also come to mind. Firstly, the reading span, symmetry span, and operation span subtests were chosen to make the task comparable to the most common adult WMC tasks (Redick et al., 2012; Unsworth et al., 2005). In the context of child assessment, however, these complex span tasks are limited by the fact that they involve complex processing skills. Processing skills tend to develop with age, especially reading and mathematical abilities, making the processing problems less difficult for older children (see Table 4). Because WMC performance depends on the tradeoff between processing and storage demands (e.g., Barrouillet et al., 2009), the increase of

WMC scores with age is partly attributable to increases in processing abilities. Certain tasks are less vulnerable to this problem, such as the listening span variant of the reading span (Daneman & Carpenter, 1980) and complex spans involving elementary processing problems (Lépine, Barrouillet, & Camos, 2005), and these may be better suited to child assessment.

Secondly, it may be worthwhile to generalize the adaptive procedure implemented in the ACCES to similar complex span batteries used with adults (e.g., Redick et al., 2012; Oswald et al., 2015). Although our development of an adaptive task was motivated by the specificities of children sample, such as their wide range of ability, similar questions can emerge in adult assessment. For example, working memory tasks in use with older adults or with clinical samples demonstrating a WMC deficit may similarly benefit from adaptive testing. Pilot data collected in a large sample of psychology students during the development of the ACCES suggests that the adaptive procedure retains its sensitivity in adults and could be used to yield appropriate WMC estimates in this population, as it does with children.

Several arguments also suggest that adaptive testing could improve the validity of WMC tasks by yielding scores that better reflect core processes underlying working memory functioning. For example, participants completing an adaptive task perform more trials close to their actual working memory span; challenging trials are likely to place more demands on executive control (Adam et al., 2015; Vandierendonck et al., 1998b). Likewise, trials performed at a high set size in a complex span place more demands on secondary memory (e.g., Rose, Myerson, Roediger III, & Hale, 2010), an important component of working memory performance that partly mediates its relationship with high-level cognition (Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth & Engle, 2007; Unsworth, Fukuda, Awh, & Vogel, 2014). Thus, adaptive testing may ultimately help improve working memory measurement in all populations.

Acknowledgments This research was supported by the Regional Council of Picardie (RCP) and the European Regional Development Fund (ERDF). The authors heartily thank Randy Engle and Anik de Ribaupierre for kindly authorizing the use of material from their versions

of the symmetry span and reading span, respectively. The authors also thank the children who participated in the study and their families, as well as the elementary schools and middle schools that authorized the study.

References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. doi:10.1037/0033-2909.131.1.30
- Adam, K. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The contribution of attentional lapses to individual differences in visual working memory capacity. *Journal Of Cognitive Neuroscience*, *27*(8), 1601–1616. doi:10.1162/jocn_a_00811
- Alloway, T. P. (2007). *Automated Working Memory Assessment (AWMA)*. London: Harcourt Assessment.
- Alloway, T. P. (2011). A comparison of working memory profiles in children with ADHD and DCD. *Child Neuropsychology*, *17*(5), 483–494. doi:10.1080/09297049.2011.553590
- Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal Of Experimental Child Psychology*, *106*(1), 20–29. doi:10.1016/j.jecp.2009.11.003
- Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology*, *45*(2), 477–490. doi:10.1037/a0014615
- Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology*, *41*(4), 579–597. doi:10.1037/0012-1649.41.4.579
- Cain, K. (2006). Children's reading comprehension: The role of working memory in normal and impaired development. In S. J. Pickering (Ed.), *Working memory and Education* (pp. 62–93). Oxford: Academic Press.
- Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal Of Experimental Child Psychology*, *33*(3), 386–404. doi:10.1016/0022-0965(82)90054-6
- Conlin, J. A., Gathercole, S. E., & Adams, J. W. (2005). Children's working memory: Investigating performance limitations in complex span tasks. *Journal Of Experimental Child Psychology*, *90*(4), 303–317. doi:10.1016/j.jecp.2004.12.001
- 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*(5), 769–786. doi:10.3758/BF03196772
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. doi:10.1016/s0022-5371(80)90312-6
- Delaloye, C., Ludwig, C., Borella, E., Chicherio, C., & De Ribaupierre, A. (2008). L'empan de lecture comme épreuve mesurant la capacité de mémoire de travail: Normes basées sur une population francophone de 775 adultes jeunes et âgés. *Revue Européenne De Psychologie Appliquée*, *58*, 89–103. doi:10.1016/j.erap.2006.12.004
- Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–236. doi:10.3758/s13421-014-0461-7
- Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, *37*(4), 581–590. doi:10.3758/BF03192728
- Gathercole, S. E. (1998). The development of memory. *Journal Of Child Psychology And Psychiatry*, *39*(1), 3–27. doi:10.1017/S0021963097001753
- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends In Cognitive Sciences*, *3*(11), 410–419. doi:10.1016/s1364-6613(99)01388-1
- Gathercole, S. E., Brown, L., & Pickering, S. J. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Educational and Child Psychology*, *20*(3), 109–122.
- Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology*, *70*(2), 177–194. doi:10.1348/000709900158047
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. doi:10.1037/0012-1649.40.2.177
- Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16. doi:10.1002/acp.934
- Gonthier, C., Thomassin, N., & Roulin, J.-L. (2016). The Composite Complex Span: French validation of a short working memory task. *Behavior Research Methods*, *48*(1), 233–242. doi:10.3758/s13428-015-0566-3
- 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*(2), 184–198. doi:10.1037/0096-3445.130.2.184
- Hoard, M. K., Geary, D. C., Byrd-Craven, J., & Nugent, L. (2008). Mathematical cognition in intellectually precocious first graders. *Developmental Neuropsychology*, *33*(3), 251–276. doi:10.1080/87565640801982338
- Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–F15. doi:10.1111/j.1467-7687.2009.00848.x
- Holmes, J., Hilton, K. A., Place, M., Alloway, T. P., Elliott, J. G., & Gathercole, S. E. (2014). Children with low working memory and children with ADHD: Same or different?. *Frontiers In Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00976
- Hutton, U. Z., & Towse, J. N. (2001). Short-term memory and working memory as indices of children's cognitive skills. *Memory*, *9*(4–6), 383–394. doi:10.1080/09658210042000058
- Jeffries, S., & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia: An International Journal Of Research And Practice*, *10*(3), 196–214. doi:10.1002/dys.278
- Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *26*(2), 336–358. doi:10.1037/0278-7393.26.2.336
- Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal Of Clinical And Experimental Neuropsychology*, *24*(6), 781–791. doi:10.1076/jcen.24.6.781.8395
- Korrmann, J., Zettler, I., Kammerer, Y., Gerjets, P., & Trautwein, U. (2015). What characterizes children nominated as gifted by teachers? A closer consideration of working memory and intelligence. *High Ability Studies*, 1–18. doi:10.1080/13598139.2015.1033513
- Lépine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin & Review*, *12*(1), 165–170. doi:10.3758/BF03196363

- Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2013). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861–883. doi:10.3758/s13423-013-0565-2
- Loosli, S. V., Rahm, B., Unterrainer, J. M., Weiller, C., & Kaller, C. P. (2014). Developmental change in proactive interference across the life span: Evidence from two working memory tasks. *Developmental Psychology*, *50*(4), 1060–1072. doi:10.1037/a0035231
- Lord, F. M. (1968). *Some test theory for tailored testing*. ETS Research Bulletin Series, *1968*(2), i–62. doi:10.1002/j.2333-8504.1968.tb00562.x
- Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal Of Experimental Psychology: General*, *130*(2), 199–207. doi:10.1037/0096-344
- Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working memory and specific language impairment: An update on the relation and perspectives on assessment and treatment. *American Journal Of Speech-Language Pathology*, *19*(1), 78–94. doi:10.1044/1058-0360(2009/09-0028)
- Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal Of Experimental Psychology: General*, *141*(4), 788–798. doi:10.1037/a0027968
- Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, *47*(4), 1343–1355. doi:10.3758/s13428-014-0543-2
- Pearce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. doi:10.1016/j.jneumeth.2006.11.017
- Pickering, S. J., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children (WMTB-C)*. London: Pearson Assessment.
- Raven, J., Raven, J. C., & Court, J. H. 2000, updated (2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3: The Standard Progressive Matrices*. San Antonio, TX: Harcourt Assessment.
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164–171. doi:10.1027/1015-5759/a000123
- Reynolds, C. R., & Voress, J. K. (2007). *Test of memory and learning* (2nd ed.). Austin: PRO-ED.
- Rose, N. S., Myerson, J., Roediger, H. L., III, & Hale, S. (2010). Similarities and differences between working memory and long-term memory: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 471–483. doi:10.1037/a0018405
- Seigneuric, A., & Ehrlich, M. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading And Writing*, *18*(7–9), 617–656. doi:10.1007/s11145-005-2038-0
- Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116–141. doi:10.1016/j.jml.2014.01.004
- St Clair-Thompson, H., & Sykes, S. (2010). Scoring methods and the predictive ability of working memory tasks. *Behavior Research Methods*, *42*(4), 969–975. doi:10.3758/BRM.42.4.969
- Swanson, H. L. (2006). Cognitive processes that underlie mathematical precociousness in young children. *Journal Of Experimental Child Psychology*, *93*(3), 239–264. doi:10.1016/j.jecp.2005.09.006
- Swanson, H. L. (2011). Working memory, attention, and mathematical problem solving: A longitudinal study of elementary school children. *Journal Of Educational Psychology*, *103*(4), 821–837. doi:10.1037/a0025114
- Thaler, N. S., Goldstein, G., Pettegrew, J. W., Luther, J. F., Reynolds, C. R., & Allen, D. N. (2013). Developmental aspects of working and associative memory. *Archives Of Clinical Neuropsychology*, *28*(4), 348–355.
- Tillman, C. M. (2011). Developmental change in the relation between simple and complex spans: A meta-analysis. *Developmental Psychology*, *47*(4), 1012–1025. doi:10.1037/a0021794
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. doi:10.1016/0749-596x(89)90040-5
- Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066. doi:10.1037/0033-2909.133.6.10387
- Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. doi:10.1016/j.cogpsych.2014.01.003
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. doi:10.3758/bf03192720
- Vandierendonck, A., De Vooght, G., & Van der Goten, K. (1998a). Interfering with the central executive by means of a random interval repetition task. *The Quarterly Journal Of Experimental Psychology A: Human Experimental Psychology*, *51A*(1), 197–218. doi:10.1080/027249898391828
- Vandierendonck, A., De Vooght, G., & Van der Goten, K. (1998b). Does random time interval generation interfere with working memory executive functions? *European Journal Of Cognitive Psychology*, *10*(4), 413–442. doi:10.1080/713752284
- Vock, M., & Holling, H. (2008). The measurement of visuo-spatial and verbal-numerical working memory: Development of IRT-based scales. *Intelligence*, *36*(2), 161–182. doi:10.1016/j.intell.2007.02.004
- Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). San Antonio: NCS Pearson.
- Weiss, D. J. (1974). *Strategies of adaptive ability measurement*. Oxford, England: University of Minnesota, Department of Psychology.
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive theory. *Applied Psychological Measurement*, *6*(4), 473–492. doi:10.1177/014662168200600408
- Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., & ... Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal Of Clinical And Experimental Neuropsychology*, *33*(1), 101–111. doi:10.1080/13803395.2010.493149