

Scoring methods and the predictive ability of working memory tasks

HELEN ST CLAIR-THOMPSON AND SARAH SYKES
University of Hull, Hull, England

Measures of working memory (WM) are useful predictors of cognitive skills and educational attainment in children. A number of scoring methods can be used for WM tasks—for example, the sum of all correctly recalled stimuli in perfectly recalled lists (absolute score) or the proportion of items recalled in the correct serial position during the task (proportion correct). The present study explored whether proportion correct scoring had an advantage over absolute scoring of WM tasks for predicting children's educational attainment. The participants were 81 primary school children aged 7–8 years. Each participant was tested on five measures of WM. Schools supplied national curriculum attainment levels for each child in reading, writing, mathematics, and science. The results revealed that proportion correct scoring resulted in WM tasks' being better predictors of children's achievement. The results are discussed in terms of both psychological theory and implications for research methods.

Working memory (WM) is a limited-capacity system responsible for storing and integrating information during complex and demanding activities (Baddeley, 1986). Measures of WM are widely used across a number of domains of psychology (see, e.g., Conway et al., 2005; Friedman & Miyake, 2005; Miyake, 2001). For example, within educational psychology, WM can be used to predict children's scholastic attainment across a number of curriculum areas (for a review, see Alloway & Gathercole, 2006; Gathercole & Alloway, 2008; Pickering, 2006).

WM can be assessed using a range of tasks (for a review, see Oberauer, 2005). Among the most commonly used are WM span measures in which participants engage in online processing while maintaining information for later recall. For example, in listening span (Daneman & Carpenter, 1980), participants make judgments about the meaning of sentences and then attempt to recall the final word of each sentence in sequence. These tasks can be distinguished from short-term memory tasks, such as digit recall or block recall, that involve the immediate recall of information with no supplementary processing. Measures of WM are consistently better predictors of cognitive skills than are measures of short-term memory (see, e.g., Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Engle, Tuholski, Laughlin, & Conway, 1999).

Despite their popularity in research and assessment, no standard administration and scoring methods exist for WM tasks (see also Conway et al., 2005; Friedman & Miyake, 2005). Usually, tasks begin with short lists of items to remember, with the number of items increasing over successive trials. In many such cases, the task is terminated when a participant fails at a list length (i.e., if a participant fails

to recall the items correctly in the majority of trials at one level of difficulty). However, some researchers prefer to administer all the trials. A number of scoring methods can also be used. Some researchers assign a span score of the highest level at which a participant successfully recalls the stimuli in the majority of trials (e.g., Daneman & Carpenter, 1980). Other methods include counting the total number of stimuli in perfectly recalled trials (see, e.g., Engle et al., 1999; McNamara & Scott, 2001), or the proportion of stimuli recalled correctly throughout the task (see, e.g., Friedman & Miyake, 2000; Kane et al., 2004; Turner & Engle, 1989).

Scores obtained using different methods are highly correlated (see, e.g., Friedman & Miyake, 2005; Klein & Fiss, 1999; Turner & Engle, 1989; Waters & Caplan, 1996). However, some scoring methods demonstrate better psychometric properties (see also Conway et al., 2005; Friedman & Miyake, 2005), and some result in higher correlations with criterion measures (e.g., Unsworth & Engle, 2007). Friedman and Miyake (2005) compared scoring methods for the reading span task, in which participants read sets of sentences aloud and attempt to remember the last word of each. One score was the proportion correct, referring to the proportion of items that a participant recalled in the correct serial position during the task. For example, if a participant recalled three out of five items on a trial, he or she scored .60 for that trial. The proportions for all the trials were then averaged, with the maximum possible score being 1.00.¹ Another score was the absolute score, referring to the sum of all correctly recalled stimuli in perfectly recalled lists. For example, if a participant recalled two items in a list of two, three items in a list of three, and three items in a list

H. St Clair-Thompson, h.st-clair-thompson@hull.ac.uk



of four, then their score was 5 (2 + 3 + 0). The proportion correct score was more closely correlated with Verbal SAT scores than the absolute score. This difference was not due to reliability; internal consistency and test-retest reliability were similar for the two methods. Friedman and Miyake (2005) therefore suggested that the differences arose because proportion correct scoring might have been more sensitive to subtle individual differences.

A more detailed explanation, however, was offered by Unsworth and Engle (2007). They demonstrated that proportion correct scoring resulted in tasks' being better predictors of criterion measures for a range of short-term and WM tasks. They suggested that WM comprises an active maintenance or short-term memory component (*primary memory*) and controlled search and retrieval processes (*secondary memory*). Items are thought to be initially maintained in primary memory but then displaced to secondary memory by other incoming items or distracting information. WM tasks thus employ both primary and secondary memory because processing activities displace items from primary memory. Short-term memory tasks typically employ only primary memory but will draw upon secondary memory when long lists of items are presented and the early items in the list have been displaced. Differences between absolute and proportion correct scoring may result from the proportion correct method's containing additional information from long list lengths. Participants recall some items at long list lengths, even though they cannot remember them all. Proportion correct scoring allows these items to contribute to a participant's score and thus may capture more variance from secondary memory. According to this view, rather than the methods' differing in sensitivity, the scoring method used indicates the cognitive resources that are being assessed.

Given that WM tasks are among the most widely used measurement tools in psychology (see also Conway et al., 2005), further research is needed to examine differences between scoring methods. Due to the importance of WM tasks in educational research and assessment (e.g., Alloway & Gathercole, 2006; Gathercole & Alloway, 2008; Pickering, 2006), one outstanding issue is whether scoring methods influence the predictive ability of WM tasks in children. One might expect the advantage of proportion correct scoring over absolute scoring to be smaller in this participant group. It has been suggested that children's performance on WM tasks is more closely related to that on short-term memory tasks due to a lack of rehearsal (Cowan et al., 2005). If differences between scoring methods result from varying degrees of primary and secondary memory involvement, then the closer relationship between short-term memory and WM in children (Cowan et al., 2005) might be expected to reduce the advantage for proportion correct scoring.

Previous studies examining scoring methods (Friedman & Miyake, 2005; Unsworth & Engle, 2007) have also neglected to examine differences between scoring methods in terms of assigning individuals to discrete groups. High-WM and low-WM groups are often created for research using median splits. Statistical analyses such as ANOVAs can then be conducted to examine how individuals with

different WM capacities differ in their performance on criterion tasks. Friedman and Miyake (2005) discussed the proportion of participants who changed between high- and low-WM groups over two testing sessions for each scoring method. However, they did not examine differences between the scoring methods themselves. It is possible that dichotomization is dependent upon the scoring method used. If proportion correct scores are better predictors of cognitive skills than absolute scores, then one might also expect a greater difference in attainment between high- and low-WM groups using this method. In addition, previous studies have also failed to examine differences between scoring methods in short-term memory tasks in the visuospatial domain.

The present study, therefore, had three main aims: to examine the effects of scoring method on the predictive ability of WM tasks in children, to assess the influence of scoring method on the assignment of individuals to discrete groups, and to explore scoring methods across a range of tasks including those assessing visuospatial short-term memory. Primary school children aged 7–8 years were tested on five measures of WM. The absolute score and the proportion correct score were calculated for each task. Analyses were conducted to examine whether the proportion correct score was a better predictor of scholastic attainment than the absolute score in each case.

METHOD

Participants

The participants were 81 children with a mean age of 7 years and 5 months ($SD = 4$ months) who were recruited from a primary school in the north of England.

Materials and Procedure

Participants were tested individually in a quiet area of the school classroom. They completed five measures of memory, designed to tap the phonological loop, visuospatial sketchpad, and central executive components of the multiple-component model of WM (Baddeley, 2000). The measures were taken from the Working Memory Test Battery for Children (WMTB-C; Pickering & Gathercole, 2001). However, the administration method differed from that which is specified in the WMTB-C manual. Participants received four trials at all the specified list lengths for each WM task. No stopping rule was employed.

In the word-recall task, participants were asked to verbally recall, in the same order, sequences of monosyllabic words spoken aloud by the experimenter. The words were presented at the rate of one per second. Following two practice trials, there were four trials at each list length of one to six. Test-retest reliability for word recall was reported as .80 for children aged 6–8 years by Pickering and Gathercole (2001).

In the block-recall task, participants were asked to recall, in the same order, sequences tapped out on blocks by the experimenter. Following two practice trials, there were four trials at each list length of one to six. Test-retest reliability for block recall was reported as .63 for children aged 6–8 years by Pickering and Gathercole (2001).

In the listening-recall task, participants heard a series of sentences and were asked to judge the veracity of each. At the end of each trial, they were asked to verbally recall the final word from each sentence in sequence. Following two practice trials, there were four trials at each list length of one to four. A maximum list length of four was employed due to the average performance of 7- to 8-year-olds on this task equating to a memory span of one or two items. Test-retest

Table 1
Descriptive Statistics of Memory Tasks Using Each Scoring Method

Task	<i>M</i>	<i>SD</i>	Range	Skewness	Kurtosis
Word Recall					
Absolute score (max = 84)	30.49	8.96	15–60	.76	.52
Proportion correct	.40	.11	.20–.67	.45	–.39
Block Recall					
Absolute score (max = 84)	36.68	12.92	12–66	.04	–.75
Proportion correct	.43	.13	.11–.69	–.41	–.17
Listening Recall					
Absolute score (max = 40)	6.80	4.66	0–18	.27	–1.06
Proportion correct	.10	.06	0–.26	.37	–.56
Counting Recall					
Absolute score (max = 60)	19.54	10.75	4–50	1.02	.62
Proportion correct	.27	.12	.08–.54	.29	–.79
Backward Digit Recall					
Absolute score (max = 80)	18.27	8.73	6–41	.64	–.33
Proportion correct	.26	.12	.06–.58	.34	–.43

reliability for listening recall was reported as .83 for children of this age by Pickering and Gathercole (2001).

In the counting-recall task, participants were asked to count the number of items in a series of arrays and then verbally recall the successive tallies of each array. Again, there were two practice trials followed by four trials at each list length. Testing continued up to a list length of five, again determined by mean scores for children on this task. Test–retest reliability for counting recall was reported as .74 by Pickering and Gathercole (2001).

In the backward-digit-recall task, participants heard series of digits and were asked to verbally recall them in reverse order. There were four trials at each list length of two to six. Test–retest reliability for backward digit recall was reported as .53 by Pickering and Gathercole (2001).

For each memory task, two scores were calculated. Each depended on items’ being recalled in the correct serial order. That is, an item was counted as correct if it was recalled in exactly the same position in which it had been presented (with the exception of backward digit recall, for which the sequence had to be reversed). The absolute score was calculated as the sum of all correctly recalled stimuli in perfectly recalled lists. The proportion correct score was calculated as the average proportional recall for each trial, and the proportions for all the trials were then averaged (see also Friedman & Miyake, 2005). Finally, a single WM score was calculated for each scoring method by averaging the *Z* scores on the five WM tasks.

The schools supplied National Curriculum attainment levels in reading, writing, mathematics, and science for each pupil. These comprised teachers’ assessments of children’s progress as measured by tasks and tests that were administered informally. The majority of children aged 7–8 years are expected to work at Levels 1–3 of the National Curriculum, with each level having three sublevels.

RESULTS

The descriptive statistics for both scoring methods are shown in Table 1. The mean performance on each task was as expected for children of this age (see, e.g., Pickering & Gathercole, 2001), with lower scores on tasks of WM than on tasks of short-term memory. The correlations between WM and attainment in reading, writing, mathematics, and science are presented in Table 2. The upper triangle displays coefficients for the absolute scoring method, and the lower triangle displays coefficients for the proportion correct scoring method. The scores on the diagonal are the correlations between the two scoring methods. For each WM task, scores using the two methods were highly correlated, ranging from .75 for counting recall to .85 for word recall. There were also significant correlations between scores on each WM task. Scores using the two methods were also significantly related to achievement in reading, writing, mathematics, and science. The highest correlations were observed for tasks assessing the central executive, counting recall, listening recall, and backward digit recall, with coefficients ranging from .35 to .62 for absolute scoring and from .41 to .66 for proportion correct scoring.²

In order to examine whether proportion correct scoring predicted additional variance in scholastic attainment over and above absolute scoring, a series of regression analyses

Table 2
Correlations Between Memory Tasks and Scholastic Attainment

	1	2	3	4	5	6	7	8	9	10
1. Word recall	.85**	.18	.32**	.37**	.34**	.63**	.24*	.33**	.22*	.29**
2. Block recall	.20	.84**	.26*	.38**	.42**	.64**	.42**	.44**	.44**	.37**
3. Counting recall	.42**	.45**	.75**	.47**	.32*	.68**	.38**	.43**	.44**	.38**
4. Listening recall	.42**	.34**	.61**	.81**	.57**	.80**	.35**	.45**	.53**	.52**
5. Backward digit recall	.31**	.43**	.55**	.53**	.83**	.76**	.42**	.53**	.52**	.62**
6. Overall working memory	.64**	.66**	.82**	.79**	.77**	.91**	.52**	.62*	.61**	.62**
7. Reading	.30**	.50**	.51**	.46**	.41**	.61**	–	.87**	.81**	.81**
8. Writing	.39**	.53**	.59**	.54**	.51**	.70**	.87**	–	.81**	.81**
9. Mathematics	.31**	.53**	.66**	.58**	.55**	.72**	.81**	.81**	–	.83**
10. Science	.41**	.44**	.62**	.61**	.57**	.72**	.81**	.81**	.83**	–

Note—The upper triangle displays coefficients for absolute score, and the lower triangle displays coefficients for proportion correct. **p* < .05. ***p* < .01.

Table 3
Summary of Multiple Regression Analyses
for Reading, Writing, Mathematics, and Science

Task	Variable	ΔR^2			
		Reading	Writing	Mathematics	Science
Word recall	Absolute score	.06*	.11**	.05*	.09**
	Proportion correct	.03	.04*	.06*	.09**
Block recall	Absolute score	.17**	.20**	.19**	.14**
	Proportion correct	.07*	.09**	.08**	.06*
Listening recall	Absolute score	.12**	.21**	.28**	.27**
	Proportion correct	.09**	.09**	.06**	.11**
Counting recall	Absolute score	.14**	.18**	.20**	.15**
	Proportion correct	.18**	.17**	.25**	.25**
Backward digit recall	Absolute score	.18**	.28**	.27**	.38**
	Proportion correct	.02	.02	.05*	.01
Overall working memory	Absolute score	.27**	.39**	.38**	.39**
	Proportion correct	.11**	.10**	.14**	.13**

* $p < .05$. ** $p < .01$.

was conducted. For each curricular domain, the absolute score was entered first, followed by the proportion correct score. The results are presented in Table 3.

Word recall was a significant predictor of achievement in reading, writing, mathematics, and science, with the proportion correct scores predicting additional variance over the absolute scores for writing, mathematics, and science ($\Delta R^2 = .04, .06, \text{ and } .09$, respectively). Block recall was a significant predictor of each curricular domain, with the proportion correct score accounting for additional variance in each case ($\Delta R^2 = .07, .09, .08, \text{ and } .06$, respectively). Listening recall was also significantly related to each curricular domain, with the proportion correct score predicting additional variance ($\Delta R^2 = .09, .09, .06, \text{ and } .11$, respectively). A similar pattern emerged for counting recall ($\Delta R^2 = .18, .17, .25, \text{ and } .25$, respectively). Finally, backward digit recall was a significant predictor of achievement in reading, writing, mathematics, and science, but the proportion correct score only predicted additional variance in mathematics ($\Delta R^2 = .05$). All the analyses were conducted again, entering proportion correct and then absolute score, and absolute score did not predict any unique variance in attainment.

To examine the possible explanation that differences between the two scoring methods was due to reliability, *split-half reliability* was calculated for each task as the correlation between scores on the first two and second two trials at each list length. Cronbach's alpha was also calculated, with the number correct summed across levels for the first trial at each level, the second trial at each

level, and so on (see also Friedman & Miyake, 2005). The results are displayed in Table 4. The reliability estimates for the two scoring methods were similar for word recall, block recall, and counting recall, but were slightly higher for proportion scoring for listening recall and backward digit recall.

To further examine differences between the two scoring procedures, we examined the effects of scoring method on creating dichotomized groups by using median splits. Using each scoring method, we assigned all individuals with scores below the median to the low-WM group, and we assigned those with scores above the median to the high-WM group. The percentage of participants changing classification between the two methods was then calculated for each task.³ The percentage who changed classification between the low-WM and high-WM groups were as follows: 25.9% for word recall, 13.6% for block recall, 17.3% for listening recall, 21.0% for counting recall, and 16.0% for backward digit recall. The percentage who changed classification using the overall WM score was 9.9%.

To further explore the superiority of proportion correct scoring for predicting attainment, a series of ANOVAs was conducted on the attainment scores for the high-WM and low-WM groups defined by each scoring method. Using the overall WM scores with absolute scoring, there were significant differences between the high- and low-WM groups in each curriculum domain [$F(1,79) = 17.86, p < .01, \eta_p^2 = .18$, for reading; $F(1,79) = 23.62, p < .01, \eta_p^2 = .23$, for writing; $F(1,79) = 27.31, p < .01, \eta_p^2 = .26$, for

Table 4
Reliability Estimates for Each Working Memory Task

Task	Split-Half Reliability		Cronbach's Alpha	
	Absolute Score	Proportion Correct	Absolute Score	Proportion Correct
Word recall	.75	.76	.86	.86
Block recall	.73	.79	.78	.82
Listening recall	.62	.74	.76	.85
Counting recall	.72	.72	.84	.84
Backward digit recall	.62	.74	.76	.85

mathematics; and $F(1,79) = 34.82, p < .01, \eta_p^2 = .31$, for science]. Using proportion correct scoring, the analyses revealed larger effect sizes [$F(1,79) = 37.42, p < .01, \eta_p^2 = .32$; $F(1,79) = 45.13, p < .01, \eta_p^2 = .36$; $F(1,79) = 52.90, p < .01, \eta_p^2 = .40$; and $F(1,79) = 60.35, p < .01, \eta_p^2 = .43$, for reading, writing, mathematics, and science, respectively]. The mean attainment scores for high- and low-WM groups using each scoring method are displayed in Figure 1. A similar pattern of findings emerged for each WM task, with larger effect sizes using proportion scoring rather than absolute scoring. The differences between the two methods were, however, less pronounced for backward digit recall.

DISCUSSION

The main aim of the present study was to explore the effects of scoring method on the ability of WM tasks to predict children’s scholastic attainment. The results revealed that proportion correct scoring resulted in WM tasks’ being better predictors of children’s scholastic achievement than absolute scoring. The findings are therefore consistent with previous suggestions that some scoring methods can result in higher correlations between WM and criterion measures (e.g., Friedman & Miyake, 2005; Unsworth & Engle, 2007) and further extend these findings to the relationship between WM and scholastic attainment in children.

An additional aim of the present study was to assess the influence of scoring method on the assignment of individuals to discrete groups. Differences between the two scoring methods were observed when we used WM scores to assign individuals to low- or high-WM groups. First, a number of participants changed from the low- to the high-WM group, or vice versa, when changing from absolute to proportion correct scoring. Second, using proportion correct scoring resulted in a better dichotomization in terms of differences in attainment between high- and low-WM

groups. Although forming high- and low-WM groups with each scoring method revealed significant differences in children’s attainment, the differences (and, thus, effect sizes) were consistently larger for proportion correct scoring. This is a further demonstration of the superiority of proportion correct scoring for this purpose. Although classification of participants into groups is problematic (see, e.g., Friedman & Miyake, 2005; Waters & Caplan, 1996, 2003), such groups are often created in research to explore how individuals differing in WM differ in performance on a criterion task. The present study suggests that the creation of such groups is heavily dependent on the scoring method used for WM tasks and that, if the purpose of creating dichotomized groups is to examine differences in scholastic skills, then groups should be formed on the basis of proportion correct scoring.

It is important to note that the differences observed between the absolute and proportion correct scoring methods were not due to reliability. Split-half reliability and Cronbach’s alpha were similar for each scoring method for word recall, block recall, and counting recall. Proportion correct scoring resulted in slightly higher reliability for listening recall and backward digit recall. However, only small differences between the two scoring methods were observed for backward digit recall, and the greatest differences were found in counting recall, for which there were no differences in reliability. Thus, the differences between absolute and proportion correct scoring cannot be attributed to reliability alone.

There are, however, several explanations of the differences between scoring methods. Friedman and Miyake (2005) proposed that proportion scoring is a more sensitive measure of the same construct than that assessed by absolute scoring. Alternatively, Unsworth and Engle (2007) suggested that differences arise due to a part-whole relationship. That is, proportion correct scoring picks up exactly the same variance as absolute scoring, plus additional variance from long lists that are not per-

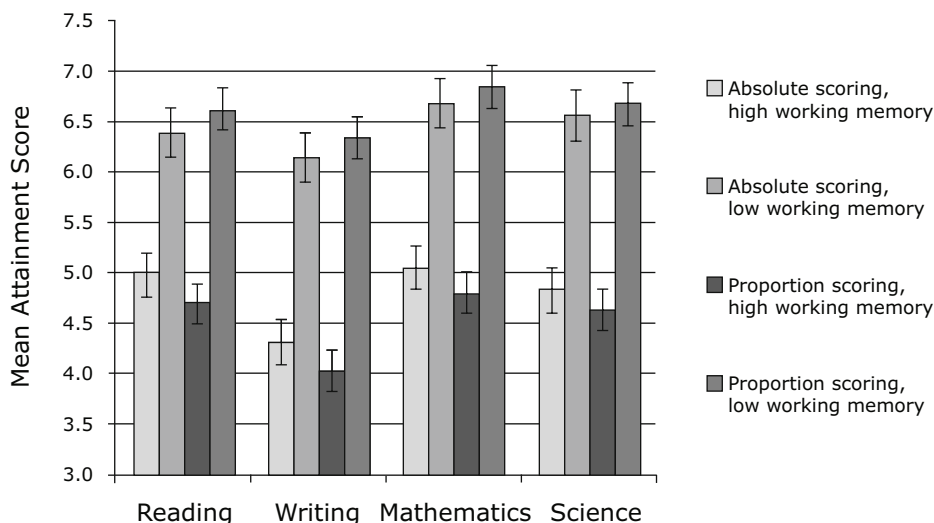


Figure 1. Mean attainment scores for the high and low working memory groups using each scoring method.

fectly recalled. Unsworth and Engle suggested that short lists require only primary memory, whereas longer lists require the use of both primary and secondary memory. Absolute scoring leads only to short lists contributing to the WM score and, thus, only to variance from primary memory. In contrast, proportion scoring leads to both short and long lists contributing to scores and, thus, to variability from both primary and secondary memory (or similarly, from both short-term memory and executive attentional resources). It is also worthy of note that findings presented by Lustig, May, and Hasher (2001) suggest that proportion scoring may predict additional variance in attainment because it captures variance from long lists that employ the ability to overcome proactive interference.

We suggest here that one might expect the difference between proportion scoring and absolute scoring to be smaller in children than in adults. Short-term memory and WM are closely related in children, possibly due to a lack of rehearsal (see, e.g., Cowan et al., 2005). The present study revealed close relationships between performance on short-term memory and WM tasks. However, this did not reduce the advantage for proportion correct scoring. This suggests that differences between absolute and proportion correct scoring do not arise due to longer lists preventing the use of rehearsal strategies. This also suggests that the superiority of proportion correct scoring may not result from varying degrees of primary and secondary memory involvement. Further research is therefore needed to elucidate the cognitive underpinnings of the differences between the two scoring methods.

The results of the present study have important implications for assessment and practice. A number of popular assessments of WM, such as the WMTB-C (Pickering & Gathercole, 2001) and the Automated Working Memory Assessment (Alloway, 2007), along with subtests of the British Ability Scales (Elliott, Smith, & McCulloch, 1997) and the Wechsler Intelligence Scale for Children (Wechsler, 1992), use scoring methods that are variations of absolute scoring. If proportion scoring were to be used, these assessments could be better at predicting complex cognition. Thus, to the extent that the purpose of using WM tasks is to predict cognitive performance, this purpose would appear to be better served using proportion correct scoring methods.

The findings of the present study also suggest that variations in administration and scoring methods may be responsible for some of the discrepancies that exist in the literature. Unsworth and Engle (2007) described some discrepancies in clinical research and suggested that these could be eliminated if researchers used identical scoring methods.

There are also discrepancies in research in educational and child psychology. For example, Swanson and Kim (2007) demonstrated that verbal short-term memory was uniquely related to children's mathematics performance. However, not all studies have reported evidence in favor of this relationship (e.g., Gathercole & Pickering, 2000; Geary, Hamson, & Hoard, 2000). Swanson and Kim employed a scoring procedure in which participants were given credit for correctly recalling some items in trials

that were not perfectly recalled, similar to proportion correct scoring. Other studies, however, have used absolute scoring. There are also some discrepancies in the literature concerned with relationships between WM and intelligence. Some researchers suggest that relationships arise due to storage capacity, whereas others propose that they arise due to executive attentional resources (e.g., Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Unsworth, Redick, Heitz, Broadway, & Engle, 2009). Some studies have employed absolute scoring (e.g., Conway et al., 2002), and others have employed proportion correct scoring (e.g., Colom et al., 2008). Consideration of the methodologies used in WM tasks may, therefore, eliminate these discrepancies.

It is, however, important to note that, in the present study, proportion correct scoring was not a significantly better predictor of attainment than was absolute scoring for all of the WM tasks. With the exception of predicting mathematics, the proportion correct did not predict additional variance in achievement for backward digit recall. This task differs from short-term memory tasks because of the requirement to reverse the sequence. However, it also differs from WM tasks because all of the items to be remembered are presented in succession, and there is no processing activity interleaved between the items (see also St Clair-Thompson, 2010). Therefore, it is possible that the differences between backward digit recall and other WM tasks are a result of individual differences in strategy use. For example, if participants delay processing until all digits are available, there may be fewer variations in cognitive resources between short and long lists. Further research is needed to explore this possibility.

A further consideration regarding the scoring of WM tasks is the ease of assessment for both participants and experimenters. Versions of absolute scoring, including truncated span or trials correct scores, can be computed without administering all of the trials within a task. Thus, testing can be terminated when a participant fails to recall correctly the stimuli in a given number of trials (see, e.g., Pickering & Gathercole, 2001). Consequently, assessments can be completed within a relatively short period of time, particularly with young children. Administering all of the trials is more time consuming. In addition, when one administers all of the trials, lower ability participants may become frustrated at higher levels. This could be minimized by randomizing the order of list lengths so that some success is experienced throughout the task. However, further research is needed to examine differences between increasing span and randomized order procedures (see also Conway et al., 2005; Lustig et al., 2001).

In conclusion, the present study suggests that using proportion correct methods of scoring for WM tasks results in tasks' being better predictors of children's scholastic attainment. Further research is needed to explore the cognitive underpinnings of the differences between the scoring methods. However, if the purpose of using WM tasks is to predict performance on criterion measures, then proportion correct scoring should be used. More generally, it is

suggested that it is important to consider administration and scoring methods in research into WM.

AUTHOR NOTE

This research was supported by Nuffield Foundation Science Bursary 36673. The authors thank the staff and pupils at Paisley Primary School in Hull, England, for their time and assistance with this research. Correspondence concerning this article should be addressed to H. St Clair-Thompson, Department of Psychology, University of Hull, Hull HU6 7RX, England (e-mail: h.st-clair-thompson@hull.ac.uk).

REFERENCES

- ALLOWAY, T. P. (2007). *Automated working memory assessment*. London: Pearson Assessment.
- ALLOWAY, T. P., & GATHERCOLE, S. E. (2006). How does working memory work in the classroom? *Educational Research & Reviews*, *1*, 134-139.
- BADDELEY, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- BADDELEY, A. [D.] (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417-423.
- COLOM, R., ABAD, F. J., QUIROGA, M. Á., SHIH, P. C., & FLORES-MENDOZA, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*, 584-606.
- CONWAY, A. R. A., COWAN, N., BUNTING, M. F., HERRIAULT, D. J., & MINKOFF, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163-184.
- CONWAY, A. R. A., KANE, M. J., BUNTING, M. F., HAMBRICK, D. Z., WILHELM, O., & ENGLE, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769-786.
- COWAN, N., ELLIOTT, E. M., SAULTS, J. S., MOREY, C. C., MATTOX, S., HISMJATULLINA, A., & CONWAY, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*, 42-100.
- DANEMAN, M., & CARPENTER, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*, 450-466.
- DANEMAN, M., & MERIKLE, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422-433.
- ELLIOTT, C. D., SMITH, P., & McCULLOCH, K. (1997). *British Ability Scales technical manual* (2nd ed.). Windsor, U.K.: NFER-Nelson.
- ENGLE, R. W., TUHOLSKI, S. W., LAUGHLIN, J. E., & CONWAY, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309-331.
- FRIEDMAN, N. P., & MIYAKE, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General*, *129*, 61-83.
- FRIEDMAN, N. P., & MIYAKE, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods, Instruments, & Computers*, *37*, 581-590.
- GATHERCOLE, S. E., & ALLOWAY, T. P. (2008). *Working memory and learning: A practical guide for teachers*. London: Sage.
- 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*, 177-194.
- GEARY, D. C., HAMSON, C. O., & HOARD, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology*, *77*, 236-263.
- KANE, M. J., HAMBRICK, D. Z., TUHOLSKI, S. W., WILHELM, O., PAYNE, T. W., & ENGLE, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189-217.
- KLEIN, K., & FISS, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments, & Computers*, *31*, 429-432.
- LUSTIG, C., MAY, C. P., & HASHER, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*, 199-207.
- MCNAMARA, D. S., & SCOTT, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, *29*, 10-17.
- MIYAKE, A. (2001). Individual differences in working memory: Introduction to the special issue. *Journal of Experimental Psychology: General*, *130*, 163-168.
- OBERAUER, K. (2005). The measurement of working memory capacity. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 393-407). Thousand Oaks, CA: Sage.
- PICKERING, S. J. (ED.) (2006). *Working memory and education*. Amsterdam: Elsevier.
- PICKERING, S. J., & GATHERCOLE, S. E. (2001). *Working Memory Test Battery for Children (WMTB-C)*. London: Psychological Corporation.
- ST CLAIR-THOMPSON, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology*, *22*, 286-297.
- SWANSON, H. L., & KIM, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence*, *35*, 151-168.
- TURNER, M. L., & ENGLE, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, *28*, 127-154.
- 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*, 1038-1066.
- UNSWORTH, N., REDICK, T. S., HEITZ, R. P., BROADWAY, J. M., & ENGLE, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635-654.
- WATERS, G. S., & CAPLAN, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, *49A*, 51-79.
- WATERS, G. S., & CAPLAN, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, *35*, 550-564.
- WECHSLER, D. (1992). *Wechsler Intelligence Scale for Children—Third Edition UK*. London: Psychological Corporation.

NOTES

1. Friedman and Miyake (2005) also calculated the total number of items recalled across all trials. The proportion correct was highly correlated with this score (.99), and the reliability estimates for the two methods were almost identical.
2. Rank correlations (Spearman's rho) were also consistently higher for proportion scoring than for absolute scoring. This suggests that the higher correlations for proportion scoring were not a result of the distribution of scores.
3. Note that this method differs from that of Friedman and Miyake (2005) and Waters and Caplan (1996), who examined changes between two testing sessions for each scoring method rather than group changes between the scoring methods themselves.

(Manuscript received December 20, 2009;
revision accepted for publication May 1, 2010.)