Psychology Science Quarterly, Volume 51, 2009 (4), pp. 403-419

Do Measures of working memory predict academic proficiency better than measures of intelligence?

KERRY LEE[1], MADELINE LEE PE[2], SU YIN ANG[2] & LAZAR STANKOV[2]

Abstract

It is often asserted that working memory predicts more variance in academic proficiency than do measures of intelligence. We used data from three studies to show that the validity of this assertion is highly dependent on the method of analysis. Using the same measures of intelligence, but different measures of working memory and algebraic proficiency, we found working memory provided better explanatory power only when analysis was conducted on the observed variable level. When the same data were analysed using structural equation models, only measures of intelligence had a direct effect on algebraic proficiency. From a theoretical viewpoint, our findings are consistent with a claim that working memory is a constituent component of (fluid) intelligence.

Key words: short term memory; intelligence measures; quantitative methods; statistical regression

[1] *Correspondence concerning this article should be addressed to:* Kerry Lee, PhD, National Institute of Education, 1 Nanyang Walk, Singapore 637616; email: kerry.lee@nie.edu.sg

[2] National Institute of Education, Nanyang Technological University, Singapore

In recent papers on the relationship between general intelligence, working memory, and academic proficiency, it has often been asserted that working memory predicts more variance in academic proficiency than do measures of intelligence (Andersson, 2008; Lee, Ng, Ng, & Lim, 2004; Swanson, 2004). Working memory is defined as a system that allows simultaneous, but temporary storage and processing of information in the service of cognitive tasks, such as reasoning and problem solving (Baddeley & Hitch, 1974; Baddeley, 2000). Working memory is deemed a good indicator of children's learning potential (Alloway, 2009a). At issue is whether working memory measures have better predictive values than do measures of intelligence. Here, we used data from three studies to examine whether the relationship between measures of intelligence, working memory, and academic proficiency are dependent on the method of analysis.

Applied issue: Intelligence, working memory as predictors of academic proficiency

Previous work on working memory and academic proficiency is largely supportive of the notion that measures of working memory are better predictors of school achievement than measures of intelligence. Andersson (2008), for example, found that measures of working memory predicted accuracy in children's performances on mathematical word problems (Grades 2 to 4) even after variation attributable to intelligence, reading ability and age differences were controlled. Similarly, Swanson and Beebe-Frankenberger (2004) found that working memory significantly predicted mathematical calculation and word problem solving accuracy in children from Grades 1 to 3 even after fluid intelligence measures were entered into the regression models. Swanson (2004) showed that working memory measures predicted accuracy on a mathematical problem solving task in 8- and 11-year-olds even after differences in age, mathematical knowledge, fluid intelligence, and reading abilities were statistically controlled. He found that working memory contributed about 5% of the unique variance in solution accuracy. In Bull and Scerif (2001), working memory span accounted for 3% more variance in mathematics performance in 7-year-olds than did measures of intelligence and reading ability.

In our own work (Lee et al., 2004; Lee, Ng, & Ng, 2009), we found that measures of central executive capacities predicted performances on algebraic problem solving tasks even after variation on a performance IQ task had been statistically controlled. The central executive measure uniquely predicted 2.6% of the variance in 10-year-olds' performance on mathematical word problem solving. Using path analysis, the central executive component contributed directly to mathematical performance and indirectly via literacy and performance IQ. Literacy and performance IQ also contributed directly to mathematical performance, but the standardized total effect of the central executive component was greater than that of both literacy and performance IQ.

Studies conducted with children with learning difficulties have arrived at similar conclusions. Alloway (2009a), for example, found that working memory provided unique and long term prediction of learning outcomes in reading and mathematics even though differences in intelligence, prior knowledge, and skills were statistically controlled. Of interest was that intelligence did not provide any more explanatory power than did working memory when variation in prior knowledge and skills were controlled. In a study conducted with children with low working memory, Alloway, Gathercole, Kirkwood and Elliot (2009) showed that measures of intelligence and working memory explained separate and unique variance in reading and mathematical learning proficiency.

Observed versus latent variable analyses

In the studies we have reviewed, researchers mostly relied on hierarchical regression analyses conducted with observed variables. In this study, we examined the relationships between measures of intelligence, working memory and academic proficiency using latent variables in structural equation models. One advantage in using latent versus observed variables is that measurement errors are modelled explicitly. It is widely acknowledged that measures of working memory and executive functioning are not pure (e.g., Rabbitt, 1997; Miyake et al., 2000). In regression analyses, measurement errors are confounded with true measures of the constructs in question. By using multiple indicators, extracting their common variance, and modelling measurement errors explicitly, latent level analyses provide a more precise examination of relationships between conceptual constructs.

Theoretical issue: Intelligence and working memory

The relationship between measures of intelligence and working memory has received extensive treatment, most recently in a paper by Ackerman, Beier, and Boyle (2005) who argued that the constructs of working memory and general intelligence were not isomorphic. Although our main focus in this paper is more applied in nature, our work is of relevance to this debate. Amongst those working in the area of human intelligence, Hunt (1978; 1980) was the first to call attention to Baddeley & Hitch's (1974) study. Both Hunt (1980) and Stankov (1983) saw working memory as an example of a limited capacity system, narrower in scope than the processing (or attentional) resources that provides an instantiation of Spearman's (1927) notion of mental energy. On a conceptual level, the construct of working memory proved useful in the explanation of performance on several fluid, primary ability intelligence tasks, such as measures of Cognition of Figural Relations as captured by the Raven's Progressive Matrices test (see Embretson, 1995) and Inductive Reasoning as captured by the Series Completion problems (see Stankov & Myors, 1990).

In parallel with attempts to employ the conceptual framework of working memory to understand fluid intelligence, there were efforts to develop new cognitive tasks based on Baddley and Hitch's (1974) ideas. One of the first Mental Counting (i.e., mental updating) tasks was described by Massaro (1975) and many more measures of working memory have been developed since (e.g., Suess, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). These tasks, including those employed in the present study, have become common measures of working memory and are frequently pitted against well-established measures of fluid or general intelligence in the prediction of academic performance.

Over the past two decades, two ways of treating these (new) measures of working memory have emerged. One way is in the spirit of Ackerman, Beier, and Boyle's (2005) approach. It considers working memory tests as measures of a new primary ability of fluid intelligence, akin to but distinct from, say, Induction (Series Completion) or Cognition of Figural Relations (Matrices), but clearly defining a fluid intelligence second-order factor (see

Stankov, 1988; Stankov & Cregan, 1993; Stankov, 2000). Under this approach, the construct of working memory is not deemed synonymous with fluid intelligence. Instead, it is deemed to be one of the many primary abilities that contribute to the measure of fluid intelligence. The second way is to treat measures of working memory as a gauge of a construct somewhat distinct from traditional measures of intelligence. This second approach may stem from the reductionistic belief that measures of intelligence need to be understood in terms of some "basic" processes like those studied in cognitive psychology (e.g., attention) or even in terms of neuropsychological processes like those classified under frontal lobe functions.

The present paper may shed some light on whether working memory measures tap processes that are different from those tapped by typical measures of intelligence. The logic is as follows. If working memory measures add to the prediction of academic proficiency in algebra, a proper conclusion should be that they tap something over and above measures of intelligence. This conclusion, like those reviewed in the preceding section will, however, be open to criticism that processes important to the measurement of intelligence were omitted in the selection of cognitive measures and therefore the issue will be hard to settle. The other outcome is that working memory does not add substantially to the prediction of algebraic proficiency over and above measures of intelligence. If this second outcome was to be obtained, the special role attributed to working memory in cognitive performance will be undermined. Note that this latter outcome does not question the validity of the working memory construct; it does question the superior predictive validity of putative measures of working memory.

Research question and hypothesis

Our aim was to examine the extent to which working memory measures predicted academic proficiency after variance attributable to intelligence has been controlled. We were specifically interested in whether these relationships are dependent on the method of analysis. Our criterion measure for academic proficiency was children's proficiency in algebraic word problems. Such problems are a standard component of the mathematical curriculum in Singapore. Children are introduced to arithmetic followed by algebraic word problems from the early primary years (i.e. 8 years of age). These problems are demanding and provide a gauge of children's proficiency in both reading comprehension and a specific branch of mathematics. We re-analysed data from a published study (Lee et al., 2004) and new data from two additional studies. Two sets of analyses were conducted. We expected the regression analyses to replicate previous findings. The structural equation modelling would provide information on whether findings from the regression analyses are biased by measurement errors.

Method

Participants

Participants in all three studies were from government schools located in middle to lower middle class areas in the western zone of Singapore. All children participated with parental consent. Power analyses for all three studies were conducted with parameters based on multi-predictor regression analyses with Type 1 error set at .05, small effect size, and power set at 85%. Because the present analysis utilised a subset of predictors found in the original studies, the power of each set of analyses was expected to exceed 85%. 151 Primary 5 children (77 boys, M_{age} of 10.7 years, SD = .65) participated in Study 1. 255 Primary 5 children (132 boys, M_{age} of 11.2 years, SD = .36) participated in Study 2. In Study 3, 151 Primary 5 children were recruited into the study (74 boys, M_{age} = 10.50 years, SD = .50). Due to absences from school, 33 children had partially missing data. To avoid a reduction in power, all missing values were replaced using the multiple imputation procedure in PASW Statistics (version 17.0.2). To minimise biases introduced by peculiarities associated with any one set of imputation, we used the imputation procedure to generate 10 complete data sets. Findings are based on results pooled across the imputations.

Instruments & procedures

In Study 1, children were administered a mathematical test, the Working Memory Test Battery for Children (Pickering & Gathercole, 2001), two subtests of intelligence, and a reading ability test. With the exception of the mathematics test, all other tests were administered on a one-to-one basis over several days. Further details about the procedure can be found in Lee et al. (2004).

Block Design and Vocabulary. The two measures of Intelligence were Block Design and Vocabulary from the Wechsler Intelligence Scale for Children (WISC, Wechsler, 1991). In the Block Design task, children were given a number of red and white blocks and were asked to use them to reproduce designs presented to them in a series of pictures. In the Vocabulary task, children were asked to give verbal explanations to a list of increasingly difficult words. Within the WISC, Block Design contributed to the Performance subscale and was known to be a measure of fluid intelligence. Vocabulary was part of a Verbal subscale and was known to be a measure of crystallized intelligence. For both Block Design (full range: 0 - 69) and Vocabulary (full range: 0 - 60), the dependent measures were scores based on the accuracy of response to each question. Reliability scores using split half-correlations corrected by the Spearman Brown formula were .87 for both measures (Wechsler, 1991).

Counting Recall & Backward Digit Recall. In Study 1, we used a subset of the central executive measures from the working memory battery: Counting Recall and Backward Digit Recall. In the Counting Recall task, children were presented with cards containing arrays of 4, 5, 6, or 7 dots. Depending on children's performance on the practice trials, the task began with trials containing 1 or 3 cards and progressed to trials containing 7 cards. When all the cards in a given trial had been counted, the children were asked to recall the total number of dots on each card in the order in which they were presented. The dependent measure was the number of trials correctly recalled (full range: 0 - 42). Test-retest reliability of this measure is .79 (Alloway, 2009b). In the Backward Digit Recall task, children were administered lists of numbers and were asked to recall the numbers in backward sequence after each list had been administered. Each trial contained 2 to 7 numbers. The dependent measure was the number of trials correctly recalled (full range: 0 - 42). Alloway (2009b) reported test-retest reliability of .69 for this measure.

Algebraic tasks. There were three parallel versions of the algebraic test. In each version, children were asked to solve 10 algebraic word problems in one of three ways: (a) using a schematic heuristic called the model method, (b) using any method but the model method, or (c) using any method. One version was administered each week over 3 consecutive weeks. The sequence in which the three versions were administered was counterbalanced across schools. The dependent measure was the accuracy score for each of the tests (full range: 0 – 10). The Kuder-Richardson-20 coefficient (KR-20) for each version was more than .80.

In Study 2, children were administered a battery of working memory, executive functioning, English reading comprehension, vocabulary, performance intelligence, arithmetic, and algebraic tasks (further details about the procedures can be found in Lee et al., 2009). Similar to the first study, we used Block Design and vocabulary from the WISC as indicators of intelligence. Counting Recall and Letter Memory served as indicators of working memory.

Letter Memory. In the Letter Memory task, children were administered lists of letters and were asked to recall the last four letters on each list. The number of letters presented in each trial (5, 7, 9, or 11) varied randomly across trials. This information was not disclosed to children to increase the likelihood that they continued updating till the end of each trial. The task was divided into different sets. In the first set, children were asked to recall the last two letters presented in the lists and were told the number of letters that would be shown in the first three trials: 2, 4 or 5. In the remaining trials, the number of letters to be shown was not revealed. In the second set, children were asked to recall the last three letters in the list. Similar to the first set, children were told only the number of letters to be shown in the first three trials: 3, 5, or 6. The dependent measure was the proportion of letters recalled correctly within each trial, summed across trials (Cronbach's $\alpha = .66$).

Algebraic tasks. Three tasks were used to gauge children's algebraic proficiency: a) Representation Formation, b) Solution Formation, and c) Overall Accuracy. The Representation Formation task is, in essence, the initial steps in solving a word problem. Children were asked to use the model method to depict information in five algebraic word problems. The resultant schematics were coded as right or wrong (range: 0 to 5, KR-20 = .76). In the Solution Formation task, children were asked to construct step-by-step procedures to solve a number of algebraic problems. Schematics for these problems were provided to the children. One mark was awarded for each correct solution (range: 0 to 5, KR-20 = .84). The Overall Accuracy test was modified from Lee et al. (2004) and contained ten algebraic word problems drawn from the Primary 4 to Primary 6 curriculum. Children were asked to use the model method to solve these questions. Responses were coded as either right or wrong (range: 0 to 10, KR-20 = .87).

In Study 3, children were administered a large battery of executive functioning, reading comprehension, intelligence, motivation, and mathematics tasks. The tasks were divided into 5 sets and were administered over several sessions. We extracted from this study the same intelligence measures as those used in Study 1 and 2. Two working memory measures were used: Mr. X and Pictorial Updating.

Mr. X. In the Mr. X task (Alloway, 2007b), the child was shown two X shaped figures, each holding a ball at one of eight cardinal positions. The child had to decide whether they were holding the ball in the same hand. At the end of each trial, the child had to point to the position at which each ball was held, in the correct order. The task progressed from a block containing one set of Mr. X figures to a block containing seven sets of figures. Each block contained six trials. The total number of positions recalled served as the dependent measure (range = 0 – 42). The test-retest reliability for this measure is .77 (Alloway, 2007a).

Pictorial Updating. In the Pictorial Updating task, children were shown a series of animal pictures, one at a time. To ensure that updating was being used in the task, the children did not know how many items were going to be presented, and were asked to recall a specified number of animals from the end of each trial. The number of animals presented was varied randomly across trials (Min = 3, Max = 11). The task began with the child recalling the last two animals. This increased to the last four. Each block contained two practice sets and twelve experimental trials. The children received a point for every animal recalled correctly. The order of recall was not taken into account (range = 0 to 108). Test-retest reliability for the measure is .69.

Algebraic task. Items for the instrument were modified from Lee et al. (2004; 2009) and contained 10 algebraic word problems. Responses were coded as either right or wrong (range = 0 - 10, KR-20 = .86). For the structural equation model analysis, we wanted to avoid ambiguity associated with the use of a latent variable generated from a single indicator. For this reason, we divided the 10 questions into three groups. Each group contained questions of varying difficulties. Group accuracy scores served as indicators for the algebraic latent factor.

Results

Descriptive statistics of Studies 1, 2 and 3 are presented in Table 1. We may note that the arithmetic means on two intelligence subtests (Block Design and Vocabulary) are comparable across the three studies. Correlations between Block Design and Vocabulary range from .42 in Study 1 to .38 in Study 2 and .29 in Study 3. These correlations are within the typical range (i.e., between .20 and about .60) for correlations among measures of fluid and crystallized intelligence. It is also useful to comment on the correlations these two tests have with the measures of working memory and algebraic proficiency across the three studies. Given that in the extant literature, working memory is treated as being more closely associated with fluid intelligence, it is important to check whether Block Design is more highly correlated with measures of working memory than does Vocabulary. Inspection of Table 1, shows that out of six measures of working memory two – Backward Recall in Study 1 and Counting Recall in Study 2 – have higher correlations with Vocabulary then they do with Block Design. This is reasonable given that at the age of 10 to 11 years the differentiation of abilities into fluid and crystallized intelligence has not been completed. It is also important to note that virtually all algebraic tasks have higher correlation with Vocabulary than they do with Block Design. This is likely a reflection of the use of algebraic word problems as our criterion measures.

We used the same analytic procedure for data from all three studies. First, we conducted a hierarchical multiple regression analysis, this was followed by a latent factor or structural equation analysis.

Table 1:
Descriptive statistics and correlations for Studies 1, 2 and 3

		M	*SD*	1	2	3	4	5	6
	Study 1 ($N = 151$)								
1	Block design	44.45	10.04						
2	Vocabulary	24.48	8.40	.42**					
3	Counting recall	25.99	4.89	.35**	.31**				
4	Backward digit recall	18.21	5.35	.16	.34**	.49**			
5	Algebraic word problems (Model method)	2.50	2.38	.52**	.58**	.45**	.36**		
6	Algebraic word problems (Any method but the model method)	2.96	2.60	.51**	.52**	.43**	.33**	.82**	
7	Algebraic word problems (Any method)	2.97	2.63	.50**	.60**	.42**	.28**	.87**	.82**
	Study 2 ($N = 255$)								
1	Block Design	43.86	11.72						
2	Vocabulary	23.76	7.81	.38**					
3	Counting recall	23.51	4.20	.20**	.28**				
4	Letter Memory	6.36	2.15	.26**	.21**	.26**			
5	Algebraic task (Representation Formation)	2.55	1.63	.51**	.56**	.33**	.38**		
6	Algebraic task (Solution Formation)	3.26	1.81	.51**	.50**	.31**	.40**	.76**	
7	Algebraic task (Overall Accuracy)	4.22	3.07	.52**	.56**	.33**	.41**	.81**	.82**
	Study 3 ($N = 151$)								
1	Block Design	47.17	11.11						
2	Vocabulary	25.34	7.89	.29**					
3	Mr. X	17.84	5.83	.31**	.16				
4	Pictorial Updating	87.28	10.36	.43**	.33**	.27**			
5	Algebraic word problems (Group 1)	1.42	1.25	.39**	.51**	.26**	.39**		
6	Algebraic word problems (Group 2)	1.09	1.04	.37**	.55**	.21*	.41**	.76**	
7	Algebraic word problems (Group 3)	1.10	1.00	.28**	.36**	.24**	.36**	.63**	.59**

** $p < .01$

Regression analyses

In the regression analysis, the intelligence measures were entered together in the first block, followed by the working memory measures in a second block. Because the aim of this study was to compare findings from analysis conducted at the observed level versus those conducted at the latent level, we used principal component analysis to compute component scores for the three algebraic performance scores in Study 1 and 2. In Study 3, we used the overall accuracy score from the algebraic test as the criterion measure.

Findings from the regression analyses for all three studies are summarised in Table 2. In Study 1, when the measures of intelligence were entered alone, they predicted 46% (i.e., $R^2 = .46$) of variance in algebraic performance, $F(2, 148) = 62.62$, $p < .01$. Addition of the working memory measures significantly increased the amount of variance explained, $\Delta R^2 = .05$,

Table 2:
Hierarchical regression analysis results for Studies 1, 2 and 3

Variables	*B*	*SE*	*β*	R^2	ΔR^2
	Study 1				
Step 1 (Intelligence)				0.46**	
Block Design	0.03	0.01	0.35**		
Vocabulary	0.05	0.01	0.46**		
Step 2 (Working Memory)					.05**
Block Design	0.03	0.01	0.29**		
Vocabulary	0.05	0.01	0.39**		
Count Recall	0.04	0.01	0.20**		
Backward Digit Recall	0.01	0.01	0.07		
	Study 2				
Step 1 (Intelligence)				0.47**	
Block Design	0.03	0.01	0.39**		
Vocabulary	0.05	0.01	0.44**		
Step 2 (Working Memory)					0.07**
Block Design	0.03	> 0.01	0.33**		
Vocabulary	0.05	0.01	0.38**		
Count Recall	0.03	0.01	0.12**		
Letter Memory	0.10	0.02	0.23**		
	Study 3				
Step 1 (Intelligence)				0.36**	
Block Design	.07	.02	0.26**		
Vocabulary	.17	.03	0.47**		
Step 2 (Working Memory)					0.05**
Block Design	.04	.02	0.16*		
Vocabulary	.15	.03	0.41**		
Mr. X	.05	.04	0.10		
Pictorial Updating	.06	.02	0.21**		

** $p < .01$

F(2, 146) = 7.16, p < .01. Similar results were found in Study 2. Measures of intelligence alone significantly predicted algebraic performance, F(2, 252) = 111.67, p < .01, R^2 = .47. Adding the working memory measures significantly improved the prediction of algebraic performance, ΔR^2 = .07, F(2, 250) = 19.27, p < .01.

Study 3 validated the results found in Study 1 and 2. When the measures of intelligence were entered alone, they significantly predicted algebraic performance, F(2, 148) = 41.01, p < .01, R^2 = .36. Adding the working memory measures significantly improved the explanatory power of the model, ΔR^2 = .05, F(2, 146) = 5.81, p < .01.

Structural equation modelling

We submitted the same data to a series of structural equation models in which the relationship between measures of working memory, intelligence, and algebraic proficiency were investigated. As a baseline model, Model 1 assumed independence between the working memory and intelligence constructs. In this and subsequent models, we modelled direct paths from both latent constructs to algebraic proficiency (see Figure 1, Model 1). Model 2 is analogous to the hierarchical regression. First, we estimated the relationship between intelligence and algebraic proficiency by drawing a direct path from intelligence to algebraic proficiency. This is analogous to the first block in the hierarchical regression mentioned earlier. Second, we introduced the working memory measures and their associated latent factor. We added paths from working memory to intelligence and from working memory to algebraic proficiency. The path between working memory and algebraic proficiency in this model provided estimates analogous to those produced by adding working memory measures to the second block of the hierarchical regression. With only three latent variables, the directional relationship between working memory and intelligence could not be statistically disambiguated. Indeed, specifying either directional or non-directional paths between working memory and intelligence results in no changes to the fit and regression estimates for the models. Model 3 postulated a direct path from the working memory latent variable to intelligence and the path from working memory to algebraic proficiency was fixed at zero.

Findings and path coefficients from the structural equation analyses for Studies 1, 2, and 3 are shown in Figure 1 and Table 3. Goodness-of-fit statistics for Models 1, 2 and 3 are presented in Table 4. In Study 1, the baseline model (Model 1) did not provide a good fit to the data, $\chi^2(12)$ = 44.46, p < .01, CFI = .95, RMSEA = .13, and SRMR = .15.

In Model 2, we first fitted a direct path from intelligence to algebraic performance. This was followed by paths from working memory to intelligence, and from working memory to algebraic performance. The model provided a good fit to the data, $\chi^2(11)$ = 16.44, p = .125, CFI = .99, RMSEA = .06, and SRMR = .03. Notably, the path from working memory to algebraic performance (β = .02) was not significant. The estimate for this path is analogous to those obtained in the second block of our hierarchical regression analysis in which the explanatory power of working memory was controlled for variation in intelligence. In Model 2, the path from intelligence to algebraic performance (β = .90) and from working memory to intelligence (β = .64) were both significant. In Model 3, the path from working memory to algebraic performance was constrained to zero. Apart from being more parsimonious, the additional constraint did not significantly alter the model fit, $\Delta\chi^2$ = .01, df = 1, p = .920, CFI = .99, RMSEA = .05, and SRMR = .03. Model 3 accounted for 82.6% of variance in algebraic performance.

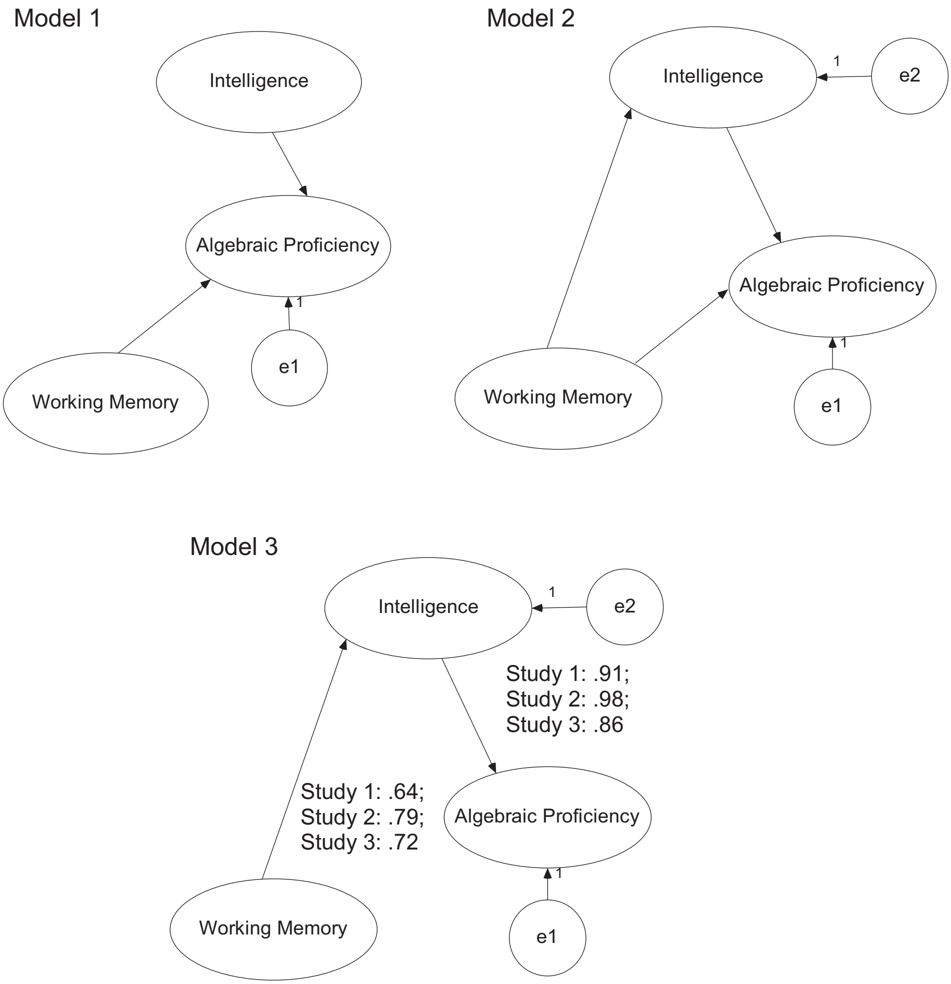

Figure 1:

Model 1 assumes independence of intelligence and working memory; only the path from working memory to intelligence was constrained to zero. Model 2 represents the full model; with working memory predicting intelligence, and both working memory and intelligence predicting algebraic proficiency. Model 3 is the best-fitting model; the direct path from working memory to algebraic proficiency was constrained to zero. Values refer to significant standardised path coefficients from Model 3 for the three studies (Study 1, Study 2 and Study 3)

Table 3:
Unstandardized and standardized regression estimates of the structural models for Study 1, 2 and 3. Estimates are based on specifications in Model 3

Study 1					
			b	*SE*	*β*
Working Memory	→	Intelligence	1.19	0.27	0.64**
Intelligence	→	Algebra	0.33	0.05	0.91**
Intelligence	→	Block Design	1.00		0.61**
Intelligence	→	Vocabulary	0.95	0.15	0.69**
Working Memory	→	Backward Digit Recall	1.00		0.61**
Working Memory	→	Counting Recall	1.20	0.25	0.81**
Algebraic Proficiency	→	Model Method	1.00		0.94**
Algebraic Proficiency	→	Any Method but Model	1.03	0.06	0.88**
Algebraic Proficiency	→	Any Method	1.10	0.05	0.93**
Study 2					
Working Memory	→	Intelligence	2.77	0.72	0.79**
Intelligence	→	Algebra	0.41	0.05	0.98**
Intelligence	→	Block Design	1.00		0.59**
Intelligence	→	Vocabulary	0.70	0.09	0.62**
Working Memory	→	Count Recall	1.00		0.47**
Working Memory	→	Letter Memory	0.61	0.12	0.56**
Algebraic Proficiency	→	Representation Formation	0.50	0.02	0.87**
Algebraic Proficiency	→	Solution Formation	0.56	0.03	0.88**
Algebraic Proficiency	→	Overall Accuracy	1.00		0.93**
Study 3[a]					
Working Memory	→	Intelligence	0.73/0.62	0.25/0.20	0.85**/0.72**
Intelligence	→	Algebra	0.16/0.14	0.03/0.02	0.85**/0.86**
Intelligence	→	Block Design	1.00/1.00		0.53**/0.59**
Intelligence	→	Vocabulary	0.85/0.76	0.17/0.12	0.63**/0.65**
Working Memory	→	Pictorial Updating	1.00/1.00		0.67**/0.76**
Working Memory	→	Mr. X	0.33/0.29	0.10/0.09	0.39**/0.38**
Algebraic Proficiency	→	Algebra Group 1	1.00/1.00		0.90**/0.89**
Algebraic Proficiency	→	Algebra Group 2	0.82/0.82	0.06/0.07	0.87**/0.86**
Algebraic Proficiency	→	Algebra Group 3	0.64/0.63	0.06/0.07	0.71**/0.70**

[a] Estimates after the slash are obtained from a model in which the disturbance estimate for intelligence was constrained.
** $p < .01$

Table 4:
Fit indices for structural models of Study 1, 2 and 3

	χ^2	*df*	CFI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	*p*
				Study 1				
Model 1	44.46**	12	0.95	0.13	0.15			
Model 2	16.44	11	0.99	0.06	0.03			
Model 3	16.45	12	0.99	0.05	0.03			
Model 2 vs. Model 3						0.01	1	0.92
				Study 2				
Model 1	43.01**	12	0.96	0.10	0.12			
Model 2	8.86	11	1.00	<0.01	0.02			
Model 3	9.28	12	1.00	<0.01	0.02			
Model 2 vs. Model 3						0.42	1	0.52
				Study 3				
Model 1	56.12**	12	0.88	0.16	0.15			
Model 2	22.56*	12	0.97	0.08	0.06			
Model $3_{constrained}$	23.05*	13	0.97	0.07	0.05			
Model $3_{unconstrained}$	20.87	12	0.97	0.07	0.05			
Model 2 vs. Model $3_{constrained}$						.49	1	0.48

*p < .05; ** p < .01

Similar results were found in Study 2. Model 1 did not provide a good fit to the data, $\chi^2(12) = 43.01$, $p < .001$, CFI = .96, RMSEA = .10, and SRMR = .12. Model 2 exhibited a good model fit, $\chi^2(11) = 8.86$, p = .635, CFI =1.00, RMSEA < .01, and SRMR = .02. With the exception of the path from working memory to algebraic proficiency (β = .12), the other two paths were significant (.84 ≥ β ≥ .74). In Model 3, the path from working memory to algebraic performance was constrained to zero. This also provided a good fit to the data, $\chi^2(12) = 9.28$, $p = .679$, CFI = 1.00, RMSEA < .01, and SRMR = .02. Similar to findings from Study 1, the additional constraint did not significantly improve model fit (χ^2 = .42, df = 1, p = .517). Model 3 accounted for 96.7% of variance in algebraic performance.

Study 3 showed the same pattern of findings as Studies 1 and 2. Model 1 resulted in a poor fit, $\chi^2(12) = 56.12$, $p < .001$, CFI = .88, RMSEA = .16, and SRMR = .15. Model 2 produced a solution that converged, but which contained negative variance in the disturbance estimate for intelligence. In the absence of outliers or signs of multicollinearity, we resolved this issue by constraining the estimate to a value obtained from the same estimates in Study 1 and 2 (Chen, Bollen, Paxton, Curran, & Kirby, 2001). The resultant model provided a good fit to the data, $\chi^2(12) = 22.56$, $p = .032$, CFI = .97, RMSEA = .08, and SRMR = .06. With the exception of the path from working memory to algebraic proficiency (β = -.14), the other two paths were significant (.99 ≥ β ≥ .72).

In Model 3, we constrained the path between working memory and algebraic performance to zero. To examine whether constraining the disturbance estimate for intelligence biased the model estimates, we ran the models with both the disturbance estimate left unconstrained and constrained as per Model 2. There were some differences in the estimates for both the measurement and structural components of the model, but the overall pattern remains the same. The constrained model provided a similar fit to the data as Model 2, $\chi^2(13) = 23.05$, $p = .031$, CFI = .97, RMSEA = .07, SRMR = .05, and explained 74.4% of the variance in algebraic performance.

In summary, our analyses of the data from all three studies show good support for Model 3. Only intelligence has a significant direct path to algebraic proficiency. At best, working memory has only an indirect effect on algebraic proficiency.

Discussion

The findings showed a clear difference between analyses conducted at the indicator versus the latent factor level. At the indicator level, the data showed that after controlling for individual differences in measures of intelligence, working memory explained an additional 5% to 7% of variance in algebraic proficiency. We found this pattern of results in all three studies despite some differences in the working memory and algebraic measures. The most striking finding was that working memory had no direct effect on algebraic proficiency in the structural equation models. Although the analyses were based on the same set of data, the finding was consistent: working memory had only an indirect effect on algebraic proficiency. Only intelligence, derived from two subtests measuring Verbal and Performance components of the WISC, had a direct effect on algebraic proficiency.

Which set of analyses do we trust and what impact has the differences in findings on theory or practice? First, we want to make clear that we are not suggesting that indicator based regression analyses are incorrect. For applied works in which researchers are interested in identifying measures that will aid in decision making, it may be desirable to stay on the indicator level. In the field, one may not have the luxury to collect the multiple measures necessary for computing latent factors. Standardisation and analytic data necessary for converting raw scores into latent scores are also not readily available.

The latent factor analyses differ from the regression analyses in that the former map more directly onto the constructs under measure. In latent factor analyses, explicit estimates of measurement errors for each latent construct are built into the models. For this reason, structural relationships amongst latent factors are deemed to be free of measurement errors and reflect only relationships amongst the theoretical constructs. In analyses involving indicator variables, estimates of relationships between variables are confounded by measurement error. Findings of significance could be caused by either overlapping relationships amongst the theoretical constructs, measurement errors, or both.

Although an over-simplification, this distinction is useful. If we are interested in the theoretical relationship between working memory, intelligence, and academic performance, latent factor analyses provide better guidance. From this perspective, an interpretation that is consistent with those commonly found in the literature is that working memory contributes to variation in intelligence, which in turn contributes to variation in academic performance. This interpretation is consistent with earlier investigations that suggested that the effect of working memory on academic performance is both direct and mediated by intelligence (Lee et al., 2004). The main difference with some previous findings is that our present findings suggest that there is no direct effect.

As mentioned in the introduction, the relationship between intelligence and working memory has been a source of active debate. Our findings do not question the theoretical status of working memory and its role in the established measures of intelligence. It is plausible to argue that working memory is involved in solving Block Design problems and must have had a "historical" role in the acquisition of language (e.g., Jensen, 1980). In fact, our findings do not challenge the usefulness of the measures of working memory for the prediction of academic performance. What they do challenge is a claim that measures of working memory are somehow different from other measures of intelligence and perhaps superior because of their predictive power. This is even more surprising given that only two subtests from WISC are used in the studies reported in the present paper.

Caveats

One limitation of this study is that we used only Vocabulary and Block Design as indicators of intelligence. Although these subtests are highly correlated with Full Scale IQ from the WISC and have high reliability (Sattler, 2001), it can argued that they do not provide a comprehensive measure of intelligence. However, given that our latent variable analyses show that working memory failed to explain directly variance in algebraic proficiency even when only the two subtests were used; it is likely the case that prediction based on the Full Scale IQ from WISC or some other scale assessing general intelligence would be even harder to improve upon.

An additional concern is that all the data used in this paper came from our laboratory. Although different measures of working memory were used in the various studies, the variety was limited. Given the modest correlations amongst working memory tasks ($r = .26$ to $.49$ in our data), it is important to replicate our findings using a wider variety of working memory tasks.

Conclusions

The findings show a clear differentiation in the relationship between intelligence, working memory, and algebraic proficiency depending on whether data were analysed on the observed or latent level. When analysis was conducted on the observed variable level, we found that working memory provided more explanatory power than intelligence. The reverse was true when the same data were used to generate latent measures and analysed using structural equation models. This divergence in findings reopens the debate on the relationship between working memory and intelligence. We argue that the findings are consistent with a view that working memory is one of the constituent measures of intelligence.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin, 131,* 30-60.

Alloway, T. P. (2007a). Working memory, reading and mathematical skills in children with developmental coordination disorder. *Journal of Experimental Child Psychology, 97,* 20-36.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development, 80,* 606-621.

Alloway, T. P. (2007b). The Automated Working Memory Assessment [Computer software]. London: Psychological Corporation.

Alloway, T. P. (2009a). Cognitive Training: Improvements in Academic Attainments. *Professional Association for Teachers of Students with Specific Learning Difficulties, 22,* 57-61.

Alloway, T. P. (2009b). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment, 25*.

Andersson, U. (2008). Working memory as a predictor of written arithmetical skills in children: The importance of central executive functions. *British Journal of Educational Psychology, 78,* 181-203.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423.

Baddeley, A., & Hitch, G. J. (1974). Working Memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation* (pp. 47-90). New York: Academic Press.

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

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models. *Sociological Methods & Research, 29,* 468-508.

Embretson, S. E. (1995). The role of working memory capacity and general control processes in intelligence. *Intelligence, 20,* 169-189.

Hunt, E. (1980). Intelligence as an information-processing concept. *British Journal of Psychology, 71*.

Hunt, E. & Washington University, S. D. o. P. (1978). The Foundations of Verbal Comprehension. http://search.ebscohost.com [On-line].

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Lee, K., Ng, E., & Ng, S. F. (2009). The contributions of working memory and executive functioning to problem representation and solution generation in algebraic word problems. *Journal of Educational Psychology, 101,* 373-387.

Lee, K., Ng, S. F., Ng, E. L., & Lim, Z. Y. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *Journal of Experimental Child Psychology, 89,* 140-158.

Massaro, D. W. (1975). *Experimental Psychology and Information Processing*. Chicago: Rand McNally.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49-100.

Pickering, S. J., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children*. Kent: The Psychological Corporation.

Rabbitt, P. (1997). Introduction: Methodologies and Models in the Study of Executive Function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1-38). Hove, East Sussex, U.K.: Psychology Press.

Sattler, J. M. (2001). *Assessment of children: cognitive applications*. (4th ed.) San Diego: J. M. Sattler.

Spearman, C. (1927). *The Abilities of Man*. Oxford England: Macmillan.

Stankov, L. (1983). Attention and Intelligence. *Journal of Educational Psychology, 75*.

Stankov, L. (1988). Aging, attention, and intelligence. *Psychology and Aging, 3*.

Stankov, L. (2000). Complexity, Metacognition, and Fluid Intelligence. *Intelligence, 28*.

Stankov, L. & Cregan, A. (1993). Quantitative and qualitative properties of an intelligence test: Series completion. *Learning and Individual Differences, 5,* 137-169.

Stankov, L. & Myors, B. (1990). The relationship between working memory and intelligence: Regression and COSAN analyses. *Personality and Individual Differences, 11,* 1059-1068.

Suess, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability--and a little bit more. *Intelligence, 30,* 261-288.

Swanson, H. L. (2004). Working memory and phonological processing as predictors of children's mathematical problem solving at different ages. *Memory and Cognition, 32,* 648-661.

Swanson, H. L. & Beebe-Frankenberger, M. (2004). The Relationship Between Working Memory and Mathematical Problem Solving in Children at Risk and Not at Risk for Serious Math Difficulties. *Journal of Educational Psychology, 96,* 471-491.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition*. San Antonio, TX: Psychological Corporation.

Author Note

Studies reported in this paper were supported by grants from the Education Research Fund, EP2/02KL and the Office of Educational Research, CRP8/05KL. We thank the children who participated in the studies and the school administrators who provided access and assistance. Views expressed in this article do not necessarily reflect those of the National Institute of Education.