Available online at www.sciencedirect.com
SCIENCE DIRECT®
Intelligence 32 (2004) 277–296
INTELLIGENCE

Working memory is (almost) perfectly predicted by *g*

Roberto Colom[a,*], Irene Rebollo[a], Antonio Palacios[a], Manuel Juan-Espinosa[a], Patrick C. Kyllonen[b]

[a] *Facultad de Psicología, Universidad Autónoma de Madrid (UAM), Madrid 28049, Spain*
[b] *Educational Testing Service (ETS), Princeton, NJ, USA*

Received 19 May 2001; received in revised form 16 December 2003; accepted 22 December 2003
Available online 20 February 2004

Abstract

This article analyzes if working memory (WM) is especially important to understand *g*. WM comprises the functions of focusing attention, conscious rehearsal, and transformation and mental manipulation of information, while *g* reflects the component variance that is common to all tests of ability. The centrality of WM in individual differences in information processing leads to some cognitive theorists to equate it with *g*. There are several studies relating WM with psychometric abilities like reasoning, fluid intelligence, spatial visualization, spatial relations, or perceptual speed, but there are very few studies relating WM with *g*, defined by several diverse tests. In three studies, we assessed crystallised intelligence (Gc), spatial ability (Gv), fluid intelligence (Gf), and psychometric speed (Gs) using various tests from the psychometric literature. Moreover, we assessed WM and processing speed (PS). WM tasks involve storage requirements, plus concurrent processing. PS tasks measure the speed by which the participants take a quick decision about the identity of some stimuli; 594 participants were tested. Confirmatory factor analyses yielded consistently high estimates of the loading of *g* over WM (.96 on average). WM is the latent factor best predicted by *g*. It is proposed that this is so because the later has much in common with the main characteristic of the former.

Keywords: Working memory; *g*; Crystallised intelligence; Spatial ability; Fluid intelligence; Psychometric speed

1. Introduction

Cognitive psychologists have found it useful to develop models of human information processing to assist in the interpretation of human abilities. Kyllonen and Christal (1990) proposed

* Corresponding author. Tel.: +34-91-497-41-14; fax: +34-91-497-52-15.
E-mail address: roberto.colom@uam.es (R. Colom).

0160-2896/$ - see front matter
doi:10.1016/j.intell.2003.12.002

that individual differences in cognition arise from four sources: processing speed (PS), working memory (WM) capacity, the breadth of declarative knowledge, and the breadth of procedural knowledge.

Tests of intelligence reflect different cognitive abilities. Although the test correlations range from .20 to .80, they are all positive and greater than zero. This empirical phenomenon means that all kinds of ability tests measure something in common: the *g* factor. The *g* factor refers to the component variance that is common to all tests of ability (Brody, 1992, 1997; Carroll, 1993, 1995, 1997, 2003; Jensen, 1998; Mackintosh, 1998). The most *g*-loaded tests involve complex cognitive operations (inductive and deductive reasoning, as well as abstraction). Tests with low *g* loadings involve less complex cognitive operations (sensory discriminations, reaction times to simple stimuli, and rote memory). However, a mental test *g* loading is not predictable as sensory modality or knowledge content from specific features; it is more based on the *complexity* of mental operations required (Arend et al., 2003; Roberts, Beh, & Stankov, 1988; Spilsbury, 1992; Vernon & Weese, 1993). *g* is not a measure of specific knowledge, skills, or strategies for problem solving. Rather, it reflects individual differences in information processing, that is, the capacity and efficiency of the mental processes by which knowledge and skills are acquired and used (Jensen, 1998).

The studies published in the last 30 years show that the speed with which people perform simple cognitive tasks (elementary cognitive tasks, ECTs) are correlated with human abilities (Ceci, 1990; Deary, 1995, 2000; Vernon, 1987). Typical ECTs consist of various forms of simple, choice, and discrimination reaction time (RT). Although different ECTs can be devised to elicit different elements of the information processing system, the sources of individual differences in various ECTs do not correspond directly to the different information processes that are hypothesized to be involved in the ECTs (Jensen, 1998).

Kyllonen and Christal (1990) have reported an influential study claiming that reasoning is a little more than WM capacity. These researchers assessed reasoning ability using psychometric tests and WM capacity using tests constructed according to the definition of Baddeley (1986). Confirmatory factor analyses yielded consistently high estimates of the correlation between WM and reasoning ability (.80 to .90), thus giving room for their main conclusion: “if working memory capacity is responsible for differences in reasoning ability, then it may be that working memory capacity affects success across the various component stages of reasoning tasks” (p. 427). However, their results could be interpreted as supportive of the hypothesis that WM is primarily determined by individual differences in reasoning ability (p. 428).

Several years later, Stauffer, Ree, and Carretta (1996) studied the relationship between *g* (obtained from the ASVAB) and a general cognitive factor (obtained from 25 computer-based cognitive-components measures). They found a correlation of .95, concluding that “our results suggest that measurement of human ability, whether by traditional paper-and-pencil tests or by cognitive components, yields, in large part, a measure of *g* (...) the amount of *g* in the common variance among the cognitive-components tests is greater than in traditional paper-and-pencil tests, indicating that the amount of reliable variance attributable to specific abilities is smaller in cognitive-components tests than in traditional paper-and-pencil tests. That finding is contrary to the expectations of some cognitive psychologists” (p. 200–01).

Engle, Tuholski, Laughlin, and Conway (1999) have analyzed several verbal-quantitative WM tasks. They also administered tests of fluid intelligence. Engle et al. have argued that WM capacity and fluid intelligence reflect *the ability to keep a representation active*. Controlled attention ability is proposed as a crucial underpinning of fluid intelligence. Storage requirements were not found as a relevant component of the correlation between WM and fluid intelligence.

Miyake, Friedman, Rettinger, Shah, and Hegarty (2001) measured the central executive component of WM, as well as three spatial abilities, namely, visualization, spatial relations, and perceptual speed. Miyake et al. found that a latent factor representing executive functioning predicted the three latent factors representing visualization, spatial relations, and perceptual speed with values of .91, .83, and .43, respectively. Thus, the three spatial ability factors differ in the degree of executive involvement—highest for visualization and lowest for perceptual speed. They argued that a WM perspective is useful in characterizing the nature of cognitive abilities and human intelligence.

Süb, Oberauer, Wittmann, Wilhelm, and Schulze (2002) administered a battery of 17 WM tasks, together with several psychometric tests of cognitive ability. They found that a good predictor of complex cognitive performance need not necessarily be a combination of storage and processing. Simple span tasks are equally related to reasoning than typical WM tasks. This finding is contrary to those reported by Engle et al. (1999) or by Conway, Cowan, Bunting, Therriault, and Minkoff (2002). Süb et al. hypothesized that WM is a causal factor for intelligence, although they recognised that "a correlational study like this one cannot decide between these alternatives" (p. 276). The S.E.M. model displayed in Fig. 8 of Süb et al. shows that WM is related to *g* in a relatively weak way: .38 with spatial WM and .58 with verbal-quantitative WM. These are weak relations compared with those observed in some previous studies. It is difficult to make a strong case from these values.

Ackerman, Beier, and Boyle (2002) administered 36 ability tests representing verbal, numerical, spatial, and PS abilities; the Progressive Matrices Test; and seven WM tests. Interestingly enough, the CFA model depicted in their Fig. 5 shows a .70 correlation between WM and *g*.

Summing up, previous research has analyzed the relation between WM measures and some cognitive abilities like reasoning ability (Kyllonen & Christal, 1990), fluid intelligence (Colom, Flores-Mendoza, & Rebollo, 2003; Conway et al., 2002; Engle et al., 1999), or spatial ability (Miyake et al., 2001). However, there are only two studies addressing the question of whether WM (as indicated by performance on tasks requiring the switching of control processes forth and back between the representation of the items and the processing component) is especially important to "mark" *g* (as indicated by several diverse conventional ability tests; see Jensen, 1998). Note that none of these two studies defined a general second-order factor predicting psychometric abilities, WM, or PS, defined as primary latent factors. To do just that, we administered a variety of tests designed to measure conventional abilities (crystallised intelligence, spatial ability, fluid intelligence, and psychometric speed), WM capacity, and PS, to samples of psychology undergraduates and Air Force recruits. We selected ability tests from well-known batteries and created WM and PS tasks.

2. Studies 1–3

We administered batteries of paper-and-pencil and computerized tests to three groups of participants over several months. Study methods were similar, and we report them together, although we separate results by study.

3. Method

3.1. Participants

The participants were 198 Psychology undergraduates (Study 1; 145 females and 43 males; mean age = 23.4), 203 U.S. Air Force recruits (Study 2; 155 males and 48 females; mean age = 22.3), and 193 U.S. Air Force recruits (Study 3; 147 males and 46 females; mean age = 22.7). Therefore, a total of 594 participants were tested.

3.2. Testing facilities and procedure

3.2.1. Universidad Autónoma de Madrid

The computerized tests were written in Borland C++. The testing facility consisted of 15 testing stations. Each station was an IBM microcomputer. The tests were applied in five separate sessions and the average testing time was 5 h. Of the 5 h, 3 h were dedicated to psychometric testing, and the remaining 2 h were dedicated to computer testing.

3.2.2. Armstrong laboratory

The computerized tests were written in Turbo Pascal. The testing facility consisted of 30 testing stations. Each station was a Zenith microcomputer. The session lasted around 2.5 h. The program presented the cognitive tasks randomly, inserting breaks at various points during the session. After the computerized session, participants left their testing stations and walked to another room where they solved the paper-and-pencil tests.

3.3. Choice of tasks

Every latent factor was defined by three measures (Table 1). Because the latent factors are what is shared by several tasks used to tap some construct, specific task requirements have less influence on the estimates of construct relations. Moreover, because the measurement error for each task is not part of the latent factor, the later provides a reliable measure of the construct.

The WM tasks were designed to reflect storage requirements, plus concurrent processing (Baddeley & Logie, 1999). In WM tasks, participants must engage in an effortful coordination of processing and storage requirements (Engle et al., 1999). The present study considers quantitative, verbal, and spatial measures of WM. The quantitative task is an adaptation from a task developed by Larson and Sacuzzo (1989), the verbal task is an adaptation from a task used by Hunt (1978), and the spatial task is an adaptation from a task developed by Lohman and Nichols (1985). Studies 1 and 2 considered single digits, letters, and points as stimuli, while Study 3 considered complex digits, words, and lines as stimuli.

The PS tasks were designed to measure the speed by which the participants take a quick decision about the identity of some stimuli. They follow a paradigm popularised by Posner (1978): two stimuli are presented sequentially, and the participant must decide and respond accordingly whether the two stimuli are same or different with respect to a criteria stated in the instructions (Carroll, 1993). Verbal, quantitative, and spatial processing tasks were designed. It must be noted that Study 3 introduced a slight variation in the PS tasks: Two probe stimuli were presented before the

Table 1
Descriptive statistics for the three studies

Tests	Study 1 ($n = 198$)	Study 2 ($n = 203$)	Study 3 ($n = 193$)
	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)
WM			
Counter	26.3 (10.3)	26.3 (15.1)	7.5 (6.8)
Sentence verification	30.1 (9.6)	12.4 (7.9)	8.1 (6.5)
Line formation	8.9 (8.6)	11.3 (9.5)	9.3 (8.3)
PS			
Rectangle–Triangle	39.6 (10.9)	28.5 (10.0)	27.8 (11.1)
Vowel–Consonant	37.8 (9.6)	29.7 (9.5)	25.2 (11.9)
Odd–Even	38.3 (10.5)	29.6 (10.9)	24.7 (10.8)
Gc			
PMA-V	35.0 (6.9)		
DAT-VR	35.2 (6.7)		
Monedas	22.9 (6.5)		
V4		5.7 (2.7)	5.7 (2.5)
V5		4.5 (3.6)	4.2 (3.5)
R4		6.4 (2.8)	6.1 (2.9)
Gv/Gf			
Identical figures	15.7 (3.4)		
Surface development	22.4 (10.6)		
APM	23.9 (4.6)		
VZ3		16.6 (7.9)	13.1 (7.8)
S1		83.9 (22.7)	82.1 (24.8)
I3		40.4 (26.3)	53.4 (20.1)
Gs			
P1		57.8 (15.7)	29.7 (8.3)
P2		53.9 (12.5)	25.4 (7.0)
P3		39.9 (6.4)	40.8 (6.4)

appearance of the target stimulus. The modification involves a small implication of WM (Smith & Jonides, 1997).

Crystallised intelligence (Gc) was measured by three measures: two verbal and one quantitative. Study 1 comprised the vocabulary subtest from the Primary Mental Abilities Battery (PMA; Thurstone, 1938), the verbal reasoning subtests from the differential aptitude test battery (DAT; Bennet, Seashore, & Wesman, 1974), and a quantitative reasoning test called "Monedas" (Seisdedos, 1978; see Colom, Juan-Espinosa, Abad, & García, 2000, for details). Studies 2 and 3 comprised three tests from the ETS kit (Ekstrom, French, & Harman, 1976): V4 (advanced vocabulary), V5 (vocabulary), and R4 (necessary arithmetic operations).

Spatial/fluid abilities (Gv/Gf) were measured by three figurative tests. Study 1 comprised the identical figures test (Manzione, 1978), the surface development test (Ekstrom et al., 1976), and the advanced progressive matrices test (APM). There are several studies noting the figurative *charge* of the APM (e.g., Dillon, Pohlman, & Lohman, 1981; Lim, 1994). DeShon, Chan, and Weissbein (1995) have demonstrated that APM items are biased by its figurative content (see, also, Abad, Colom, Rebollo, & Escorial, in press). Nevertheless, the defined latent factor can be considered as a mixture of spatial ability and fluid intelligence. Studies 2 and 3 comprised the surface development, the card rotations, and the figure classification tests (Ekstrom et al., 1976). As with the APM, the figure classification test can be considered as a measure of spatial ability as well as a measure of fluid intelligence (Wothke et al., 1991). Thus, the defined latent factor can be considered as a mixture of Gv and Gf.

Finally, psychometric speed (Gs) was measured by the finding A's, the number comparison, and the identical picture tests (Ekstrom et al., 1976). The three measures rely on speed, but each measure is based on a different content domain: letters, numbers, and pictures. The consideration of these content domains helps to define a representative latent factor.

Appendix A describes the tests and tasks in much more detail.

3.4. Analyses

The present investigation is based on an assessment using traditional tests of intelligence, and PS and WM computerized tasks. The participants' performance on these tasks is analyzed using confirmatory factor analysis (CFA) techniques. CFA analyses were performed through LISREL 8.5 (Jöreskog & Sörbom, 2001).

LISREL 8.5 uses maximum-likelihood estimation to derive the specific parameters, based on the correlation matrix (see Appendix B). Because there is no clear consensus as to the best-fit indices for the evaluation of CFA models, we followed the recommendation of Browne and Cudeck (1992). Those researchers focus on the root-mean square error of approximation (RMSEA) fit index. RMSEA is an estimate of the discrepancy between the model and the data per degree of freedom for the model; values less than .05 constitute good fit, values in the .05 to .08 range acceptable fit, values in the .08 to .10 range marginal fit, and values greater than .10 poor fit. The goodness of fit index (GFI) is also considered in the present investigation. GFI must be higher than .9; lower values reflect marginal fit. Attending to the requirement of one reviewer, we also report AGFI and CFI indexes (see Figs. 1, 2, and 3).

The model tested is aligned with the criteria used for selecting the psychometric tests and the cognitive tasks. Therefore, it was postulated a structure with four first-order latent factors in Study 1 and five first-order latent factors in Studies 2 and 3 (see Table 1). The main goal is to test the status of WM within a structure, in which a second-order factor (*g*) predicts the first-order latent factors (Yung, Thissen, & McCleod, 1999). If the tested model ranges acceptable fit, then we can look at the loading of *g* over working memory. If the loading is high enough (.9 or higher; see Kyllonen & Christal, 1990), then, it must be concluded that WM is (almost) perfectly predicted by *g*.

It is important to note that we are not interested in testing alternative models. We have a clear prediction to test: Is WM the latent factor best predicted by a higher order factor representing *g*? If this is the case, then it would be reasonable to state that this is because *g* has much in common with the key characteristic of WM, namely, storage requirements, plus concurrent processing. People differ in their ability to perform those cognitive activities, thus, the best measures of *g* could be those that impose a greater stress over the person's WM, as suggested by Jensen (1998).

One reviewer stated that we adopted a questionable view. Why do we predict first-order factors from a general second-order factor? It would be equally interesting to use WM as a predictor of *g*. Although we think that this approach is valuable, we still think that confirmatory models were designed with the purpose of testing theoretical predictions. Süb et al. (2002) considered WM as a predictor of *g*, Ackerman et al. (2002) studied the correlation between WM and *g*, while the present studies considered *g* as the predictor of WM and other latent factors to answer the question of whether WM is the latent factor best predicted by *g*. Finally, changing the direction of the causal path or transforming it to a correlation path would not change the quantity of the loading or the general fit of the model. It must be remembered that since the program works with correlations, the three theoretical models are equivalent, and we cannot infer causality.

4. Results

4.1. Study 1

Table 1 shows the descriptive statistics for the tests administered in Study 1. The tested model is a higher order factor model. There are four first-order latent factors: WM, PS, crystallised intelligence (Gc), and spatial/fluid intelligence (Gv/Gf). Every latent factor is defined by three measures. A higher order factor identified with *g* predicts the four first-order factors. The RMSEA value was .074, lower than .08, suggesting an acceptable fit. The GFI value was .92, higher than .9, suggesting a good fit.

The model is displayed in Fig. 1. The highest weigh of *g* over the first-order factors corresponds to WM (1.04), while the lowest loading is for PS (.62). The remaining first-order factors show relatively high loadings. The results indicate that WM is predicted (almost) perfectly by *g*.

4.2. Study 2

Table 1 shows the descriptive statistics for the tests administered in Study 2. The same higher order factor model was tested in Study 2, although five first-order factors were postulated this time: WM, PS, crystallised intelligence, spatial/fluid intelligence, and psychometric speed (Gs). A higher order factor thought to represent *g* predicts the five first-order factors. The RMSEA value was .044, lower than .5, suggesting a good fit. The GFI value was .93, higher than .9, suggesting a good fit.

The model is displayed in Fig. 2. Like in Study 1, the highest loading of *g* over the five first-order factors corresponds to WM (.90), while the lowest loading corresponds to psychometric speed (.49). Crystallised intelligence and PS show relatively low loadings (.59 and .52, respectively) while spatial/fluid intelligence show a high loading (.83). Therefore, like in Study 1, *g* predicts very nicely the WM latent factor.

4.3. Study 3

Table 1 shows the descriptive statistics for the tests administered in Study 3. The same model as in the previous studies was tested. The RMSEA value was .06. Because the value is lower than .08, the model ranges acceptable fit. The GFI value was .91, higher than .9, suggesting a good fit.

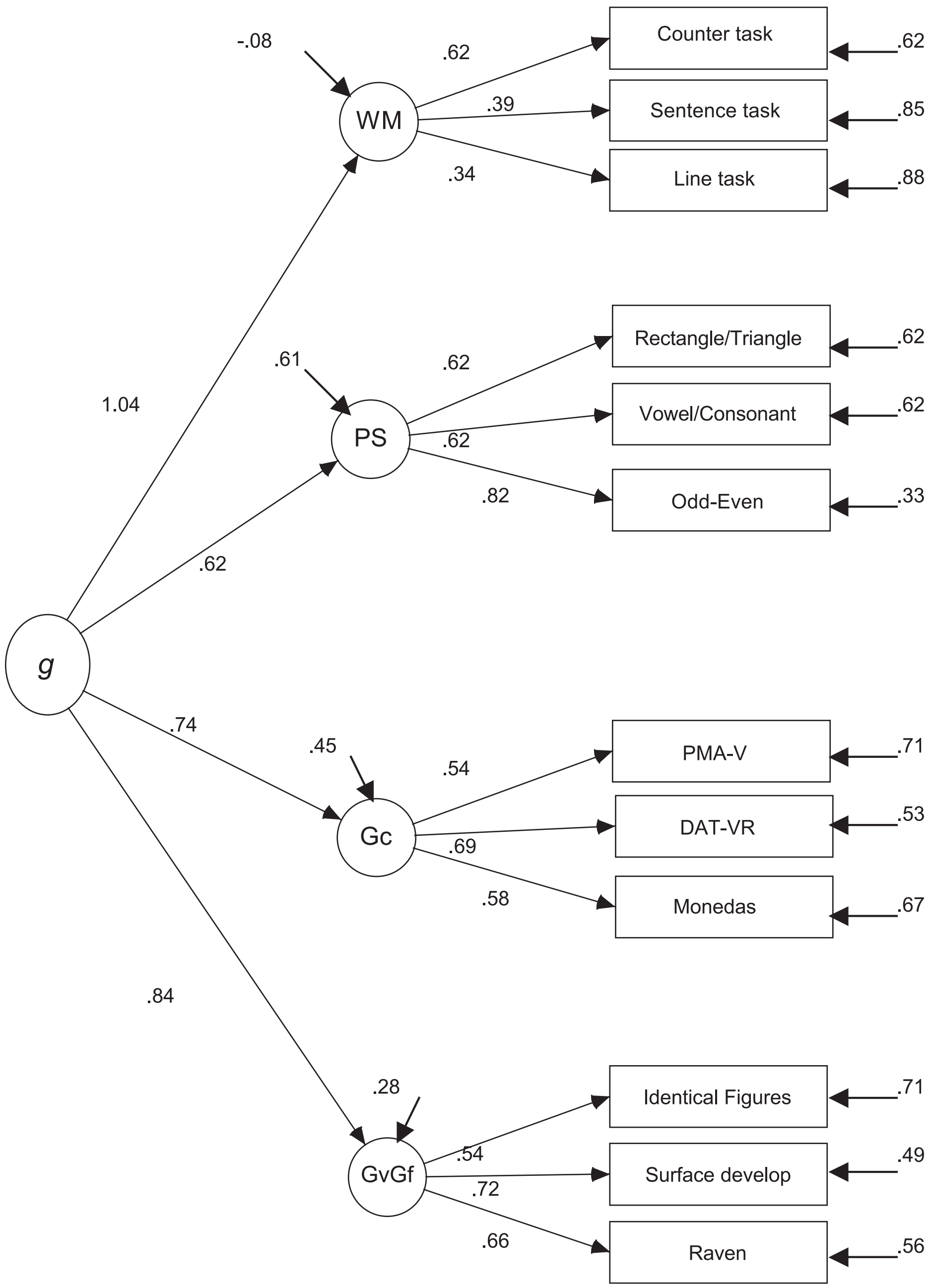

Fig. 1. CFA model of Study 1. $\chi^2 = 105.898$, $df = 50$, CMIN/DF = 2.12, GFI=.92, RMSEA=.073 (range=.053/.094). AGFI=.874. CFI=.889. Standardized parameters are shown.

The model is displayed in Fig. 3. The highest loading of *g* over the five first-order factors corresponds to WM (.93), while the lowest loading corresponds to psychometric speed (.45). Crystallised intelligence and spatial/fluid intelligence show relatively low loadings (.48 and .48, respectively).

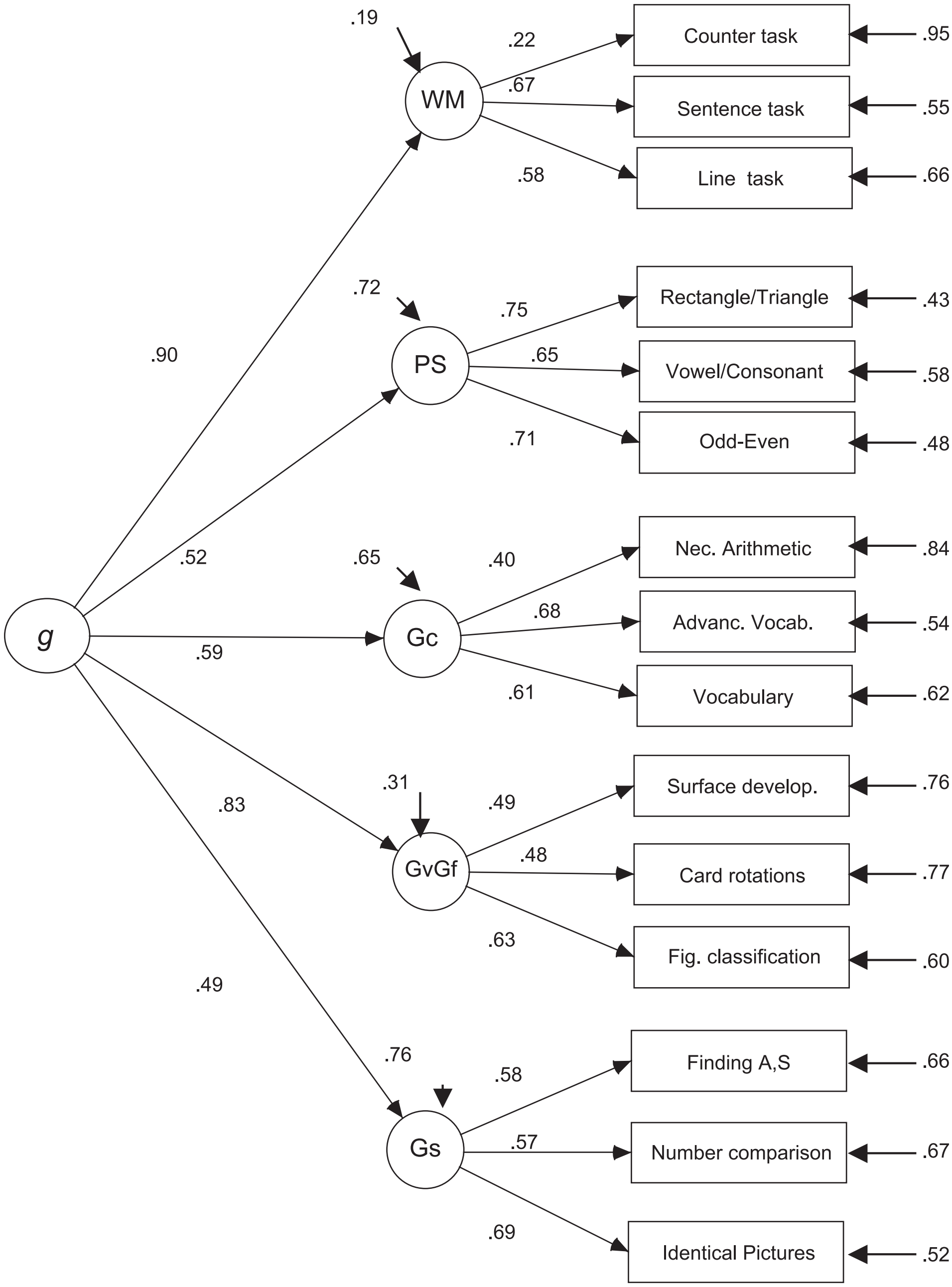

Fig. 2. CFA model of Study 2. $\chi^2 = 121.859$, $df = 85$, CMIN/DF = 1.43, GFI=.927, RMSEA=.044 (range=.023/.062). AGFI=.897. CFI=.925. Standardized parameters are shown.

There is a thought-provoking change in this model when it is compared with the models in the previous studies. The loading of *g* over the PS factor increases. Now, the loading is .76, a value only surpassed by the loading of *g* over the WM factor.

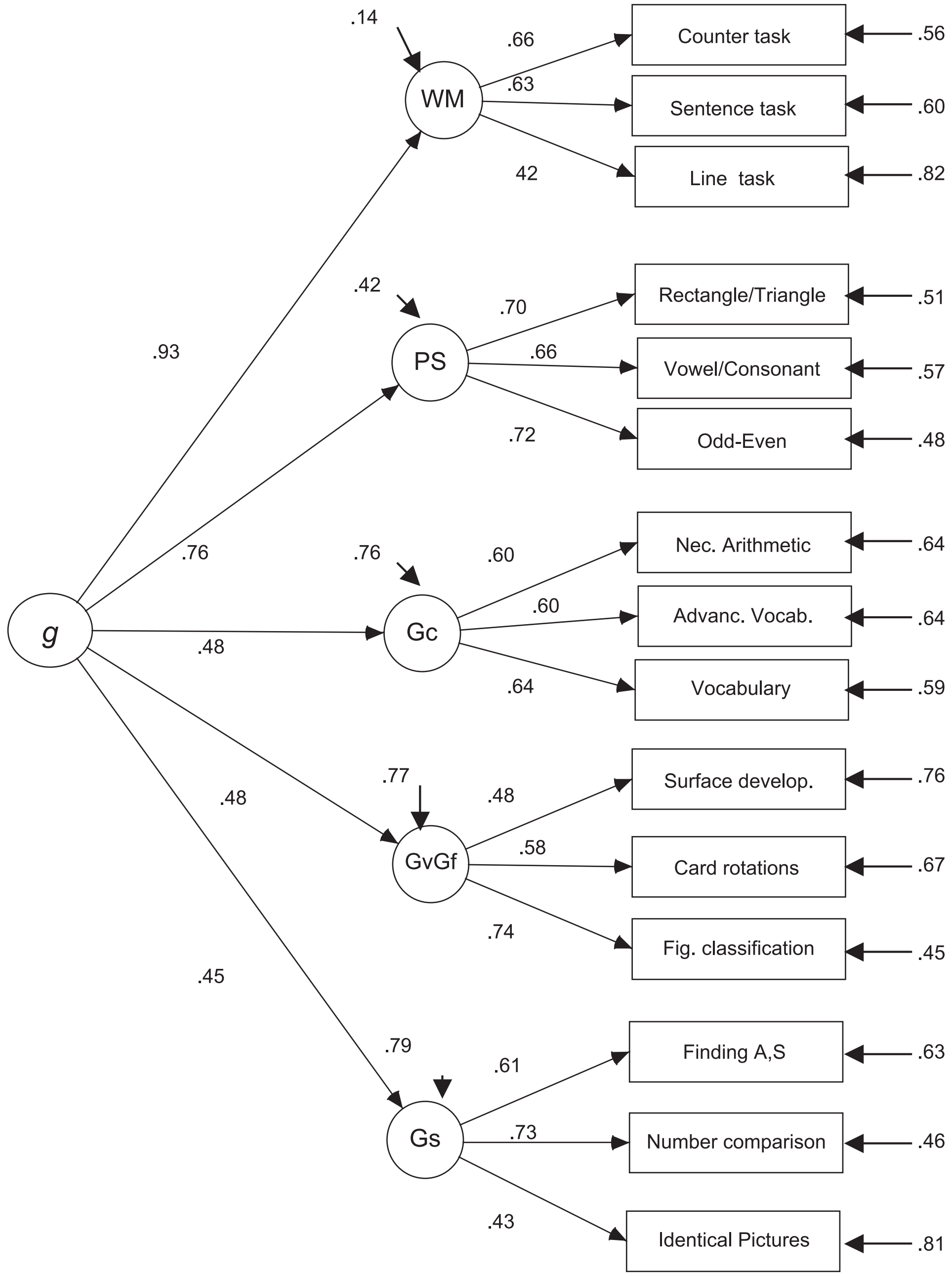

Fig. 3. CFA model of Study 3. $\chi^2 = 158.981$, $df = 85$, CMIN/DF = 1.9, GFI=.912, RMSEA=.059 (range=.043/.076). AGFI=.875. CFI=.872. Standardized parameters are shown.

It is important to discuss the nature of the change observed by the PS factor. Remember that the PS tasks were modified in Study 3: The participant must consider one or two probe stimuli before taking a quick decision about their identity with the test stimulus. Therefore, there is a small WM "ingredient" within the modified PS tasks. Thus, the introduction of a WM load, even very soft, produces an increase in the *g* loading of exactly the same task.

In whole, Study 3 agrees with the previous studies: WM is predicted (almost) perfectly by *g*. Moreover, Study 3 is a demonstration of the type of manipulation that could change the cognitive requirements imposed over the person: It essentially consists in an increase of the task cognitive complexity, that is, an increase in the stress imposed over the person's WM.

5. General discussion

Kyllonen and Christal (1990) found correlations around .9 between WM and reasoning ability. Engle et al. (1999) found a correlation of .49 between WM and fluid intelligence. Conway et al. (2002) found a correlation of .60 between verbal-quantitative WM and fluid intelligence. Miyake et al. (2001) found a correlation of .91 between executive functioning (EC) and spatial visualization, of .83 between EC and spatial relations, and of .43 between EC and perceptual speed. Colom et al. (2003) found a correlation of .70 between WM and fluid intelligence. Thus, previous research has studied the relationships between the construct of WM and some cognitive abilities. These studies do not test a more direct hypothesis about the relationship between the central factors of the human-abilities model and the theories of human information processing. The central factor in the human-abilities model is *g*, not fluid intelligence, spatial visualization, spatial relations, or perceptual speed, while the central factor in the theories of human information processing is WM (Jensen, 1998; Lohman, 2000). An adequate representation of psychometric *g* requires a number and variety of tests (Jensen & Weng, 1994). The only two studies relating WM with *g* are the ones performed by Ackerman et al. (2002) and Süb et al. (2002). The former study found a correlation of .70 between WM and *g*, while the later study found a correlation of .38 between spatial WM and *g*, and of .58 between verbal-quantitative WM and *g*. However, these two studies did not test the predictive power of *g* over WM and against other first-order latent factors. CFA models are informative with respect to the predictive power of *g* over WM. Considering the three reported studies in the present article, the loading of *g* over the WM latent factor averages .96. Therefore, WM is (almost) perfectly predicted by *g* (92% of explained variance).

Widaman notes in his review of the present article that there are low correlations among the WM tasks in some instances. True, but Miyake et al. (2001) have noticed that although the correlation among WM tasks are frequently lower than with other within-construct correlations "zero-order correlations of this magnitude (often .30 or less) are common among executive (WM) tasks, partly because these complex tasks often involve a good deal of variance related to non-executive processes as well as measurement error" (p. 630). Fortunately, latent-variable analysis is particularly useful in these circumstances because the analysis extracts the common variance between the tasks chosen to tap working memory.

Another reviewer highlights the fact that the loadings of individual tasks fluctuate across studies. The reviewer thinks that this fact could help to support the results observed with the datasets considered in the present article.

For comparative purposes, we reanalyzed several tests and the tasks taken from the Ackerman et al. (2002) study. The selection was made to test a model as close as possible to the ones considered in the three studies reported here. The analyzed tasks were the following: verbal ability—vocabulary, MAB similarities, and completion; quantitative ability—math knowledge, arithmetic, and math approximation; spatial ability—paper folding, spatial analogy, and cube comparison; working memory—ABCD order, computation span, and spatial span; processing speed—number sorting, digit/symbol, finding € and ¥, and directional headings I. The tested model was an hierarchical one: *g* predicted WM, PS, verbal ability, quantitative ability, and spatial ability. The results strongly supported the observed here: WM was the latent first-order factor best predicted by *g* (.94). The remaining regression loadings were .72 for PS, .53 for the verbal factor, .48 for the quantitative factor, and .84 for the spatial factor. The RMSEA value was .058, and the Chi-square/*df* was 1.456, which suggests a good fit for the model.

WM comprises the functions of focusing attention, conscious rehearsal, and transformation and mental manipulation of information received from external sources or retrieved from long-term memory traces. WM can deal only with a limited amount of information. The capacity of WM is expressed as the maximum amount of available activation to support storage and processing. When the task demands exceed the available resources, both the storage and computational functions experiment a breakdown. Individual differences in the amount of capacity to be allocated can account for the systematic differences in performance. Within any task domain, individual differences will emerge when the task demands consume sufficient capacity to exhaust some person's resources. These individual differences have much to do with individual differences in the overall capacity and efficiency of mental processes, as was proposed by Jensen (1998). This could help to explain why WM is (almost) perfectly predicted by *g*. In fact, across the three reported studies, WM is the latent factor best predicted by *g*.

The results in Study 3 are in strong agreement with that view. Study 3 introduced a subtle change in the PS tasks: The participant must consider one or two probe stimuli before taking a quick decision about their identity with the test stimulus. This supposed a small WM "charge." The effect was a higher loading of *g* over the PS factor. Therefore, the greater the strain imposed over WM, the higher the *g* required.

Functions like monitoring the contents of WM, switching between tasks requiring WM, applying a rule that may be kept active in WM, and planning a set of actions in a problem-solving task have some component in common. The component could be something that monitors operations performed on WM or something responsible for the maintenance of the goal structure needed to guide processing in any cognitive task. Is this component the essence of *g*?

Although the findings reported in the present study support a positive answer, more research is obviously needed. We agree with the statement of Miyake et al. (2001) that "a better understanding is likely to illuminate the nature of human intelligence and help bring cognitive theories and psychometric theories into closer alignment" (p. 639). However, WM tasks able to separate temporary storage and controlled processing are needed. There are numerous WM tasks reflecting temporary storage plus controlled processing, but STM tasks cannot be considered as reflecting temporary storage with no executive involvement (Colom, Abad, Rebollo, Flores-Mendoza, & Botella, 2002; Süb et al., 2002). New tasks should be developed to separate the effect of temporary storage from controlled processing. Then, those facets of WM must be correlated with *g*. If the perspective of Engle et al. (1999) is right, then controlled processing, but not temporary storage, should be responsible of the fact that WM is almost perfectly predicted by *g*. But if the alternative perspective is right, then, the predictive power of *g* will increase as the demands on both temporary storage and concurrent processing increase (Bayliss, Jarrold, Gunn, & Baddeley, 2003). Testing those predictions will surely help to answer one key question: Why is WM the latent factor best predicted by *g*?

Acknowledgements

The research referred to in this article was supported by grants funded by the European Office of Aerospace Research and Development (EOARD, WOS 96-2006), the Spanish's Ministerio de Ciencia y Tecnología (Grants PB98-0066 and BSO-2002-01455) and the Spanish's Ministerio de Educación, Cultura y Deportes (Grant AP2001-0016). We would like to thank Francisco J. Abad for his helpful comments.

Appendix A. Task description

Following are task descriptions organized by factor: WM, PS, crystallised intelligence (Gc), spatial/fluid intelligence (Gv/Gf), and psychometric speed (Gs). The computerized WM and PS tasks were originally programmed in Borland 3.0 C++, although they were rewritten in Turbo Pascal for their application in the Armstrong Laboratory.

A.1. Working memory tasks

The WM tasks were preceded by practice sessions. The person does not go to the task itself until he/she reaches 80% of correct responses.

Counter Task (adapted from Larson & Sacuzzo, 1989). Fig. A.1.1 shows a typical display.

Let us take a look at one example (Fig. A.1.1). The task begins with a screen with three horizontal lines (counters) and the number 1 at the top left half of the computer screen. The second screen shows the number 1 at the bottom left half of the computer screen. Because this number is preceded by a minus sign, the participant must subtract both numbers. The third screen shows the number 1 at the top right half of the computer screen. The answer screen shows the results of the computations the participant is required to perform. He/She must decide if the answer screen is correct or incorrect.

The task consists of 68 trials varying in the number of counters and in the computations required. Twenty trials consider two counters and four to eight computations, 24 trials consider three counters and three to eight computations, and 24 trials consider four counters and three to eight computations. The person responds YES when the result of the computations corresponds to the number displayed in the answer screen, and responds NOT in any other case.

Studies 1 and 2 displayed the same screens and numbers (single digits), while Study 3 displayed complex digits (two-number digits) to make the computations harder.

The reliability estimates (Cronbach's alpha) were .70 (Study 1), .88 (Study 2), and .76 (Study 3).

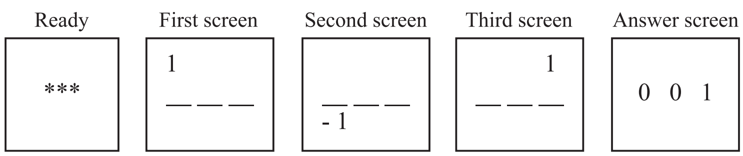

Fig. A.1.1. Counter task.

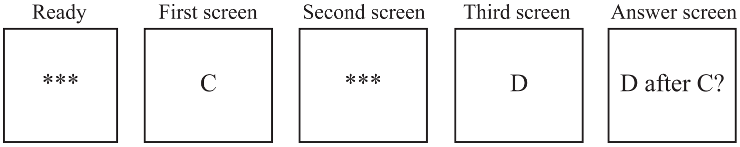

Fig. A.1.2. Sentence-verification task.

Sentence-verification task (adapted from Hunt, 1978). Fig. A.1.2 shows a typical display.

Let us take a look at one example (Fig. A.1.2). The first screen shows a given letter (C). The third screen shows another letter (D). The answer screen displays a question about the relative positions of the letters shown previously (D after C?). The participant must decide if the answer is yes or no.

The task consists of 80 trials, with an amount of sequentially displayed single letters ranging from two to eight. The sentence types were the result of the combination of the terms "before," "after," and "between" in affirmative or negative statements (after or not after, before or not before, and so forth).

Studies 1 and 2 displayed the same stimuli, but in Study 3, the letters were substituted by words.

The reliability estimates (Cronbach's alpha) were .79 (Study 1), .82 (Study 2), and .83 (Study 3).

Line formation task (adapted from Lohman & Nichols, 1985). Fig. A.1.3 shows a typical display.

Let us take a look at one example (Fig. A.1.3). The first screen shows a point at the top left half of the computer screen. The second screen shows a point at the bottom left half of the computer screen. The third screen shows a point at the middle bottom half of the computer screen. The answer screen displays the line that could be represented following the path of the points presented previously. The participant must decide if the line corresponds to the sequence of the points presented.

The task includes 70 trials, 10 with three points, 30 with four points, and 30 with five points. The person responds YES if the test line is simply rotated or displaced, but responds NO if the test line is inverted or transformed.

Studies 1 and 2 displayed points, while Study 3 displayed simple lines to be mentally combined in a complex line. The number of simple lines was between two and four.

The reliability estimates (Cronbach's alpha) were .75 (Study 1), .78 (Study 2), and .72 (Study 3).

A.2. Processing speed tasks

The PS tasks were inspired by the paradigm popularised by Posner (1978). The person must take a decision about the identity of spatial, verbal, and quantitative stimuli. Fig. A.1.4 shows a typical sequence.

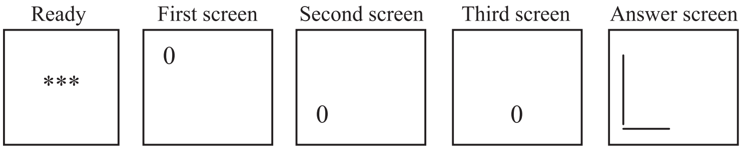

Fig. A.1.3. Line formation task.

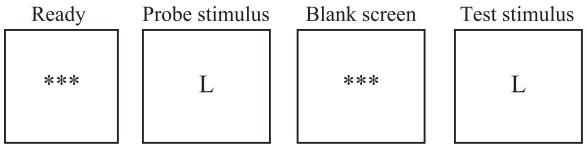

Fig. A.1.4. Processing speed tasks (example of the vowel–consonant PS tasks).

The PS tasks were preceded by practice sessions. The person does not go to the test itself until he/she reaches 80% of correct responses.

The identity tasks were based on the squares' filling type (rectangle or triangle), the type of letter (vowel–consonant), and the type of number (odd–even). Therefore, the person must take a quick decision about the identity between the probe figure and the test figure (square filled by a rectangle or by a triangle), if successive letters are vowel or consonant, and if successive numbers are odd or even (rectangle–triangle, vowel–consonant, and odd–even tasks, respectively).

Every PS task includes 60 trials, of which 50% claim for a YES response. In Study 1, the person does not receive feedback, while in Studies 2 and 3, there is an immediate feedback (a beep signal) about the response correctness. Moreover, in Study 3, the PS tasks can present two or three sequential squares, letters, or numbers.

The reliability estimates (Cronbach's alpha) for the spatial tasks were .86 (Study 1), .88 (Study 2), and .91 (Study 3). The reliability estimates for the verbal tasks were .57 (Study 1), .79 (Study 2), and .91 (Study 3). The reliability estimates for the quantitative tasks were .83 (Study 1), .88 (Study 2), and .91 (Study 3).

A.3. Crystallised tests (Gc)

PMA-V (Thurstone, 1938). This is a vocabulary test, in which the person must decide which one of several probe words has the same meaning as a target word. This test is extracted from the primary mental abilities (PMA) battery.

DAT-VR (Bennett et al., 1974). This is a test of verbal reasoning extracted from the differential aptitude test (DAT). The person must decide which pair of words fits a given sentence where the first and the last words are deleted.

Monedas (Seisdedos, 1978). This test is based on the combination of the size of a series of coins, the digits put inside the coins to specify the number of them that the subject must take into account, and some numerical operations to make the necessary calculations to arrive at a given response (adding, subtracting, etc.). It correlates r=+.64 with the numerical ability (NA) scale from the DAT.

Advanced vocabulary (V4; Ekstrom et al., 1976). This is a test about the person's knowledge of word meanings. The items' structure is the same as in the PMA-V.

Vocabulary (V5; Ekstrom et al., 1976). This is a test based on the knowledge of word meanings. It is very similar to V4.

Necessary arithmetic operations (R4; Ekstrom et al., 1976). This test consists of problems in mathematics. Instead of solving the problem and finding the answer, the person must indicate which arithmetic operation could be used if the problem is to be solved.

A.4. Spatial/Fluid tests (Gv/Gf)

Identical figures (Manzione, 1978). This is a test in which the person is asked to compare a given abstract figure with several abstract test figures. The person must decide which of the test figures is identical to the probe figure.

Raven progressive matrices. The APM was used in the present investigation.

Surface development (VZ3; Ekstrom et al., 1976). The person must visualize how a piece of paper can be folded to form a given object.

Card rotations (S1; Ekstrom et al., 1976). This is a test of the ability to see differences in figures. There is a probe figure and several test figures. Each problem consists of one figure on the left of a vertical line and eight figures on the right. The person must decide whether each of the eight cards on the right is the "same as" or "different from" the card at the left.

Figure classification (I3; Ekstrom et al., 1976). This is a test of the ability to discover rules that explain things. In each problem, there are either two or three groups, each consisting of three figures. The person must look for something that is the same about the three figures in any one group and for things that make the groups different from one another.

A.5. Psychometric speed tests (Gs)

Finding A's (P1; Ekstrom et al., 1976). This is a test of the person's speed in finding the letter "a" in words. The person must mark the words including an "a."

Number comparison (P2; Ekstrom et al., 1976). This test asks the person to quickly compare two numbers and decide whether they are the same.

Identical pictures (P3; Ekstrom et al., 1976). This test asks the person to match quickly a given object. This test assesses the ability to pick a correct object quickly. It is very similar to the previous test.

Appendix B. Correlation matrices

B.1. Study 1

Tests	1	2	3	4	5	6	7	8	9	10	11	12
1. Counter		.21	.24	.34	.42	.41	.19	.19	.32	.30	.35	.32
2. Sentence verification			.11	.14	.12	.19	.19	.22	.17	.14	.23	.29
3. Line formation				.07	.09	.16	.17	.09	.04	.26	.30	.12
4. Rectangle–Triangle					.40	.50	.18	.13	.06	.20	.19	.10
5. Vowel–Consonant						.49	.20	.04	.16	.13	.06	.17
6. Odd–Even							.26	.22	.29	.26	.21	.24
7. PMA-V								.43	.21	.28	.20	.18
8. DAT-VR									.39	.20	.32	.42
9. Monedas										.25	.35	.33
10. Identical figures											.39	.30
11. Surface development												.49
12. Raven												

B.2. Study 2, Top half, Study 3, Bottom half

Tests	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Counter		.14	.11	.07	.04	.12	.07	.13	.12	.11	.14	.07	.04	.06	.04
2. Sentence verification	.42		.40	.28	.21	.32	.25	.25	.22	.26	.17	.29	.22	.07	.13
3. Line formation	.21	.33		.25	.23	.27	.13	.10	.11	.32	.21	.22	.14	.12	.12
4. Rectangle–Triangle	.37	.26	.18		.51	.53	.17	.06	.00	.07	.18	.05	.22	.11	.02
5. Vowel–Consonant	.37	.38	.19	.46		.43	.13	.12	.07	.12	.04	.11	.15	.15	.10
6. Odd–Even	.35	.32	.26	.52	.45		.25	.15	.16	.13	.17	.13	.22	.24	.11
7. R4	.30	.16	.15	.19	.15	.19		.23	.17	.37	.21	.31	.11	.06	.14
8. V4	.07	.13	.20	.04	.01	.03	.33		.46	.20	.10	.22	.01	.11	.11
9. V5	.19	.11	.16	.10	.00	.15	.35	.44		.14	.14	.16	.07	.13	.17
10. VZ3	.24	.19	.11	.08	.11	.10	.42	.13	.19		.33	.31	.14	.06	.27
11. S1	.12	.13	.14	.15	.10	.10	.26	.13	.06	.37		.18	.19	.15	.35
12. I3	.17	.16	.14	.04	.06	.11	.40	.29	.28	.43	.24		.03	.03	.18
13. P1	.18	.13	.01	.24	.22	.23	.19	.12	.03	.12	.15	.10		.34	.39
14. P2	.16	.15	.07	.21	.25	.22	.18	.02	.03	.03	.15	.13	.44		.41
15. P3	.02	.01	.06	.06	.15	.21	.06	.16	.02	.14	.27	.23	.23	.34	

Appendix C. Unstandardized parameter estimates and standard errors for Models 1, 2, and 3

Factor	Parameter	Model 1		Model 2		Model 3	
		Unstandardized estimate	S.E.	Unstandardized estimate	S.E.	Unstandardized estimate	S.E.
WM	*g*—WM	0.405	0.083	0.598	0.084	0.586	0.086
	WM—counter task	1.589	0.356	0.337	0.132	1.050	0.169
	WM—sentence task	1.000		1.000		1.000	
	WM—line task	0.876	0.261	0.876	0.159	0.665	0.144
	Counter task residual	0.618	0.095	0.950	0.097	0.561	0.083
	Sentence task residual	0.849	0.092	0.557	0.093	0.601	0.083
	Line task residual	0.884	0.093	0.661	0.088	0.824	0.090
	WM residual variance	− **0.013**	0.031	0.085	0.080	**0.055**	0.066
PS	*g*—PS	0.509	0.075	0.372	0.073	0.543	0.079
	PS—rectangle–triangle	0.757	0.108	1.050	0.140	0.978	0.129
	PS—vowel–constant	0.752	0.108	0.903	0.126	0.918	0.125
	PS—Odd–even	1.000		1.000		1.000	
	Rectangle–triangle residual	0.616	0.078	0.437	0.076	0.510	0.073
	Vowel–constant residual	0.621	0.078	0.583	0.076	0.569	0.074
	Odd–even residual	0.330	0.081	0.489	0.075	0.488	0.073
	PS residual variance	0.411	0.094	0.373	0.084	0.217	0.072
Gc	*g*—Gc	0.424	0.073	0.400	0.079	0.289	0.069
	Gc—Necessary arithmetic			0.591	0.148	0.994	0.193
	Gc—advanced vocabulary			1.000		1.000	
	Gc—vocabulary			0.899	0.187	1.072	0.207
	Gc—PMA-V	0.939	0.181				
	Gc—DAT-VR	1.194	0.210				

Appendix C (*continued*)

Factor	Parameter	Model 1		Model 2		Model 3	
		Unstandardized estimate	S.E.	Unstandardized estimate	S.E.	Unstandardized estimate	S.E.
Gc	Gc—Monedas	1.000					
	Necessary arithmetic residual			0.840	0.093	0.645	0.091
	Advanced vocabulary residual			0.542	0.104	0.641	0.091
	Vocabulary residual			0.629	0.096	0.587	0.094
	PMA-V residual	0.708	0.087				
	DAT-VR residual	0.528	0.087				
	Monedas residual	0.669	0.086				
	Gc Residual variance	0.151	0.055	0.298	0.096	0.276	0.082
GvGf	*g*—Gv/Gf	0.561	0.076	0.405	0.080	0.232	0.061
	Gv/Gf—surface development			1.000		1.000	
	Gv/Gf—card rotations			0.986	0.233	1.193	0.251
	Gv/Gf—figure classification			1.288	0.274	1.537	0.332
	Gv/Gf—identical figures	0.807	0.136				
	Gv/Gf—surface development	1.076	0.151				
	Gv/Gf—raven	1.000					
	Surface development residual			0.762	0.091	0.766	0.090
	Card rotations residual			0.769	0.091	0.668	0.092
	Figure classification residual			0.606	0.094	0.448	0.110
	Identical figures residual	0.712	0.083				
	Surface development residual	0.488	0.076				
	Raven residual	0.558	0.077				
	Gv/Gf residual variance	0.127	0.056	0.074	0.045	0.180	0.063
Gs	*g*—Gs			0.278	0.068	0.331	0.078
	Gs—finding A's			1.012	0.198	0.830	0.175
	Gs—number comparison			1.000		1.000	
	Gs—identical pictures			1.216	0.236	0.591	0.140
	Finding A's residual			0.666	0.090	0.633	0.097
	Num comparison residual			0.674	0.090	0.466	0.114
	Identic pictures residual			0.518	0.098	0.814	0.092
	Gs residual variance			0.248	0.076	0.424	0.119

Parameters nonsignificantly different from zero are in bold face.

References

Abad, F. J., Colom, R., Rebollo, I., & Escorial, S. (in press). Sex differential item functioning in the Raven's Advanced Progressive Matrices: Evidence for bias. *Personality and Individual Differences.*

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology. General*, *131*(4), 567–589.

Arend, I., Colom, R., Botella, J., Contreras, Ma. J., Rubio, V., & Santacreu, J. (2003). Quantifying cognitive complexity: Evidence from a reasoning task. *Personality and Individual Differences*, *35*(3), 659–669.

Baddeley, A. (1986). *Working memory.* London: Oxford Univ. Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: Cambridge Univ. Press.

Bayliss, D. M., Jarrold, C., Gunn, D., & Baddeley, A. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology. General*, *132*(1), 71–92.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). *Differential aptitude test* (p. 1995). Madrid: TEA, S.A.

Brody, N. (1992). *Intelligence* (2nd edition). San Diego: Academic Press.

Brody, N. (1997). Intelligence, schooling, and society. *American Psychologist*, *52*(10), 1046–1050.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, *21*, 230–258.

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge: Cambridge Univ. Press.

Carroll, J. B. (1994). Cognitive abilities: Constructing a theory from data. In D. K. Detterman (Ed.), *Current topics in human intelligence*. New Jersey: Ablex.

Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence*, *24*(1), 25–52.

Carroll, J.B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). Elsevier/Pergamon.

Ceci, S. (1990). *On intelligence...more or less, a bioecological treatise on intellectual development*. Englewood Cliffs, NJ: Prentice-Hall.

Colom, R., Abad, F. J., Rebollo, I., Flores-Mendoza, C., & Botella, M. (2002). *Short-term memory and working memory are conditioned by the same underlying capacity limitations.* Paper presented at the 43rd Annual Meeting of the Psychonomic Society. Kansas City, November 21–24.

Colom, R., Flores-Mendoza, C., & Rebollo, I. (2003). Working memory and intelligence. *Personality and Individual Differences*, *34*, 33–39.

Colom, R., Juan-Espinosa, M., Abad, F., & García, L. F. (2000). Negligible sex differences in general intelligence. *Intelligence*, *28*(1), 57–68.

Conway, A., Cowan, N., Bunting, M., Therriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183.

Deary, I. (1995). Auditory inspection time and intelligence: What is the causal direction. *Developmental Psychology*, *31*, 237–250.

Deary, I. (2000). Simple information processing and intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 267–284). Cambridge: Cambridge Univ. Press.

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. *Intelligence*, *21*, 135–155.

Dillon, R. F., Pohlmann, J. T., & Hohman, D. F. (1981). A factor analysis of Raven's Advanced Progressive Matrices freed of difficulty factors. *Educational and Psychological Measurement*, *41*, 1295–1302.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

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*(3), 309–331.

Hunt, E. B. (1978). Mechanics of verbal ability. *Psychological Review*, *85*(2), 109–130.

Jensen, A. (1998). *The g factor*. Westport: Praeger.

Jensen, A., & Weng, L. (1994). What is a good *g*? *Intelligence*, *18*, 231–258.

Jöreskog, K., & Söborn, D. (2001). *LISREL 8: User reference guide* (2nd edition). Lincolnwood, IL: Scientific Software International.

Kyllonen, P., & Christal, R. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence*, *14*, 389–433.

Larson, G. E., & Sacuzzo, D. P. (1989). Cognitive correlated of general intelligence: Toward a process theory of *g*. *Intelligence*, *13*, 5–31.

Lim, T. K. (1994). Gender related differences in intelligence: Application of confirmatory factor analysis. *Intelligence*, *19*, 179–192.

Lohman, D., & Nichols, P. D. (1985). Spatial ability: The effects of encoding processes and representational quality on mental synthesis. *Annual Meeting of the American Association for Educational Research, San Francisco*.

Lohman, D. F. (2000). Complex information processing and intelligence. In R. J. Sternberg (Ed.), *Handbook human intelligence* (pp. 285–340). Cambridge: Cambridge Univ. Press.

Mackintosh, N. J. (1998). *IQ and human intelligence*. Oxford: Oxford Univ. Press.

Manzione, J. M. (1978). *B.F.A.—Batería Factorial de Aptitudes* [*Factorial abilities battery*]. Madrid: TEA, S.A.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology. General*, *130*(4), 621–640.

Posner, M. (1978). *Chronometric explorations of mind*. Hillsdae, NJ: Erlbaum.

Roberts, R., Beh, H. C., & Stankov, L. (1988). Hicks' law, competing-task performance, and intelligence. *Intelligence*, *12*, 111–130.

Seisdedos, N. (1978). *Manual del test de Monedas* [*Handbook of the Monedas test*]. Madrid: TEA, S.A.

Smith, E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*, 5–42.

Spilsbury, G. A. (1992). Complexity as a reflection of the dimensionality of a task. *Intelligence*, *16*, 31–45.

Stauffer, J., Ree, M., & Carretta, T. (1996). Cognitive-components tests are not much more than *g*: An extension of Kyllonen's analyses. *The Journal of General Psychology*, *123*(3), 193–205.

Süb, H., Oberauer, K., Wittman, W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—And a little bit more. *Intelligence*, *30*, 261–288.

Thurstone, L. (1938). Primary mental abilities. *Psychometric Monographs*, *1*.

Vernon, P. (1987). *Speed of information processing and intelligence*. Norwood, NJ: Ablex.

Vernon, P. A., & Weese, S. E. (1993). Predicting intelligence with multiple speed of information processing tests. *Personality and Individual Differences*, *14*, 413–419.

Wothke, W., Bock, R., Curran, L., Fairbank, B., Augustin, J., Gillet, A., & Guerrero, C. (1991). *Factor analytic examination of the armed services vocational aptitude battery (ASVAB) and the KIT of factor-referenced tests.* (Technical Report) Texas: Air Force Human Resources Laboratory.

Yung, Y., Thissen, D., & McCleod, L. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, *64*(2), 113–128.