Available online at www.sciencedirect.com

SCIENCE DIRECT®

Intelligence 33 (2005) 623–642

INTELLIGENCE

Memory span and general intelligence: A latent-variable approach

Roberto Colom *, Francisco J. Abad, Irene Rebollo, Pei Chun Shih

Facultad de Psicología, Universidad Autónoma de Madrid, 28049 Madrid, Spain

Received 6 October 2004; received in revised form 5 March 2005; accepted 15 May 2005
Available online 14 July 2005

Abstract

There are several studies showing that working memory and intelligence are strongly related. However, working memory tasks require simultaneous processing and storage, so the causes of their relationship with intelligence are currently a matter of discussion. The present study examined the simultaneous relationships among short-term memory (STM), working memory (WM), and general intelligence (g). Two hundred and eight participants performed six verbal, quantitative, and spatial STM tasks, six verbal, quantitative, and spatial WM tasks, and eight tests measuring fluid, crystallized, spatial, and quantitative intelligence. Especial care is taken to avoid misrepresenting the relations among the constructs being studied because of specific task variance. Structural equation modelling (SEM) results revealed that (a) WM and g are (almost) isomorphic constructs, (b) the isomorphism vanishes when the storage component of WM is partialed out, and (c) STM *and* WM (with its storage component partialed out) predict g.

Keywords: Memory span tasks; Short-term memory; Working memory; General intelligence; Structural equation modelling

---

Working memory (WM) tasks are strongly related to g (Ackerman, Beier, & Boyle, 2002; Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Colom & Shih, 2004; Süβ, Oberauer, Wittman, Wilhelm, & Schulze, 2002), reasoning ability (Kyllonen & Christal, 1990), fluid intelligence (Colom, Flores-Mendoza, & Rebollo, 2003; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004), spatial ability (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001), and reading comprehension (Daneman & Merikle, 1996).

---

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

0160-2896/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2005.05.006

Indeed, several studies suggest that WM and intelligence are indistinguishable (isomorphic) constructs. Kyllonen and Christal (1990) found structural coefficients of .80 through .88 between WM and reasoning ability. Colom et al. (2003) found a correlation of .70 between a composite measure of WM and measures of fluid intelligence. Ackerman et al. (2002) found a structural coefficient of .70 between WM and *g*. In three separate studies, Colom et al. (2004) found a mean structural coefficient of .96 between general intelligence (*g*) and WM. Finally, Colom and Shih (2004) reported a structural coefficient of .86 between *g* and WM.

However, Ackerman, Beier, and Boyle (2005) conducted a meta-analysis examining the relationship between WM and *g*, as well as between STM and *g*. This study was based on a literature search ranging between 1872 and 2002. The meta-analytically derived correlation between cognitive ability and WM was .36, whereas the meta-analytically derived correlation between STM and cognitive ability was .28. Further, after a SEM analysis, the correlation found between STM and *g* was .49, whereas the correlation between WM and *g* was .50, which suggested that STM and WM were equally related to *g*.

Nevertheless, there are some studies claiming that WM is a much better predictor than STM. These studies consider the simultaneous estimation of relationships between the three constructs of interest, as opposed to examining WM and intelligence, or STM and intelligence, in separate analyses. Engle et al. (1999) and Conway, Cowan, Bunting, Therriault and Minkoff (2002) reported that when those relations are estimated simultaneously, the correlation between WM and intelligence is large and significant, whereas the correlation between STM and intelligence is negligible.

However, Ackerman et al. (2005) indicate that these studies are relatively limited in their assessment of the constructs of interest. For example, only two tests were used as indicators of intelligence. Furthermore, the indicators for WM and STM were also limited (Beier & Ackerman, 2004). This would be seen with some reservations, because the relations among the constructs being studied could be misrepresented. A better approach for sampling the construct space would be to include heterogeneous tasks to control for the effect of unwanted variance. Although the results of these previous studies are suggestive, they may not constitute the best evidence for examining the relations among the constructs of interest (Beier & Ackerman, 2004).

Recently, Kane et al. (2004) take a latent variable approach that resembles the study to be reported in the present article. Several measures of verbal and visuo-spatial WM and STM span were employed, as well as several diverse cognitive ability measures. Four main results can be highlighted. First, the correlation among WM and STM latent factors across content domains ranged from .63 to .89. Although those researchers did not report the results of a model where STM and WM were represented as two correlated higher order factors, we did this analysis after their correlation matrix and the resulting correlation was almost perfect ($r = .99$). Second, STM was found more domain-specific than WM. The correlation between STM-Verbal and STM-Spatial was .63, whereas the correlation between WM-Verbal and WM-Spatial was .83. Testing the structure of WM, those researchers found that a single factor model did not provide a good fit to the data, whereas a two-factor model distinguishing WM-V and WM-S did. Nevertheless, they treated WM as a unitary latent factor, while they preferred to treat the STM construct distinguishing verbal and spatial short-term storage. Third, the general structural model relating STM, WM and reasoning suffered from the well-known multicollinearity problem. Nevertheless, the primary interest was to test for the relation between what is shared among WM tasks and what is shared among reasoning tasks. WM span tasks were thought to be multiple determined by both domain-general executive attention processes and domain-specific coding and storage (STM) processes. Therefore, Kane et al. addressed the relative contribution of WM, verbal STM, and visuo-spatial STM processes to the relation between memory span and reasoning. The factor model consisted of an executive attention factor (WM), with loadings from all memory variables, reflecting the domain-general, executive variance shared by all the WM and STM span tasks. The model also consisted of domain-specific factors, with loadings from the verbal and spatial tasks on the storage-V and storage-S factors, respectively. Kane et al. presumed that the common variance among span tasks reflects executive rather than storage processes. This is surprising, because in two key previous studies, Engle et al. (1999) and Conway et al. (2002) proposed that the common variance between WM and STM reflect primarily storage and the residual WM variance reflects primarily executive control processes. Furthermore, it seems risky to assume that a latent factor clearly mixing storage-plus-processing can be seen *primarily* as a clear-cut representation of the central executive (see below). Finally, Kane et al. found that fluid intelligence was predicted by the executive attention factor (.52) and by the STM spatial factor (.54). Especially noteworthy is that the contribution of the STM-S latent factor was *independent* of the contribution of the executive attention factor, a fact that questions the likelihood of the view that executive attention processes drive primarily the predictive utility of memory span measures (Conway et al., 2002; Engle et al., 1999).

The purpose of the present study is to examine the simultaneous relationship among STM, WM, and *g* within a SEM framework. Fig. 1 depicts the general model considered in the current study. This model indicates that *g* is predicted by WM and STM. Those 3 constructs of interest are defined as broadly sampled latent factors. Note that STM, WM, and *g* are higher order factors, which means that they are intended to extract the *common true variance* underlying all the memory span measures, as well as the cognitive ability measures.

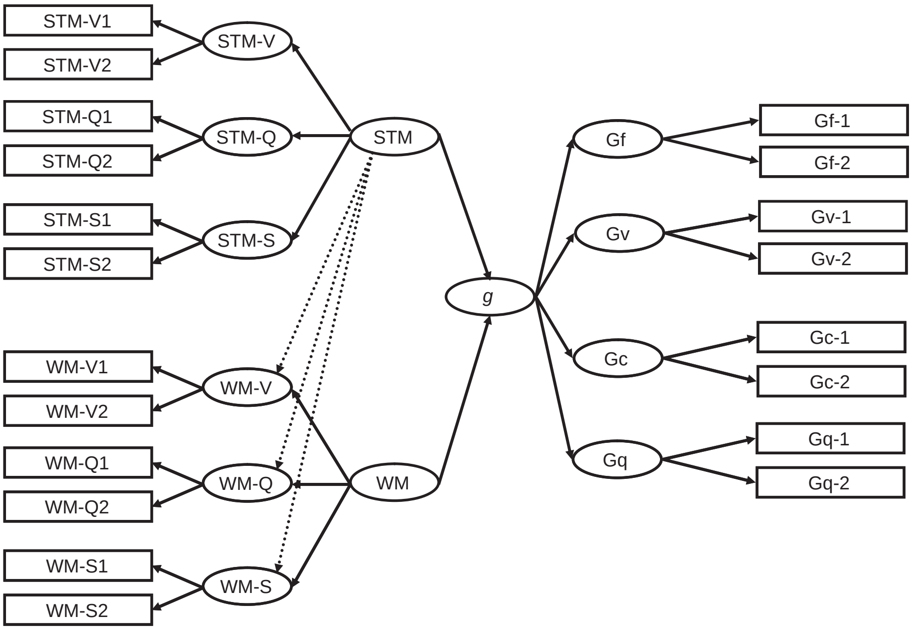

Fig. 1. General Structural Hierarchical Model representing the 3-way relationships among STM, WM, and general intelligence (*g*).

The meta-analysis of Ackerman et al. (2005) carefully considers the extreme importance of the choice of tasks. Those researchers wrote: "resolution of the question of how and how much WM and intelligence are related ultimately requires additional research. In our opinion, the issue cannot be ultimately settled until studies are conducted that provide multiple tests of a wide range of ability factors (reasoning, spatial, verbal, numerical, perceptual speed), multiple tests of WM in each of the different content domains (verbal, numerical, spatial), (and) (. . .) WM tests that do not depend on time sharing performance".

Furthermore, and especially germane for the present study, they consider crucial the so-called *principle of aggregation* which indicates that multiple measures of the same factor will yield a more stable and representative estimator than single behaviour measures. *Aggregating* across multiple measures will provide a more robust assessment of WM/STM/*g* relations, and might result in a more informative association between the measures.

These considerations conducted us to *aggregate* specific measures into primary latent factors and those factors into secondary factors, following the widely accepted hierarchical model by Carroll (1993) of cognitive abilities. This hierarchical model was applied to the three constructs of interest: STM was defined by three primary latent factors for verbal, numerical, and spatial contents, WM was defined by three primary latent factors for verbal, numerical, and spatial contents, and general intelligence (g) was defined by four primary factors for fluid, spatial, crystallized, and quantitative intelligence (Fig. 1). Moreover, the specific measures were as diverse as possible (see below). Further, we follow the meta-analytic results provided by Ackerman et al. (2005) stating that *content differentiation was not especially germane neither for STM or WM*, at least from an individual-differences perspective.

The proposed hierarchical organization of STM, WM, and *g* does not implicate that the specific primary factors are unimportant. The higher order factors representing STM, WM, and *g* capture (indirectly) what is shared by all the corresponding measures, but the first-order factors representing STM-V, STM-Q, STM-S, WM-V, WM-Q, WM-S, Gf, Gv, Gc, and Gq accumulate unique variance unpredicted by the corresponding higher order factors. Therefore, it is important to notice that our general hierarchical model *does not* assume that WM, STM, or *g* are (or are not) unitary constructs. It is assumed that the first-order latent factors share something in common, but that they have a given amount of uniqueness also. From that point of view, our general hierarchical model is less restrictive than those models presuming the unitary (or non-unitary) nature of STM or WM (see, for instance, Kane et al., 2004).

In order to avoid the recognised multicollinearity problem that characterizes WM and STM measures (Engle et al., 1999; Kane et al., 2004; Miyake et al., 2001) we took the following approach (see Conway et al., 2002, p. 176): WM tasks place great demands on STM, but (theoretically) WM tasks also require the *simultaneous* processing and storage of information. Therefore, in order to compare the relative contribution of STM and WM to *g*, we specified all the memory span tasks as indicators of STM, but only WM measures as indicators of WM.

Kane et al. (2004) adopted an alternative approach to that employed by Engle et al. (1999) and Conway et al. (2002). Their general model consisted of an executive attention factor, with loadings from all their memory variables, presumably reflecting the domain-general executive variance shared by all the span tasks (as previously noted). Instead, Engle et al. (1999) and Conway et al. (2002) proposed that the common variance between WM and STM reflect primarily storage, whereas the residual WM variance reflects primarily executive control processes. It is hard to admit that the general factor of Kane et al. (2004) represents appropriately *mainly* the presumed general control processes underlying all their memory span tasks. Our skepticism relies in the fact that their STM tasks were carefully designed to measure short-term storage: “we choose clarity of interpretation over breadth of measurement, and modelled all the verbal and spatial WM tasks after reading span (...) performance in these tasks, with either verbal or visuo-spatial materials, reflects one’s ability to encode, maintain, and retrieve lists of isolated stimuli in the face of a regularly occurring, highly interfering distracter task (...) Yet an additional benefit of the span procedure is that we could create STM versions of each of our WM tasks that presented the same to-be-remembered stimuli, *but without the additional processing demand of the secondary task* (emphasis added)”. Furthermore, their test of the structure of WM did not support its unitary nature (see above). Although the correlation between WM-Verbal and WM-Spatial was quite high (.83), the one factor model did not provide a good fit to the data. Nevertheless, they treated WM as a single latent factor predicting intelligence. However, they considered unlikely that the shared variance among STM-Verbal and STM-Spatial tasks would reflect common storage mechanisms. But there are two sources of evidence that are not consistent with that consideration. Firstly, the correlation between STM-V and STM-S was quite high (.63). Secondly, the inspection of their correlation matrix reveals that the average raw correlation among the STM measures is .52, whereas the average raw correlation among the WM measures is .60. From an individual-differences perspective, those average raw correlations do not suggest that the shared variance among WM measures, but not among STM measures, reflects common mechanisms. Those correlations are consistent with the view that both STM and WM share common mechanisms (and both constructs have unique mechanisms also). The hierarchical model depicted in Fig. 1 takes into consideration this empirical fact. We must emphasize that the predictive power of STM and WM should be addressed explicitly; if WM is more predictive than STM, then the findings will tell just that.

1. Method

1.1. Participants

208 Psychology undergraduates participated in the study to fulfill a course requirement. Their mean age was 20.73 (SD = 3.7).

1.2. Administered memory span tasks

Verbal STM was measured by the forward and backward letter span tasks, while quantitative STM was measured by the forward and backward digit span tasks. The selection of those verbal and quantitative STM tasks was made following the study by Engle et al. (1999): “tasks thought to be good STM tasks were simple word span with dissimilar words, simple word span with similar words, and backward word span with dissimilar words” (p. 314). In order to make the tasks *as simple as possible*, we have used single letters and single digits, instead of words or complex digits. It is imperative to keep in mind that, according to Rosen and Engle (1997), backward span tasks fit with other tasks of STM, because *a simple transposition of order* does not change their short-term storage nature (Cantor, Engle, & Hamilton, 1991). Spatial STM was measured by the Corsi block and the dot memory tasks, both requiring the maintenance of spatial information (location sequences and dot configurations, respectively), without involving any explicit concurrent processing. The spatial STM tasks employed here modelled those administered by Miyake et al. (2001).

Importantly, not all WM tasks were modelled after the reading span task, because we wanted to define a higher order WM factor going beyond time sharing performance (Ackerman et al., 2005). Verbal WM was measured by the ABCD and the Alphabet tasks, and they were modelled after the CAM Battery (Kyllonen & Christal, 1990). The ABCD was employed by Ackerman et al. (2002) as well as by Engle et al. (1999) as a measure of verbal WM. Quantitative WM was measured by the Mental Counters and Computation Span tasks. The mental counters task was modelled after Larson and Saccuzzo (1989) and it was employed by Mackintosh and Bennett (2003) as a measure of quantitative WM. The computation span task was modelled after Ackerman et al. (2002). Finally, spatial WM was measured by the dot matrix and the letter rotation tasks. Both tasks were modelled after Miyake et al. (2001) study and they require visuo-spatial storage (dot locations and spatial orientations, respectively) with a concurrent visuo-spatial processing (verification of spatial matrix equations or mental rotation).

1.3. Administered intelligence tests

There are two basic criteria for obtaining a good representation of the construct of general intelligence (Ackerman et al., 2002; Carroll, 1993; Jensen & Weng, 1994): number and variety of cognitive ability tests. The tests in the present study' battery call for more than three primary factors (variety) which is clearly desirable to obtain a reliable higher order factor representing *g*. Ackerman et al. (2005) wrote: "high-quality estimates of *g* are generated from the average across multiple tests of differing formats, contents, and processes".

The primary cognitive ability factors were fluid intelligence (Gf), spatial ability (Gv), crystallized intelligence (Gc), and quantitative ability (Gq). Gf was measured by the reasoning subtest from the Primary Mental Abilities Battery (PMA-R; Thurstone, 1938) and the Culture Fair Intelligence Test (subtests 1 and 3; TEA, 1997). Gv was measured by the mental rotation test from the PMA (S; Thurstone, 1938) and the surface development test (Yela, 1969). Gc was measured by the vocabulary subtest from the PMA (V; Thurstone, 1938) and the verbal reasoning subtest from the Differential Abilities Battery (DAT-VR; Bennett, Seashore, & Wesman, 1974). Finally, Gq was measured by the computation subtest from the PMA (N; Thurstone, 1938) and a test of quantitative reasoning called "Monedas" (Coins).

A more detailed tasks and tests description can be seen at the Appendix.

1.4. Procedure

Testing took place in four sessions, administered collectively (in groups of no more than 20 participants) for a total of four hours approximately. The order of task administrations was fixed for all the participants to minimize any measurement error due to an order by participant interaction. The memory span tasks administered in session 1 were FLSPAN, Mental Counters, Corsi Block, ABCD, FDSPAN, and Letter Rotation. The memory span tasks administered in session 2 were BLSPAN, Computation Span, Dot Memory, Alphabet, BDSPAN, and Dot Matrix. The intelligence tests administered in session 3 were: PMA-R, Culture Fair Intelligence Test, PMA-S, and Surface Development. The intelligence tests administered in session 4 were: PMA-V, DAT-VR, PMA-N, and Monedas.

2. Results

The descriptive statistics are shown in Table 1. Tasks' correlations and reliability indices (Cronbach's alpha) are also presented in Table 1. The measures meet standard criteria for uni-variate normality with skew values less than 3 and kurtosis values less than 4 (Kline, 1998). Reliability indices show appropriate values. CFA and SEM analyses were conducted using LISREL 8.54 (Jöreskog & Söborn, 2001).

2.1. Measurement models

The first step was to establish the *measurement models* by defining latent factors separately for both memory span and general intelligence.

STM and WM were postulated as correlated higher order latent factors. STM was considered as a second-order latent factor defined by 3 primary latent factors for verbal, quantitative, and spatial STM tasks. WM was also considered as a second-order latent factor defined by the corresponding 3 primary latent factors for verbal, quantitative, and spatial WM tasks.[1] The fit of this model was appropriate: $\chi^2_{(46)}=87.28$, CMIN/DF = 1.89, TLI = .97, RMSEA = .066.

Fig. 2 displays the standardized regression weights for this model. It is noteworthy both the high regression weights from the higher order factors to the first-order factors, as well as the large correlation between STM and WM (.89). This correlation implies that 80% of the variance is shared by those higher order factors that presumably are an appropriate statistical representation of the constructs of interest. Nevertheless, 20% of the variance is non-shared between those higher order latent factors, a fact that opens the door to answer the main question of the present study.

With respect to general intelligence (*g*), fluid (Gf), spatial (Gv), crystallized (Gc), and quantitative (Gq) cognitive abilities were postulated as primary latent factors. The higher order factor representing general intelligence (*g*) predicted those primary factors.

Fig. 3 depicts the standardized regression weights for this model. The fit of this model was appropriate: $\chi^2_{(16)}=33.48$, CMIN/DF = 2.09, TLI = .96, RMSEA = .073. This model shows that Gv and Gf are the best predicted first-order factors. Nevertheless, the regression weights from *g* to the first-order factors are all quite large.

2.2. Structural models

The second step was to establish *structural models* for the memory span and intelligence higher order latent factors. The regression weights obtained from the already tested measurement models were considered from now on.

We proceed first to compute the regression weights from STM to *g*, as well as from WM to *g*, but in separate analyses. The estimates from the measurement models are copied and fixed in order to preserve both the original independent structure and the nature of the latent variables intact. Only the structural paths from STM to *g* and from WM to *g* are estimated. Nevertheless, to evaluate the goodness of fit of

---

[1] The correlated error variances between BLSPAN and BDSPAN were intended to partial out their shared variance derived from the transposition of order that distinguishes them from FLSPAN and FDSPAN, thus strengthening the short-term storage nature of the STM latent factor.

Table 1
Correlation matrix, descriptive statistics, and reliability indices

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. FLSPAN		.577	.537	.505	.382	.244	.262	.411	.227	.347	.267	.347	.315	.204	.231	.224	.290	.271	.194	.347
2. BLSPAN			.473	.601	.301	.169	.234	.455	.163	.435	.296	.254	.260	.123	.292	.209	.156	.240	.243	.345
3. FDSPAN				.631	.334	.266	.235	.373	.258	.486	.225	.322	.317	.115	.145	.179	.146	.193	.237	.286
4. BDSPAN					.345	.243	.226	.440	.222	.454	.297	.394	.293	.164	.197	.187	.213	.235	.203	.298
5. Corsi						.454	.310	.276	.350	.272	.262	.377	.297	.169	.259	.230	.065	.138	.189	.257
6. Dot Memory							.169	.173	.343	.268	.234	.408	.283	.205	.307	.336	.218	.199	.119	.152
7. ABCD								.316	.210	.184	.279	.205	.348	.280	.147	.247	.140	.279	.173	.381
8. Alphabet									.222	.437	.324	.402	.379	.180	.264	.214	.288	.342	.190	.431
9. Counters										.335	.236	.371	.368	.297	.400	.317	.224	.316	.102	.406
10. Computation											.298	.411	.395	.259	.301	.320	.217	.257	.347	.516
11. Letter Rotation												.371	.318	.256	.322	.293	.233	.376	.178	.401
12. Dot Matrix													.394	.270	.349	.377	.170	.275	.146	.336
13. PMA-R														.499	.426	.513	.406	.447	.349	.506
14. Cattell															.303	.419	.281	.273	.237	.390
15. PMA-S																.429	.397	.357	.231	.406
16. Surface																	.211	.395	.105	.426
17. PMA-V																		.573	.175	.359
18. DAT-VR																			.184	.457
19. PMA-N																				.489
20. Monedas																				
Mean	9.5	9.2	13.5	12.3	10.2	12.1	8.3	7.7	9.5	17.5	46.3	76.5	21.1 (20.4)	13.5	27.2 (26.3)	24.6 (28.1)	32.6 (34.03)	26.9 (26.9)	17.7 (22.9)	22.3 (22.9)
SD	2.4	3.2	2.7	3.5	2.5	2.6	4.4	3.4	3.8	5.1	8.8	8.6	4.4 (4.03)	2.1	11.7 (12.1)	12.9 (13.2)	6.7 (7.8)	7.2 (7.0)	7.6 (7.8)	6.6 (7.2)
Skew	.42	.23	.14	−.01	.26	.10	−.35	.40	−.28	−.20	−.07	−.19	−.33	−.47	.24	.64	.05	.16	.26	−.26
Kurt	−.24	−.13	−.35	−.34	.02	−.59	−1.3	−.10	−.53	−.42	.34	1.4	.00	.50	−.58	−.12	−.43	−.55	−.08	−.62
Reliability	.75	.87	.76	.83	.80	.70	.87	.76	.80	.83	.77	.79	.80	.83	.73	.93	.85	.82	.84	.89

Available population descriptive statistics (Mean and SD) for the intelligence tests are shown in parenthesis.

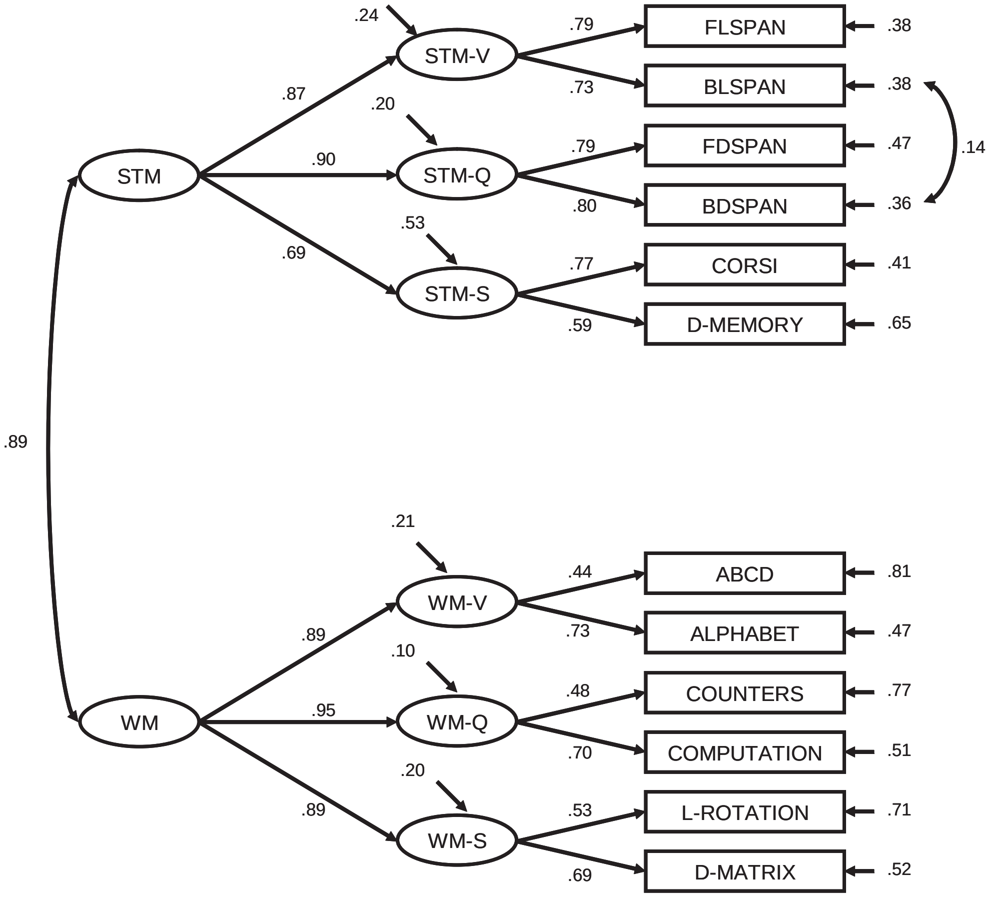

Fig. 2. CFA for STM and WM defined as higher order factors.

the model, the degrees of freedom are adjusted as if all the copied parameters had been estimated. The results inform about the *isolated* predictive power of the memory span constructs.

Fig. 4 shows that the structural coefficient from STM to *g* was high (.58). The fit for this model was excellent: $\chi^2_{(68)}=100.49$, CMIN/DF = 1.48, TLI = .97, RMSEA = .048. This finding indicates that the temporary storage component captured by the STM latent factor behaves as a good predictor of *g*.

Fig. 5 shows that the structural coefficient from WM to *g* was much higher (.89). The fit for this model was reasonable: $\chi^2_{(69)}=126.51$, CMIN/DF = 1.83, TLI = .96, RMSEA = .063. This finding indicates that the WM latent factor representing the switching between the representation of the items and the processing component constitutes a better predictor than the latent factor requiring little switch of attention from the representation of the items (STM). Additionally, this result is consistent with the view that *g* and WM are (almost) isomorphic constructs (Colom et al., 2004; Colom & Shih, 2004; Engle, 2002; Kyllonen & Christal, 1990).

Nevertheless, the WM latent factor mixes the contribution of short-term storage and other presumed general control processes operating across the WM tasks. Given that the WM latent factor comprises processing and storage requirements, whereas the STM latent factor comprises *primarily* (although not

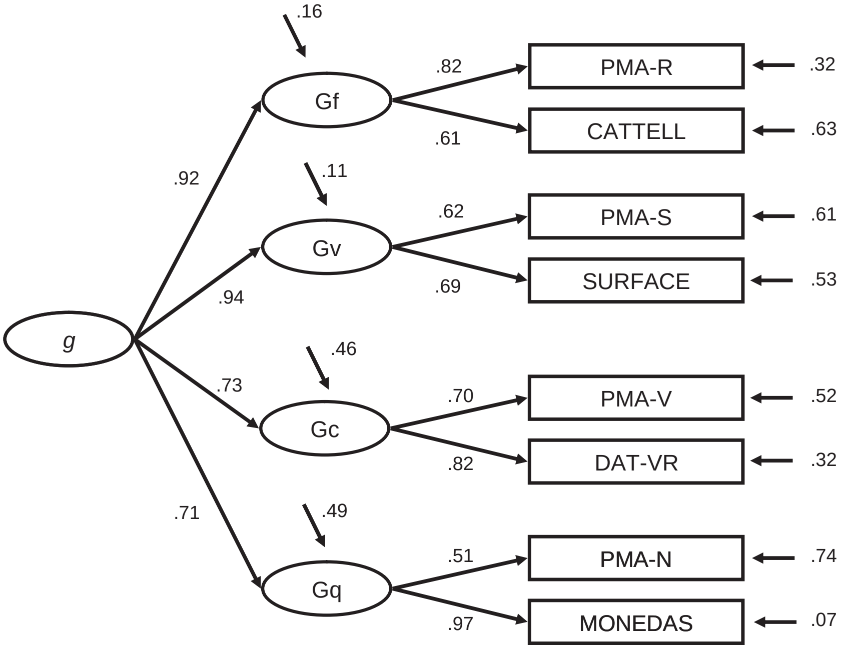

Fig. 3. CFA for general intelligence (*g*).

exclusively) storage requirements, the final phase consisted in a structural hierarchical model in which STM and WM predicted *g*. Taking into account the observed high correlation between STM and WM, we specified all the memory span tasks as indicators of STM, but only WM measures as indicators of WM, in order to compare their relative joint contribution to *g* (Conway et al., 2002).

It is imperative to keep in mind several cautions, mainly because Kane et al. (2004) expressed some reservations about the logic underlying this general model. First, *theoretically*, the resulting STM latent

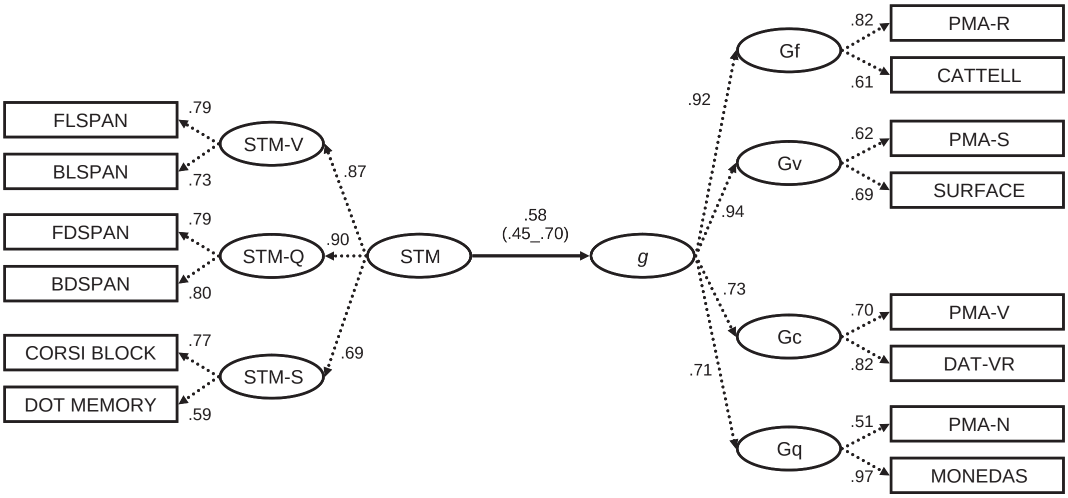

Fig. 4. SEM for 2-way relationship between STM and *g*.

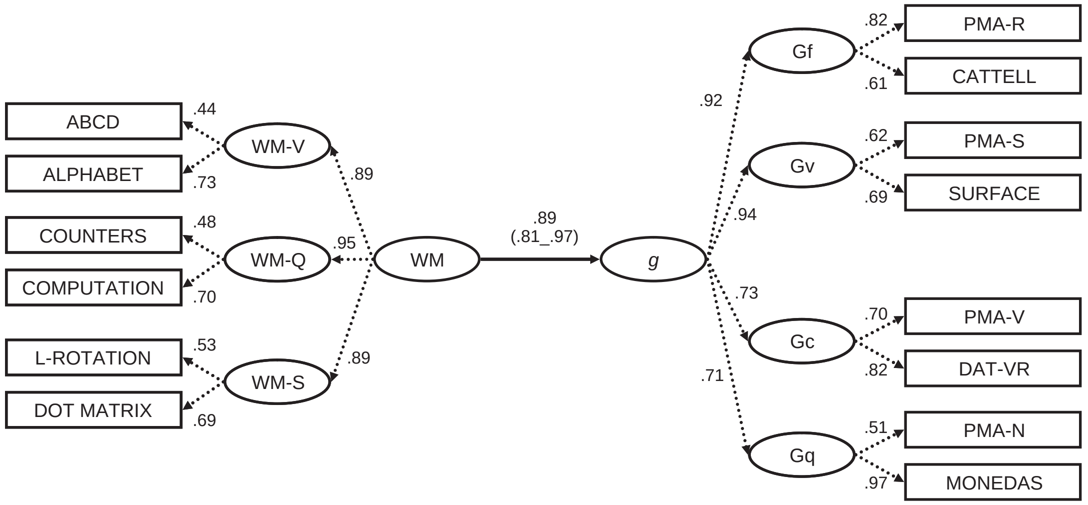

Fig. 5. SEM for 2-way relationship between WM and *g*.

factor comprises the shared storage component among all the memory span tasks, whereas the resulting WM latent factor comprises the specific component of the WM tasks with their storage component partialed out. Second, in this general hierarchical model, all parameters from the measurement model for *g*, for the STM and WM first-order factors, as well as for the second-order STM factor were *fixed*. Third, given that the variance of the first-order WM factors was partitioned into that due to storage and that due to other (unknown) processes, the nature of the second-order WM factor would change and thus its path coefficients were estimated. Fourth, the structural coefficients from STM and WM to *g* were also freely estimated in this general model. Fifth, Kane et al. (2004) assume that when a latent factor is defined by STM *and* WM measures, the resulting factor extracts non-storage components from the latter measures. However, this assumption will be supported if, and only if, a significant change is observed on the structural coefficient linking STM to *g*. Finally, the degrees of freedom were corrected as if all parameters had been estimated to avoid artificial increase of the fit.

Fig. 6 displays the standardized regression weights for the general structural model. The fit of this model was quite good: $\chi^2_{(154)}=260.59$, CMIN/DF = 1.69, TLI = .97, RMSEA = .058. The structural model depicted in Fig. 6 shows that both STM and WM predict *g*. The structural coefficient for STM is .58, for WM is .79, and the residual of the *g* factor was .12 (confidence interval $-.07$ to .32). Nevertheless, we tested a new model in order to reach a more clear-cut conclusion regarding the 3-way relationships among *g*, STM, and WM. This model constrained the WM and STM structural coefficients on *g* to be equal. Interestingly, this constrain lead to a non-significant change of fit ($\Delta\chi^2_{(1)}=1.7$; $p=.19$; the resulting regression weight for both STM and WM was .66). Thus, both latent constructs should be considered *equally* germane to predict (and presumably to understand) individual-differences in *g*.

The models depicted in Figs. 4–6 comprise three noteworthy points of interest. First, the structural coefficient from STM to *g* does not change from Figs. 4–6. The value of the corresponding structural coefficient (.58) appears to be robust across models. Moreover, the confidence interval of that estimated value does not change in those two models (.45 to .70). This is partly a result of the fixation of the STM loadings, indicating that adding the WM loadings on this latent factor did not change its nature. Further, this fact reinforces its interpretation as a storage component and is not consistent with the assumption by

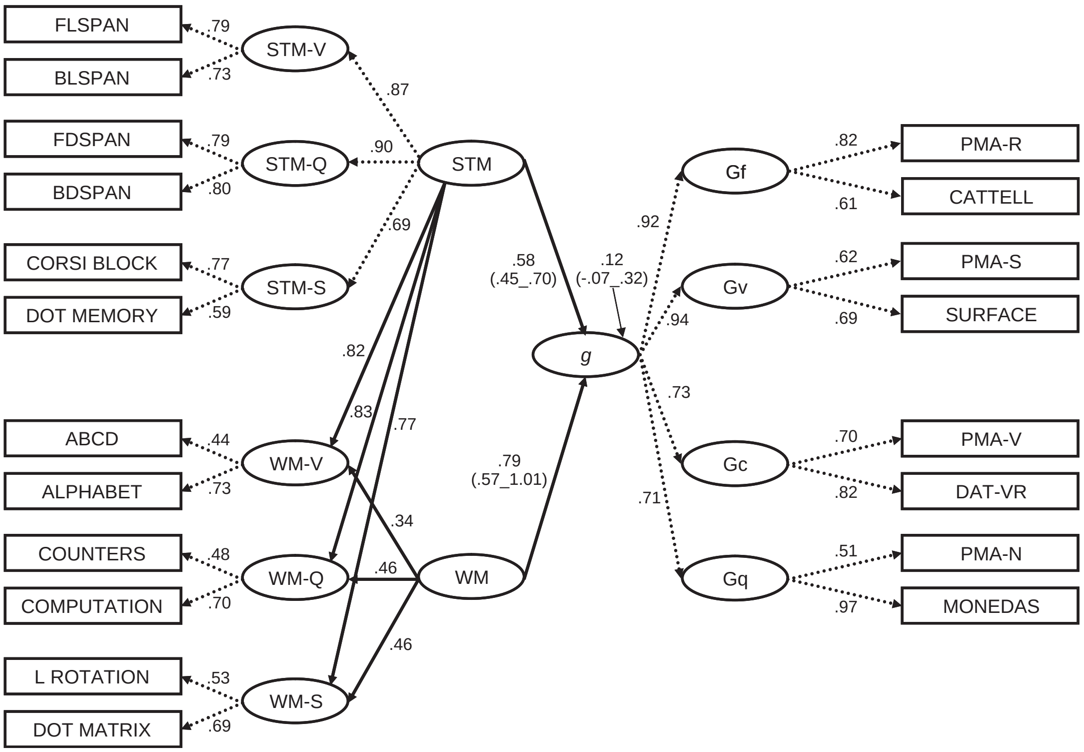

Fig. 6. SEM for the 3-way relationships among STM, WM, and *g*.

Kane et al. (2004) (see above). Second, the structural coefficient from WM to *g* does change from Figs. 5 and 6, from a value of .89 to a value of .79. That difference suggests that the storage component of the WM latent factor is necessary to account for the isomorphism between WM and *g*. That isomorphism is very robust, given that the confidence interval of the corresponding structural coefficient is narrow (from .81 to .97; see Fig. 5). However, the predictive power of the WM latent factor, once its storage component is partialed out, is much less compelling: the confidence interval for that structural coefficient is much more unstable, ranging from .57 to 1.01. This result suggests that the non-storage component of the WM latent factor predicts *g* beyond its storage component, but with a greater instability. Further, taking the value of .66 from the model constraining STM and WM to be equal, there is an 80% of straight shared variance between WM and *g*, whereas there is a 43% of shared variance between *g* and WM (with its storage component partialed out). Finally, the relevance of the storage component of the WM latent factor can be highlighted by the large values of the estimated structural coefficients observed between the STM higher order factor and the first-order WM factors, ranging from .77 to .83. Interestingly, the coefficients from the higher order WM factor to its first-order factors range from .89 to .95 in the model depicted in Fig. 5, but those values range from .34 to .46 in the general hierarchical model depicted in Fig. 6 (once the storage component is partialed out).[2]

Finally, in order to cross-validate the results of the SEM analyses, we computed a regression analysis in which a WM composite score was predicted by a STM composite score. WM variance unpredicted by

---

[2] We have fitted again the models depicted in Figs. 4–6 without fixing any value derived from the measurement models, and the results were almost identical.

STM was taken to compute a residual WM score. Then, STM scores and WM residual scores were correlated with a *g* composite score. The resulting correlations were .47 and .48, respectively. Furthermore, a new STM composite score was computed excluding the two backward measures (BLSPAN and BDSPAN). The new STM composite was employed to predict the WM composite score. Again, WM variance unpredicted by the new STM score was taken to compute a residual WM score. The new STM score and the WM residual score were correlated with *g*. The correlations were .46 and .49, respectively. Those values were virtually the same as those obtained considering the six STM measures, which supports the statement that backward measures do not change the nature of the STM construct considered in the present article.

3. General discussion

The findings show that WM and *g* are (almost) isomorphic constructs, but this is only supported when the WM latent factor comprises storage-plus-processing requirements. Once the storage component of the WM system is partialed out, the relationship between WM and *g* reveals much more unstable. Further, the simultaneous analysis of the relationships among STM, WM, and *g* showed that both STM and WM (with its storage component partialed out) predict *g* with the same power.

Kane et al. (2004) claimed that WM is a much better predictor of intelligence than STM. However, we have re-analyzed their correlation matrix to test hierarchical models equivalent to those considered in the present study. First, a higher order STM factor was defined taking STM-Verbal and STM-Spatial as primary latent factors defined by their corresponding tasks, whereas *g* was represented as a higher order latent factor from the three primary latent factors representing verbal reasoning, spatial visualization, and fluid intelligence. The resulting structural coefficient from the STM higher order factor to *g* was a large .87. Second, a higher order WM factor was defined taking WM-Verbal and WM-Spatial as primary latent factors defined by their corresponding tasks, whereas *g* was defined as in the previous analysis. The structural coefficient from the WM higher order factor to *g* was .72. Third, the 3-way relationship among STM, WM, and *g* was tested. The two sets of reported re-analyses showed that both STM and WM predicted *g*, a result entirely consistent with those derived from the present study. However, the WM latent factor comprises storage and processing requirements, so the storage component must be partialed out to test the specific predictive power that should be attributed to WM. With this goal in mind, all memory span primary latent factors were predicted by the STM higher order latent factor, whereas WM primary latent factor were predicted only by the WM higher order latent factor. This model is equivalent to the general hierarchical model tested in the present study (see Fig. 1). However, the results were disappointing, because a high structural coefficient was found from STM to *g* (.90), but a high negative coefficient was found from WM to *g* (−.61). This result is clearly unreasonable, so any clear-cut conclusion should be extracted from this later analysis.

Those re-analyses are consistent with the findings reported in the present article and reject the assumption that the short-term storage component of the WM system plays a negligible role to understand the strong relationships usually found between WM and *g*. Further, it should be remembered that the STM tasks of Kane et al. (2004) were expressly designed to partial out the processing requirements that characterized their WM tasks. Ironically, the reported re-analyses revealed that the structural coefficient between STM and *g* was greater (.87) than the structural coefficient between WM and *g* (.72). This is hardly a strong base to support their main theoretical perspective, namely, that executive functioning drives primarily the strong relationship between WM and intelligence.

Nevertheless, we endorse the statement by Kane et al. (2004) supporting the existence of a *general ability* to maintain stimulus representations (irrespective of content domain) in an active state. However, while Kane et al. (2004) propose that this general ability must be attributed to a presumed controlled attention ability, we suggest the more parsimonious view that this ability derives from a single capacity coping with several diverse memory challenges (Colom, Rebollo, Abad, & Shih, in press; Colom & Shih, 2004).

In this same vein, Baddeley (2002) initially conceived the central executive "in the vaguest possible terms as a *limited capacity pool of general processing resources*" (p. 89, emphasis added). Further, Cowan (2004) suggests that the processing component of WM tasks prevent covert rehearsal, and, therefore, presumably temporary storage requirements are greater in WM tasks than in STM tasks. The difference between those memory span tasks is that the former comprises additional processing requirements fighting against the reliable temporary maintenance of the relevant information. Therefore, "there is no reason to remain obsessed with storage-plus-processing tasks as a means to measure WM capacity. Measures designed to estimate *how much information* can be brought into the focus of attention at once are conceptually simpler" (Cowan, 2004).

What about the predictive power of the latent factor unrelated to the short-term storage component of the WM system? We found that its structural coefficient was equivalent to that of STM. Engle et al. (1999) and Conway et al. (2002) claim that the WM latent factor (with its storage component partialed out) represents appropriately a controlled attention ability closely tied to executive functioning. However, the extensive analysis by Miyake, Friedman, Emerson, Witzki and Howerter (2000) does not support that view.

Miyake et al. (2000) studied three recognised key executive functions: *shifting* between tasks or mental sets, *inhibition* of dominant or prepotent responses, and *updating* and monitoring of working memory representation. Each function was measured by several tasks. The corresponding latent factors were correlated with several contrast measures: the Wisconsin Card Sorting Test, Tower of Hanoi, random number generation, operation span, and dual tasking. Among their results, four are especially relevant for the current discussion. First, the operation span task showed low correlations with executive measures. The correlations between this task and the most reliable measures of executive functioning were: .13 with the number–letter task (shifting), .28 with the letter monitoring task (updating), and .13 with the stop signal task (inhibition). The operation span task is employed as a key measure of WM capacity by Engle et al. (1999), Conway et al. (2002), and Kane et al. (2004). Second, Miyake et al. (2000) considered a dual task in which participants performed maze tracing and word generation tasks simultaneously. Performance in this dual task correlated $-.14$ with the operation span task. Third, the random number generation task was found to draw on multiple executive functions, requiring the inhibition ability to suppress stereotyped responses, as well as the updating ability to monitor response distribution. The operation span task showed a low correlation with the random number generation task (.17). Finally, Miyake et al. (2000) did not find evidence for the proposal that the ability to efficiently switch back and forth between the processing component (equation verification) and the storage component (word span) is a crucial aspect of the operation span task.

Therefore, the study by Miyake et al. (2000) shows that one key measure of WM capacity employed by Engle et al. (1999), Conway et al. (2002), and Kane et al. (2004) is not informatively related to executive functioning. Therefore, it seems risky to assume that if the short-term memory component of the operation span task is partialed out, the remaining variance reflects mainly executive functioning. It is much more conservative and data driven to assume that the nature of this remaining variance is currently unknown.

Kane and Engle (2002) state that there are too many strong correlations among diverse memory span tasks and diverse higher order tasks (*g* demanding) to deny that some *general mechanism* is involved. We strongly endorse this statement. The ability to maintain a reliable memory representation in an active state could be crucial to general success across complex cognition. The fact that increases in *complexity* on memory span tasks and on intelligence tests recruit more activation from the dorsolateral prefrontal cortex (Duncan et al., 2000; Kane & Engle, 2002) is consistent with the view of the single capacity: memory span and intelligence do not need to be associated by the intervention of one presumed controlled attention ability (*the third factor!?*). Instead, the causal link could be provided by both memory span and intelligence exhausting the same single capacity.

The results observed in the present study are largely consistent with those reported by Embretson (1995). She classified several problems that parallel the Raven Progressive Matrices Test (and that served as a proxy measure of fluid intelligence) on their memory load and then examined how individual-differences in performance on the problems were related to both general control processes (WM) and memory load (STM). She argued that the need for working memory capacity (STM) should change with problems variations in memory load, whereas general control processes (WM) would be needed on all problems irrespective of memory load. Those general control processes presumably comprise selecting an effective strategy, monitoring solution processes, or allocating resources to processing. She found that the latent factors representing general control processes and the temporary maintenance of the information accounted for 92% of the variance on the problems. Thus, both factors were found as important for fluid intelligence.

In summary, WM and *g* are (almost) isomorphic constructs, although that isomorphism vanishes when the storage component of WM is partialed out. This suggests that the short-term storage component of the WM system is a crucial underpinning of *g*. Finally, WM components unrelated to short-term storage predict *g* to a significant degree, although the specific nature of these non-storage components is still unknown.

Acknowledgement

The research referred to in this article was supported by a grant funded by the Spanish *Ministerio de Ciencia y Tecnología* (Grant N° BSO2002-01455). We thank Phillip L. Ackerman and Arthur R. Jensen for their helpful comments to a previous version of the present article.

Appendix A

A.1. Forward Letter Span (FLSPAN), Backward Letter Span (BLSPAN), Forward Digit Span (FDSPAN), and Backward Digit Span (BDSPAN)

Single letters or digits (from 1 to 9) were presented on the computer screen at the rate of one letter or digit per second and were randomly grouped to form trials. Unlimited time was allowed to type in *direct* *or reverse order* the letters or digits presented. Set size ranged from three to nine (7 levels × 3 trials each = 21 trials total). The score was the number of accurately reproduced trials.

A.2. Corsi Block

Nine boxes are shown on the computer screen and one box at a time turned orange for 650 ms each. Three different configurations of boxes that changed on each trial were used. Immediately after a sequence of taps, their order must be repeated by clicking on the boxes with the mouse and there was unlimited time to respond. The sequences increased from 3 to 9 taps (7 levels × 3 trials each = 21 trials total). The score was the number of taps reproduced appropriately.

A.3. Dot Memory

One 5 × 5 grid was displayed for 750 ms at the computer screen. Each grid had between two and seven spaces comprising solid dots. After the grid presentation, the locations that contained dots must be recalled, clicking with the mouse on an empty grid. The trials increased from two to seven dots (6 levels × 3 trials each = 18 trials total). The score was the number of trials correctly reproduced.

A.4. ABCD

Two categories and five words in each category were used. The categories were *Trees* and *Food*. Three study frames were displayed on the computer screen for 3 seconds each. The first frame indicated the order of two members from one category (*Cedar before Oak*), the second frame indicated the order of two members from the other category (*Garlic not Before Salt*) and the third frame indicated the order of the categories (*Trees not Before Food*). After the third study screen, an eight choice answer screen was presented in order to select the correct order of the words obtained from the information given by the displayed sentences. A maximum of 10 s to select a response were allowed. The use and ordering of category members were balanced across items, as well as the variations of order (before, not before). There were 14 trials. The score was the number of hits.

A.5. Alphabet

This task was based on successor and predecessor operations applied to a string with a given number of letters. If the first screen presents the letters S, C, P and the second screen displays the operation +1, then the correct response would be T, D, Q. The string of letters is presented for 3 s, the operation to apply is presented for 1.5 s, and there is unlimited time to enter a response. The trials increased from three to seven letters (5 levels × 4 trials each = 20 trials total). For two trials within a given block 1 or 2 positions must be added, while for the other two trials 1 or 2 positions must be subtracted. The number of additions and subtractions are randomized within a given block of trials. The score was the number of correct trials.

A.6. Mental Counters

3 boxes representing counters appeared on the computer screen. At the beginning of each trial, the value of the three counters was set to 0, 0, 0. A yellow star appeared above or below one counter for

500 ms. If the star appeared above the box, one must be added (+1) to that counter, but if the star appeared below the box, one must be subtracted (−1) to that counter. The task requires keeping a running track of the value of the three counters. At the end of each trial, the participant must report the cumulative total of all three counters. There was unlimited time to enter a response. There were 10 trials with five counter changes and 10 trials with seven counter changes. The score was the number of correct trials.

A.7. Computation Span

6 s were allowed to verify the accuracy of several math equations. After the final equation of the trial was displayed, the solutions from the equations must be remembered in their correct serial order irrespective of their accuracy. Each math equation included two operations using digits from 1 to 10. The solutions were always single digit numbers. The trials ranged from three to seven equation/solutions (5 levels × 3 trials each = 15 trials total). The score was the number of hits in the verification and remembering tasks.

A.8. Dot Matrix

A matrix equation must be verified and then a dot location displayed in a 5 × 5 grid must be temporarily retained. The matrix equation requires adding or subtracting simple line drawings and is presented for a maximum of 4.5 s. Once the response is delivered, the computer displayed the grid for 1.5 s. After a given sequence of equation-grid pairs, the grid spaces that contained dots must be recalled clicking with the mouse on an empty grid. The trials increased in size from two to five equations and dots (4 levels × 3 trials = 12 trials total). The score was the number of hits in the verification and remembering tasks.

A.9. Letter Rotation

Several capital letters are presented sequentially normal or mirror imaged, and they can be rotated in one of seven orientations (multiples of 45°). There is a verification task (is the letter normal or mirror imaged?) and a recall task (the orientation of the displayed letters — where was the top of each letter pointing?). The letters are presented for a maximum of 3 s, but no time limit was set to deliver the normal or mirror imaged response. After each set, a grid was depicted to mark the places corresponding to the positions of the tops of the presented letters in their correct serial order. The trials increased in size from two to five letters (4 levels × 3 trials, 12 trials total). The score was the number of hits in the verification and remembering tasks.

A.10. PMA-R

This test comprises 30 letters' series items. The rule (or rules) underlying a given sequence of letters [*a-b-a-b-a-b-a-b*] must be extracted in order to select a given letter from a set of six possible alternatives [*a-b-c-d-e-f*]. Only one alternative is correct. The score was the total number of correct responses.

A.11. Culture Fair Intelligence Test (Scale 3)

This test comprises four subtests, but only subtests 1 and 3 were administered. Subtest 1 is a series test based on a figurative content. Several figures serve as model and the underlying rule (or rules) must be extracted to select a given figure from a set of six possible alternatives. Subtest 3 is a matrices test based on several figures where one figure is missing. The figure filling the matrix must be selected from a set of six possible alternatives. Only one alternative is correct. The score was the total number of correct responses on both subtests.

A.12. PMA-S

This test comprises 20 items. Each item includes a model figure and six alternatives must be evaluated against it. Some alternatives are simply rotated versions of the model figure, whereas the remaining figures are mirror imaged. Only the rotated figures must be selected. Several alternatives could be correct for each item. The score was the total number of correct responses (appropriately selected figures — simply rotated) minus the total number of incorrect responses (inappropriately selected figures — mirror imaged).

A.13. Surface Development

This test comprises 12 items. Each item is composed by two figures. The figure at the left is unfolded, whereas the figure at the right depicts its folded version. The unfolded figure includes the numbers 1, 2, 3, 4, and 5 displayed at several lines, while the folded version includes ten letters also displayed at several lines. The unfolded figure must be folded mentally to decide the correspondence among the five numbers and five of the ten letters. The score was the total number of correct responses (well chosen letter–number pairs across the twelve items).

A.14. PMA-V

This is a synonym test that comprises 50 items. The meaning of four alternative letters must be evaluated against a given letter that serves as model. For instance, *STOUT: Sick–Fat–Short–Rude*. Only one alternative is correct. The score was the total number of correct responses.

A.15. DAT-VR

This is a reasoning test that comprises 50 items. A given sentence stated like an analogy must be completed. The first and last letters from the sentence are missing, so a pair of letters must be selected to complete the sentence from five possible alternative pairs of letters. For instance: *…is a water like eating is a…(A) Travelling–Driving, (B) Foot–Enemy, (C) Drinking–Bread, (D) Girl–Industry, (E) Drinking–Enemy*. Only one alternative is correct. The score was the total number of correct responses.

A.16. PMA-N

This is a calculation test that comprises 40 items. The participant must simply evaluate if a given sum is correctly or badly solved. For instance, $16+38+45=99$? The score was the total number of correct responses.

A.17. Monedas (coins)

This is a quantitative reasoning test comprising 40 items. The items are based on the combination of the size of a series of coins (large, medium, and small), the digits put inside the coins to specify the number of them that the participants must take into account, and some numerical operations to make the necessary calculations to arrive at a given response (adding, subtracting, and so forth). Only one alternative is correct. The score obtained was the total number of correct responses.

References

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.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(4), 567–589.

Baddeley, A. (2002). Is working memory still working? *European Psychologist*, *7*(2), 85–97.

Beier, M. E., & Ackerman, P. L. (2004). A reappraisal of the relationship between span memory and intelligence via "best evidence synthesis". *Intelligence*, *32*(6), 607–619.

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

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence*, *15*, 229–246.

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

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

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (in press). Complex span tasks, simple span tasks, and cognitive abilities: A re-analysis of key studies. *Memory and Cognition*.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. (2004). Working memory is (almost) perfectly predicted by *g*. *Intelligence*, *32*, 277–296.

Colom, R., & Shih, P. (2004). Is working memory fractionated onto different components of intelligence? *Intelligence*, *32*, 431–444.

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

Cowan, N. (2004). Working memory capacity limits in a theoretical context. In C. Izawa, & N. Ohta (Eds.), *Human learning and memory: Advances in theory and application* (pp. 155–175). Erlbaum.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, *3*, 422–433.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., et al. (2000, July). A neural basis for general intelligence. *Science*, *289*(21), 457–460.

Embretson, S. E. (1995). Working memory capacity versus general central processes in intelligence. *Intelligence*, *20*, 169–189.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23.

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.

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 ed.). Lincolnwood, IL: Scientific Software International.

Kane, M., & Engle, R. (2002). The role of prefrontal cortex in WMC, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review*, *9*(4), 637–671.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology. General*, *133*, 189–217.

Kline, P. (1998). *Principles and practice of structural equation modelling*. New York: Guilford Press.

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

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

Mackintosh, N. J., & Bennett, E. S. (2003). The fractionation of working memory maps onto different components of intelligence. *Intelligence*, *31*, 519–531.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 9–100.

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.

Rosen, V. M., & Engle, R. W. (1997). Forward and backward serial recall. *Intelligence*, *25*, 37–47.

Süβ, 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.

TEA (1997). Manual del test de factor *g* de Cattell [Handbook of the Cattell's Culture-Fair Intelligence Test]. Madrid: TEA.

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

Yela, M. (1969). *Desarrollo de superficies* [Surface development]. Madrid: TEA.