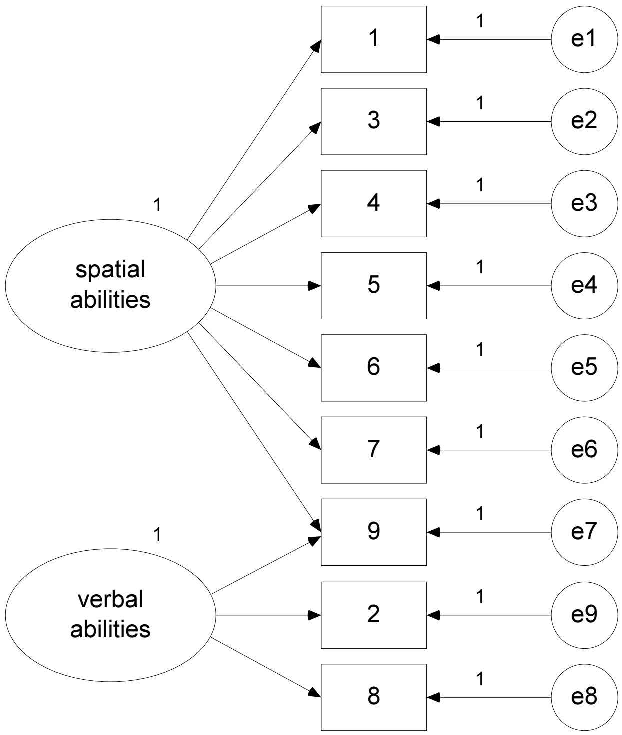
Structural Equation Modeling (SEM)

Christopher B. Germann
Department of Psychology Ψ
University of Plymouth



Legend:
1 Spatial Span
2 Reading Span
3 Simple Arrow Span
4 Paper Form Board Test
5 Space Relations Test
6 Clocks Test
7 Quantitative SAT
8 Verbal SAT
9 Identical Pictures Test

Figure1. Confirmatory factor analysis factor model with nine measured variables and two uncorrelated factors.

Contact: mail@christopher-germann.de

We employed AMOS to conduct a confirmatory factor analysis (CFA). CFA is a latent variable analysis in which "a hypothesized model is fit to a correlation matrix[1]" (Loehlin, 2004, p.94). Contrary to exploratory factor analysis CFA tests an explicitly a priori specified factor structure. Gorsuch (1983) notes that in comparison to exploratory factor analysis (EFA) "confirmatory factor analysis is powerful because it provides explicit hypothesis testing for factor analytic problems ... [and it] is the more theoretical important - and should be the much more widely used - of the two major factor analytic approaches" (p. 134). It has been argued that the use of exploratory factor analysis is only rational when the investigation using this technique is truly exploratory. A situation in which EFA is appropriate is for example when a researcher is interested in an area where no prior research has been carried out. In most other situation CFA should be employed and relevant literature should be consulted in order to specify hypothesis in advance.
The current analysis is based on the work of Shah & Miyake (1996) who experimentally demonstrated the separability of spatial and verbal working memory resources among college students. The researchers where interested in the question "whether spatial thinking and language processing are supported by separate pools of working memory resources" (p. 10). The investigators conducted an EFA which indicated a clear separation of spatial and verbal factors. We tested the hypothesized factor structure depicted in figure 1 using CFA. We decided to fix the factor variances of both factors to one in order to identify the model. Our analysis was based on the maximum likelihood estimation[2] and yielded results which supported the a priori specified factor structure indicating an acceptable overall fit of the model[3], $\chi^2(26, N = 54) = 26.56, p = .43$). (The χ^2 statistic for model fit is not statistically significant, meaning that the null hypothesis of a good fit to the data cannot be rejected.[4])

Until the present time it remains a controversial question which fit statistic(s) to use. It has been remarked that "although structural equation modeling is by now quite a mature field ... it is surprising that one of the basic elements of the modeling process ... the ability to evaluate hypothesized process models ... remains an immature art form rather than a science" (Bentler, 1994, p.257). However, in the past decades a plethora of fit indices

[1] This technique is not restricted to correlation matrices. For example, it is also possible to factor analyze covariance matrices.

[2] This is probably the most commonly used estimation in statistics. It was developed by R. A. Fisher in the 1920s.

[3] The null hypothesis (H_0) "is that the postulated model holds in the population" (Byrne, 2001, p.78) (i.e., $\sum = \sum(\theta)$) where $\sum$ represents the population covariance matrix and θ a vector that consists of the model parameters. Consequently the bigger the p-value associated with $\chi^2$the closer the fit between the under H_0 postulated model and the perfect fit (Bollen, 1989).

[4] It has been noted (i.e. Dickey, 1996) that χ^2 is not a really trustworthy fit index in many circumstances because it is influenced by the following factors: (a) Sample size: with larger samples the probability of α-errors increases, whereas small samples increase the likelihood to approve a poor model (inflation of β-error). (b) Model size: large models with many variables tend to be accompanied by larger χ^2values. (c) Distribution of variables: Significant non-normal skewness and kurtosis of variables increases χ^2values (i.c. multivariate normality assumption). (d) Omission of variables: By leaving out certain variables reproduction of the correlation matrix can become difficult.

have been developed (Raykov, 2005). Moreover, Thompson (2004) proposes to use multiple fit statistics in order to examine different aspects of fit because interpreting only one fit statistic can result in problematic interpretations. This model will consult the χ^2statistic, the Bentler Comparative Fit Index (CFI), the Goodness-of-Fit Index (GFI or $\hat{\gamma}$), and the Root Mean Square Error of Approximation (RMSEA). We also include statistics that take the number of parameters into account (PNFI and PCFI)[5] because adding parameters to a model will almost guarantee a better fit. Consequently more parsimonious (and more generalizable) models should be preferred (Dickey, 1996). In addition we employ the Browne-Cudeck Criterion (BCC) in order to facilitate model comparison (when comparing two models the one with the smaller BCC is prefered).

Our analysis yielded a Goodness-of-Fit Index (GFI) of .91, a Comparative Fit Index (CFI) of 1.00, and a root-mean-square error of approximation (RMSEA) of .02 (the 90% confidence interval[6] ranged from .00 to .11). According to the criteria suggested by (Hu & Bentler, 1999) (adequate fit indices are GFI and CFI of .95 or greater and RMSEA values of .05 or lower[7]) our results, indicate a good fit for the model (see Zhang, 2008, for a recent discussion on model fit indices). The higher the CFI and GFI the better the model fit to the data. The results of our analysis are summarized in table 1.

Table 1. Model fit indices

Test statistic	Result	Good fit?
χ^2	26.56	Yes
GFI	0.91	Yes
CFI	1.00	Yes
RMSEA	0.02	Yes
PNFI	0.62	-
PCFI	0.72	-
BCC	73.40	-

After inspection of the fit indices we focused on the individual regression coefficients. Regression weights that have a critical ratio[8] lower than $\mid 1.96 \mid$ are not statistically significant different from zero at an α-level of .05. Non-significant parameters can be regarded as unimportant in the interest of scientific parsimony (Byrne, 2001). We will report standardized regression weights because they allow "apples-to-apples comparisons of related parameters within a single model"(Thompson, 2004, p.133). As can be seen in table

[5] PNFI is the acronym for Parsimony Normed Fit Index and PCFI stands for Parsimony Comparative Fit Index. Both indices reward parsimonous models which have relatively few parameters to estimate.

[6] According to Herzog & Boomsma (2009) confidence intervals should be used to verify the precision of RMSEA values.The reported confidence interval is not really narrow indicating mediocre precision of the RMSEA estimate in expressing model fit in the population. However, it has been outlined that with small samples sizes RMSEA tends to overreject acceptable models (Hu & Bentler, 1999; Herzog & Boomsma, 2009) and confidence interval by their very nature become wider.

[7] In agreement with the literature, model fit is good if RMSEA $\in$ [0, 0.05], is acceptable if RMSEA [0.05, 0.08], and is unacceptable if RMSEA $>$ [0, 0.10].

[8] The critical ratio can be calculated by dividing the parameter estimate by its standard error. The resulting statistic is a t (or Wald) statistic (Thompson, 2004).

2 this is the case for the correlations between *Verbal SAT* and the latent construct labeled *verbal* and *Reading Span* and *verbal* and the *Identical Pictures Test* and *verbal.* [9].

Table 2. Regression weights

		Standardized Estimate	S.E.	C.R.	P
Spatial Span	← spatial	.74	.10	6.05	***
Simple Arrow Span	← spatial	.69	.10	5.49	***
Paper Form Board Test	← spatial	.60	.97	4.59	***
Space Relations Test	← spatial	.90	.643	8.10	***
Clocks Test	← spatial	.71	.96	5.77	***
Quantitative SAT	← spatial	.75	12.61	6.143	***
Identical Pictures Test	← spatial	.32	1.22	2.31	.02
Verbal SAT	← verbal	.83	38.69	1.79	.07
Reading Span	← verbal	.54	.27	1.69	.09
Identical Pictures Test	← verbal	.20	1.49	1.19	.24

Note: $^{***}p < .05$

Table 3. Squared multiple correlations

	Estimate
Spatial Span	.55
Simple Arrow Span	.47
Paper Form Board Test	.36
Space Relations Test	.81
Clocks Test	.51
Quantitative SAT	.56
Verbal SAT	.69
Reading Span	.29
Identical Pictures Test	.14

The squared multiple correlations (R^2) can be interpreted as follows: To take Space Relations Test as an example, 81% of its variance is accounted for by spatial ability. The remaining 19% of its variance is accounted for by the unique factor e4. If e4 represented measurement error only, we could say that the estimated reliability of Space Relations Test is 0.81.

(ii) Repeat the analysis but this time allow the two factors to correlate. Write a brief account of the results reporting the statistics that you consider to be most important and comment on the comparison between the results of this analysis and analysis (i).

We modified the first model by allowing the two factors to correlate. The obtained χ^2

[9] It should be noted that non-significant parameters can be symptomatic of too small sample size(Byrne, 2001).

indicated, as expected, an improvement of the model fit, $\chi^2(25, N = 54) = 22.79, p = .59$). Our computations resulted in a GFI of .92, a CFI of 1.00, and a RMSEA of .00 (95% CI ranged from .00 to .10). However, as expected, the modified model's PNFI and PCFI values were a bit lower than in the first model (0.61 vs. 0.62 and 0.69 vs. 0.72, respectively). The BCC indicated that the second model should be prefered. Taken together, the model fit indices suggest that the second model which allows for the two factors to correlate has a better fit than the first model.

Table 4. Model fit indices

Test statistic	Result	Good fit?	Model improvement?
χ^2	22.79	Yes	Yes
GFI	0.92	Yes	Yes
CFI	1.00	Yes	Yes
RMSEA	0.00	Yes	Yes
PNFI	0.61	-	No
PCFI	0.69	-	No
BCC	72.09	-	Yes

Subsequently we paid attention to the individual regression weights of the parameters. As depicted in table 4 the only non-significant correlations are between the Identical Picture Tests and both factors. By allowing the factors to correlate the other afore mentioned non-significant correlation reached statistical significance. However, the estimated factor intercorrelation of .33 was not significant (critical ration = 1.94, $p > .05$). It should be noted that the p-value was close to statistical significance (exactly .052). It could be argued that the sample size of 54 was not big enough to obtain enough power and precision (Meade & Bauer, 2007).

Table 5. Regression weights

		Standardized Estimates	S.E.	C.R.	P
Spatial Span	← spatial	.74	.10	6.01	***
Simple Arrow Span	← spatial	.68	.11	5.43	***
Paper Form Board Test	← spatial	.59	.97	4.56	***
Space Relations Test	← spatial	.90	.64	8.14	***
Clocks Test	← spatial	.72	.96	5.82	***
Quantitative SAT	← spatial	.75	12.57	6.19	***
Identical Pictures Test	← spatial	.30	1.35	1.96	.05
Verbal SAT	← verbal	.81	22.05	3.06	***
Reading Span	← verbal	.55	.172	2.68	***
Identical Pictures Test	← verbal	.20	1.53	1.15	.25

Note: ***$p < .05$

Table 6. Squared multiple correlations

	Estimate
Spatial Span	.54
Simple Arrow Span	.47
Paper Form Board Test	.35
Space Relations Test	.81
Clocks Test	.51
Quantitative SAT	.56
Verbal SAT	.66
Reading Span	.31
Identical Pictures Test	.16

Although we concluded that the hypothesized two factor model allowing for intercorrelations of the factors (see appendix) fit the data well we tried to identify any misfit in the postulated model. First we examined the matrix of the standardized residuals and found no values greater than 2.58 (a rule of thumb suggested by Joreskog & Sorbom, 1988)[10]. In addition, the modification indices gave no indication with regard to theoretical meaningful changes.

From these results one may conclude that the hypothesized model is appropriate and that the Identical Picture Test is redundant. Moreover, the current analysis confirmed, the separability of spatial and verbal memory resources Shah & Miyake (1996) postulated.

However, when drawing conclusions from CFA investigators should never be convinced that they have discovered the best fitting model. In fact it is always possible that other models fit a given dataset equally well or better. As (Kline, 1998) points out, terms like "good-fit" should be used with caution and rather conservatively. Instead terminology with more neutral connotations (i.e.: adequate, satisfactory or acceptable) should be used. Or as Arturo Rosenblueth (Rosenblueth & Wiener, 1945) formulated it: "The best model of a cat is another or, preferably, the same cat."

[10]The values in residual matrix indicate the discrepancy between the covariance matrix of the model $\sum(\theta)$ and the sample covariance matrix (S) (i.e., $\sum(\theta) - S$). Consequently, there is one residual for each pair of observed variables (Byrne, 2001).

References

Bentler, P. M. (1994). On the quality of test statistics in covariance structure analysis: Caveat emptor. In C. R. Reynolds (Ed.), *Cognitive assesment: A multidisciplinary perspective* (p. 237-260). New York: Plenum Press.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Byrne, B. M. (2001). *Structural equation modeling with AMOS : basic concepts, applications and programming.* London: Lawrence Erlbaum.

Dickey, D. (1996). Testing the fit of our models of psychological dynamics using confirmatory methods: An introductory primer. In D. Thompson (Ed.), *Advances in social science methodology* (p. 219-227). Greenwich, CT: JAI Press.

Gorsuch, R. L. (1983). *Factor analysis (2nd ed.).* Hillsdale, NJ: Erlbaum.

Herzog, W., & Boomsma, A. (2009). Small-Sample Robust Estimators of Noncentrality-Based and Incremental Model Fit. *Structural Equation Modeling*, *16*(1), 1-27.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling*, *6*, 1-55.

Joreskog, K. G., & Sorbom, D. (1988). *Lisrael 7: A guide to the program and applications.* Chicago: SPSS, Inc.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* NY: Guilford Press.

Loehlin, J. C. (2004). *Latent variable models : An introduction to factor, path, and structural equation analysis.* London: Lawrence Erlbaum.

Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling*, *14*(4), 611-635.

Raykov, T. (2005). Bias-corrected estimation of noncentrality parameters of covariance structure models. *Structural Equation Modeling*, *12*(1), 120-129.

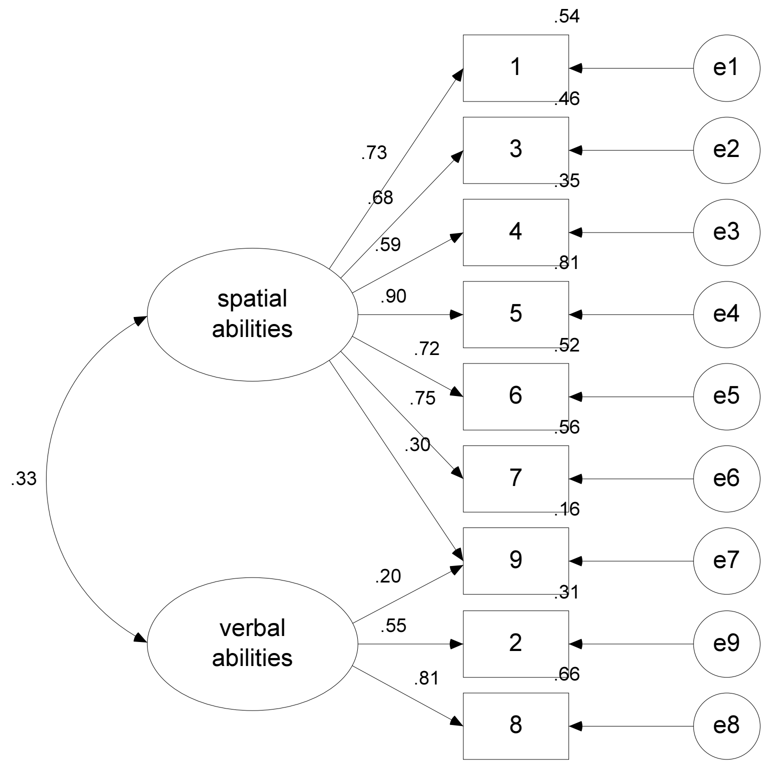
Rosenblueth, A., & Wiener, N. (1945). The role of models in science. *Philosophy of Science*, *12*, 316-321.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology*, *125*, 4-27.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis : Understanding concepts and applications.* Washington, DC: American Psychological Association.

Zhang, W. (2008). A comparison of four estimators of a population measure of model fit in covariance structure analysis. *Structural Equation Modeling*, *15*(2), 301-326.

Appendix



Legend:
1 Spatial Span
2 Reading Span
3 Simple Arrow Span
4 Paper Form Board Test
5 Space Relations Test
6 Clocks Test
7 Quantitative SAT
8 Verbal SAT
9 Identical Pictures Test

Figure 2. Revised hypothesized model with intercorrelated factors and standardized parameter estimates.