Psychology Science, Volume 49, 2007 (4), p. 326-342

Raven's Progressive Matrices, manipulations of complexity and measures of accuracy, speed and confidence

LAZAR STANKOV[1] & KARL SCHWEIZER[2]

Abstract

This paper examines the effects of complexity-enhancing manipulations of two cognitive tasks – Swaps and Triplet Numbers tests (Stankov, 2000) – on their relationship with Raven's Progressive Matrices test representing aspects of fluid intelligence. The complexity manipulations involved four treatment levels, each requiring an increasing number of components and relationships among these components. The accuracy, speed of processing, and confidence measures were decomposed into experimental and non-experimental parts and represented by the latent variables within a structural equation model. In the fitted model, four latent predictor variables had substantial path coefficients to Raven's Progressive Matrices test. Experimental accuracy scores for both Swaps and Triplet Numbers tests have significant predictive validity. Thus, complexity-enhancing manipulations affect correlations fluid intelligence captured by the Raven's test. In addition, two non-experimental latent variables (speed from Triplet Numbers and confidence from Swaps) have significant path coefficients.

Key words: intelligence, ability, complexity, Fixed-links Model

[1] The University of Sydney, Australia

[2] Correspondence should be addressed to: Karl Schweizer, Institute of Psychology, Johann Wolfgang Goethe University, Senckenberganlage 31, 60325 Frankfurt a. M., Germany; Tel.: 0049-69-798-23350, Fax.: 0049-69-798-23847, email: K.Schweizer@psych.uni-frankfurt.de

Introduction

From the individual differences point of view, being more intelligent implies the ability to solve more complex problems (see Lohman, 2000; Stankov, 1999). These are defined as problems that contain many elements and many inter-relationships among these elements (Crawford, 1991). There are, however, complications in the association of ability and task complexity. One is that the relationship between task complexity and ability is not linear. A human can automatically solve some problems that are computationally complex (Logan, 1985), and problems that are difficult for humans may be easy for machines (Carpenter, Just, & Shell, 1990). The perception of objects in space and working memory tasks are cases in point. It is for this reason that we need a method for choosing complex tasks that are clearly related to variations in human ability. This is due to the wiring of our nervous system that has emerged in the course of evolution and to the development of special skills for solving complex problems (Anderson, Fincham, & Douglass, 1997). Halford, Wilson and Phillips (1998) provide formal schemes for quantifying complexity in humans.

The method we suggest is empirical in nature, and it relies on the use of both experimental and correlational techniques (Cronbach, 1957). To understand intelligence, we need tasks that can be manipulated in a systematic way so that lower levels of the task place lower demands on the human processor, and higher levels of the task place higher demand. Given this design feature, there are two main statistics – means and variances/covariances – that are crucial for the examination of changes in task complexity. They each tell us about different issues related to task complexity as the term is defined above. Changes in arithmetic means reflect changes in the overall levels of performance. In psychometric literature, such changes are treated as indicative of task difficulty. From the individual differences point of view, changes (e.g., increase) in variances and covariances (or in correlations) are truly indicative of complexity. There is an assumption, often expressed by some experimental cognitive psychologists, that difficulty and complexity are the same constructs. This is not the case. For example, an infrequently used word that is present in a vocabulary test may be difficult when used for the assessment of crystallized intelligence (Hunt, 2000), but it is not complex. Since means and variances/covariances are statistically independent, it is logical to assume that one can have tasks that show an increase in one but not in the other. A good example is Raven's Progressive Matrices test. In our work, five difficulty levels of this test do not show systematic changes in correlation with other measures of fluid intelligence (see also Raven, Raven, & Court, 2003).

Thus, the complexity of a task is related to the increase in its correlation with other measures of, say, fluid intelligence or in the increase in factor loadings on a fluid intelligence factor. A rationale for this can be found in Thomson's sampling theory of intelligence which states that "each test calls upon a sample of the bonds which the mind can form, and that some of these bonds are common to two tests and cause their correlation." (Thomson, 1956; p.309). A contemporary exponent of this view is Detterman (1986). We wish to turn this argument around and claim that an increase in correlation is an indication of complexity whether one subscribes to Thomson's theory or not.

There have been several recent attempts to study cognitive sources of fluid intelligence. Some of the constructs that have been examined are attention (Schweizer, Moosbrugger, & Goldhammer, 2005; Stankov, 1983;1988), working memory (Ackerman, Beier, & Boyle, 2005; Buehner, Krumm, & Pick, 2005; Stankov & Myors, 1990), speed (Helmbold &

Rammsayer, 2006; Roberts & Stankov, 1999) and properties of the nervous system (Neubauer, Grabner, Fink, & Neuper, 2005; Stankov, Danthiir, Williams, Pallier, Roberts, & Gordon, 2006). Our approach to complexity is compatible with all these putative sources since complexity in this paper represents the description of behavior expressed in terms of test scores.

There exist constructs that have properties reflecting complexity in the sense we treat it in this paper. An example is working memory which is postulated to involve several interacting units (Baddeley, 1986) that serve the processing according to the demands of complex plans that require simultaneous transformation and storage of information (Bayliss, Jarrold, Gunn, & Baddeley, 2003). Working memory usually plays an important role in problem solving that is characterized by a long and complicated sequence of processing steps (Carpenter, Just, & Shell, 1990). The complexity that characterizes working memory and problem solving also applies to intelligence tests that show high loadings on factors of fluid and general intelligence (Stankov, 1983; 1999).

The scientific investigation of complexity requires the construction of tasks that can lay claim on being true measures of complexity (see also Raykov & Stankov, 1993; Roberts, Beh, & Stankov, 1988; Schweizer, 1993, 1996, 1998; Schweizer & Koch, 2003; Stankov, 1994, 2000, 2003; Stankov & Crawford, 1993; Stankov & Raykov, 1995). Stankov (2000) and Stankov and Raykov (1995) examined evidence for two such tasks – Swaps and Triplet Numbers tests – that are described in the method section below. The focus of that work was on showing that changes in task complexity lead to an increase in loadings on a factor of fluid intelligence. A positive finding was interpreted as a proof that fluid intelligence, at least in part, depends on the ability to carry out activities that require an increasing number of relatively simple steps to the solution and may be interrupted by momentary lapses of attention. Similar findings were reported by Schweizer (1996, 2007a) with the Exchange task that is akin to the Swaps test. The interpretation is, obviously, in terms of processing resources. The interpretation is also with respect to one dependent measure – accuracy of the response provided. It is important to note that in addition to the fluid intelligence factor, Stankov (2000) postulated two task-related factors (Swaps and Triplet Numbers factors) which, as it turned out, did not show the increasing trend on the size of loadings.

In this paper we use the data reported by Stankov (2000) to further validate previous findings. This extension of validation has methodological and substantive aspects. Figure 1 depicts the model that will be fitted.

Methodological Extensions. We are applying Structural Equation Modeling (SEM) procedures proposed by Schweizer (2006a, 2006b, 2007a) for the examination of the effects of complexity-enhancing manipulations of the kind considered in this paper. These procedures originated in the literature on growth modeling (see McArdle, 1986, 1988; McArdle & Epstein, 1987) and are referred to as fixed-links models. Consider the four left-hand manifest variables for the Swaps accuracy scores (SA-I to SA-IV). The essential feature of the model is the breakdown of variance and covariance into components associated with latent variables that represent different cognitive processes such as those produced through experimental treatments. Treatment levels guide the breakdown of variance and covariance. Thus, the levels of the manipulated task (i.e., four levels of the Swaps test) give rise to two latent components: non-experimental ("ne" in Fig. 1) and experimental ("ex" in Fig. 1). The model, of course, works on covariance matrices and, in its structural part, all loadings on the "ne" latent variable are constrained to be equal to 1. On the other hand, loadings on the "ex"

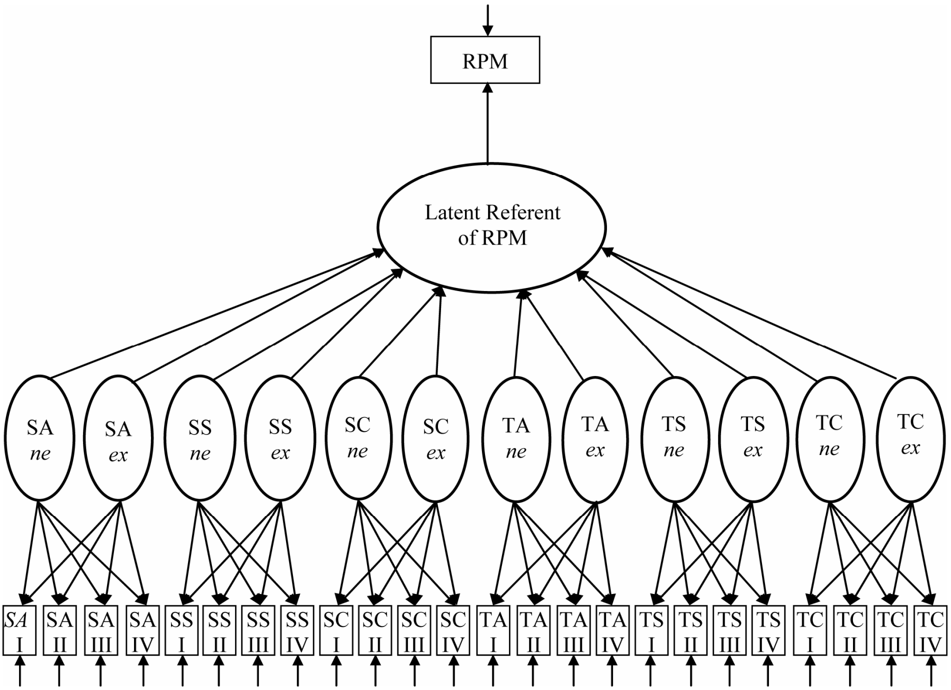

Figure 1:
Structural equation model that relates the latent variables representing the experimental and non-experimental parts of the measures of accuracy, speed of processing, and confidence to fluid intelligence (SA *ne* = Swaps – accuracy – non-experimental component; SA *ex* = Swaps – accuracy – experimental component; SS *ne* = Swaps – speed– non-experimental component; SS *ex* = Swaps – speed – experimental component; SC *ne* = Swaps – confidence – non-experimental component; SC *ex* = Swaps – confidence – experimental component; TA *ne* = Triplet numbers – accuracy – non-experimental component; TA *ex* = Triplet numbers – accuracy – experimental component; TS *ne* = Triplet numbers – speed – non-experimental component; TS *ex* = Triplet numbers – speed–experimental component; TC *ne* = Triplet numbers – confidence – non-experimental component; TC *ex* = Triplet numbers – confidence – experimental component; RPM= Raven's Progressive Matrices Test)

latent variable are allowed to vary and can be chosen to reflect linear, quadratic or, indeed, any other kind of trend[3]. These "ex" variables are expected to correlate with measures of fluid intelligence such as Raven's Progressive Matrices test. Latent variable "ne" may be a part of fluid intelligence, but it is constant across complexity-enhancing manipulations, and

[3] In effect, loadings on the fluid intelligence and task-related factors in Stankov's (2000) analyses parallel experimental and non-experimental components of the present approach. The main difference is in the requirement that loadings on the non-experimental latent variable be the same in the procedures of this paper; Stankov (2000) did not constrain loadings on the task-related factor(s) to be the same.

therefore it does not convey information about the effectiveness of the manipulations. Thus, "ne" latent variables capture information processing features of the task that are independent of changes in task complexity similar to what was captured by the Swaps and Triplets factors in Stankov's (2000) study.

Although the procedures outlined above originate in the growth curve modeling, there are differences as well. As is already obvious, the emphasis is not on the growth curve but on variances and covariances. Furthermore, mean structures are not a part of the fixed-links models. On the other hand there are similarities that may not be obvious. For example, there is similarity between the latent variable representing slope and the experimental latent variable and also similarity between the latent variable representing intercept and the non-experimental latent variable. All in all, the procedures enable the separation of the effects of different processes according to theoretical assumptions.

The other important feature of our model is the presence of a separate measure of fluid intelligence, Raven's Progressive Matrices (RPM) test. Unlike Stankov's (2000) analyses, Swaps and Triplet Numbers tests are not included in this factor. A hypothesis to be examined in this paper is that, for accuracy scores, experimental effects on Raven's test will be stronger than non-experimental effects, and the "ex" path will be significant for both Swaps and Triplet Numbers tests.

Substantive Extensions. In addition to typical *accuracy* scores, we employ two other measures that can be derived from the same cognitive act – i.e., the answer to a test item. These are the *speed* of providing an answer and *confidence* in the accuracy of the answer provided. Stankov (2000) shows that speed is indeed a separate factor from fluid intelligence that is measured by the accuracy scores. Confidence, too, is a separate factor from both fluid intelligence and speed, and it was interpreted as an aspect of a metacognitive skill of self-monitoring. Stankov (2000) did not examine the effects of complexity-enhancing manipulations on speed and confidence scores.

Are speed and confidence likely to show the same trend as accuracy scores? We believe that it is reasonable to assume that non-experimental and experimental latent variables for speed and confidence will behave in the same way as do the accuracy scores. After all, the three dependent variables are linked to the same cognitive act. We therefore fit identical experimental (ex) and non-experimental (ne) models to speed and confidence data.

As for the relationship to Raven's Progressive Matrices test (i.e., arrows leading from the 12 middle layer ovals to the top rectangle in Fig. 1), three possibilities are the most salient. One option is to have the same hypothesis as with accuracy scores – i.e., experimental (ex) rather than non-experimental (ne) latent variables for speed and confidence will affect RPM. The second option would be that non-experimental (ne) rather than experimental (ex) latent variables for speed and confidence will affect RPM. The third option is in-between the first two and allows for any pattern of experimental and non-experimental path coefficients of accuracy, speed, and confidence scores from Swaps and Triplet Numbers tests to lead to the RPM.

A theoretically interesting outcome would be to have significant paths to RPM from experimental accuracy latent variables and no paths from non-experimental accuracy latent variables. If, at the same time, speed and confidence measures have significant paths only from the non-experimental latent variables, an argument can be developed for the existence of two types of complexity that are captured by the size of path coefficients to RPM. Accuracy measures can be said to reflect changes to essentially the same task that imposes higher demands at each level. Speed and confidence, on the other hand, would reflect differences in complexity across different tasks or dependent measures. In other words, speed and confidence would be like any two measures of fluid intelligence, e.g., Raven's Progressive Matrices and Letter Series. To the extent that Raven's Progressive Matrices test has higher loading on the latent referent of RPM than does Letter Series test, we can claim that this test is a more complex task.

We shall use SEM to fit the model in Figure 1. We propose to use the pattern of significant and non-significant (zero) path coefficients to examine the effects of complexity-enhancing manipulations on fluid intelligence.

Aims

In this paper we examine the effects of complexity-enhancing manipulations of Swaps and Triplet Numbers tasks on a well-known measure of fluid intelligence, Raven's Progressive Matrices test. The manipulations involve four levels of increasing task complexity for each task. The causal model postulates the existence of two latent variables that capture experimental (ex) and non-experimental (ne) effects for each task. We employ three dependent measures: accuracy, speed of providing answers, and confidence. Our hypothesis is that complexity-enhancing manipulations from both tasks will affect RPM.

Method

Participants

The sample consisted of 345 participants. These were first-year psychology students at the Universities of Melbourne and Sydney, Australia. They took part in the experiment to satisfy their course requirements.

Study design

The study design included one criterion variable and 24 predictor variables. The set of predictor variables resulted from the systematic combinations of two types of test demands (Swaps and Triplet Numbers Tests), three observational methods (Accuracy, Speed, Confidence) and four treatment levels. The treatment levels differed systematically from each other in that each subsequent treatment imposed higher cognitive processing demands.

Criterion variable

Raven's Progressive Matrices[4]. The computerized version of Raven's Matrices included 40 items. Twenty items were adapted from "Standard Matrices" and the other items from "Advanced Matrices." The mean and standard deviation of the scores obtained from the 40-items Raven's Progressive Matrices test were 26.92 and 7.40 respectively.

Predictor variables

Swaps Test. The stimulus material for all versions of the Swaps test consisted of a set of three letters – J, K, and L – presented simultaneously on the computer screen, though not necessarily in that order. The instructions were to mentally interchange, or "swap," the positions of two of the letters. The four versions of the task differed in the number of such instructions. There were four blocks of 12 items with an equal number of swaps. For items consisting of two or more swaps, the participant had to keep track of the concurrent sequence. These items were randomly mixed to form a 48-item test. The participants did not know how many swaps would be required on any given trial. To preclude the possibility of memory for instructions limiting performance on this task, the required swap instructions were also kept visible throughout the participants' work. The answer consisted of typing the three letters in the order resulting from all the swaps. Examples of the four increasingly more involved sub-tasks are as follows:

Stimuli: J K L.
Swap1. "Swap 2 and 3." (Answer: J L K.)
Swap2. "Swap 2 and 3," "Swap 1 and 3." (Answer: K L J.)
Swap3. "Swap 2 and 3," "Swap 1 and 3," "Swap 1 and 2." (Answer: L K J.)
Swap4. "Swap 2 and 3," "Swap 1 and 3," "Swap 1 and 2," "Swap 1 and 3." (Answer: J K L.)

For each item, the accuracy of the response and time for responding was recorded and the result stored on hard disk. In addition, participants were asked for a confidence rating. They had to indicate how confident they were that the answer was correct using a 0 to 100 percent scale. For each treatment level, a confidence score and average speed of responding were computed for further analyses.

The Triplet Numbers test. The stimulus material for all versions of the Triplet Numbers Test employed in this study consisted of a randomly chosen set of three different digits presented simultaneously on the computer screen. These digits changed after each response. The four versions of this test differed with respect to the instructions given to the participants and time limits[5]. Instructions for the increasingly complex versions were as follows:

---

4 Stankov's (2000) study contained two additional tests of Gf: Letter Series and Counting Letters. Following on the request of an unknown reviewer, these two tests were excluded from the analyses presented in this paper.

5 Our experiences with this test indicate that Triplet1, Triplet2, and – to some extent – Triplet3 are so easy that participants start experiencing boredom and frustration if a 6-minute time limit (employed with Triplet4) is imposed. In our previous work, shorter versions of these tests showed satisfactory psychometric properties.

Triplet1. Press the "Yes" key if a particular number – e. g., 3 – is present within the triplet. Otherwise, press the "No" key. Maximum time allowed: 2 minutes

Triplet2. Press the "Yes" key if the second digit is the largest within the triplet. Otherwise, press the "No" key. Maximum time allowed: 3 minutes

Triplet3. Press the "Yes" key if the second digit is the largest and the third digit is the smallest . Otherwise, press the "No" key. Maximum time allowed: 5 minutes

Triplet4. Press the "Yes" key if the first digit is the largest and the second digit is the smallest *or* the third digit is the largest and the first digit is the smallest. Otherwise, press the "No" key. Maximum time allowed: 6 minutes

Three aspects of the responses were recorded for each item: (a) Accuracy, (b) Speed (time between the onset of the stimulus and response given), and (c) Confidence rating on a percentage scale. For each treatment level, a confidence score and average speed of responding were computed for further analyses. The accuracy scores of the fourth treatment level substantially deviated from the other levels and, therefore, was adjusted.

Statistical analyses

Structural Aspects of the Model. There were 25 manifest variables in this study (rectangles in Figure 1). One manifest variable, Raven's Progressive Matrices test, served as the indicator of the latent criterion variable (i.e., fluid intelligence), and the remaining manifest variables served as indicators of the latent predictor variables. Since there were no other manifest criterion variables and the internal consistency of Raven's Matrices was about .80, the loading and error component were fixed in such a way that the standardized loading was .80 and the standardized error component was .36. The 24 indicators of the latent predictor variables resulted from a systematic combination of the two tests (Swaps Test and Triplet Numbers Test), the three observational methods (accuracy, reaction time, confidence,) and four treatment levels (2 x 3 x 4 = 24).

In our model, there were 12 latent predictor variables. Two uncorrelated latent predictor variables were postulated for each quadruple (i.e., four treatment levels) of manifest variables associated with the same test and observational method. For example, two uncorrelated latent predictor variables were associated with four accuracy levels of the Swaps test. As we describe below, the two latent variables for each quadruple captured non-experimental effects ("level" in growth curves terminology) and experimental effects (i.e., change due to treatment). These latent variables were not allowed to correlate with each other. In contrast, the latent predictor variables were allowed to correlate across quadruples. Thus, latent variable for non-experimental effect on accuracy was allowed to correlate with latent variables for non-experimental effects for speed of responding and confidence.

The Fixing of the Links. The conventional model was transformed into fixed-links models in order to be able to decompose the observed measurements into components that resulted from the complexity-enhancing manipulation associated with the treatment levels and components that were independent of the complexity-enhancing manipulation. The transformation into a fixed-links model required the fixation of the links relating the latent variables to the manifest variables. This was achieved by assigning fixed numbers to the loadings, i.e., free loadings were replaced by constrained loadings.

The representation of non-experimental components (“level”) had to be achieved by the selection of a constant. The number “one” was selected for this purpose. The representation of the experimental effect was not expected to be associated with a specific curve (e.g. linear increase, quadratic increase) since there were indications of ceiling effects and equal distances between the treatment levels could not be assumed. Consequently, mixtures of curves were considered. This was achieved in successive steps. At first, the best-fitting type was selected and included into the model (e.g., 1, 2, 3, 4 for the linear increase due to the experimental treatment). Subsequently, the appropriateness of the representation was investigated, and adjustments were made in considering the modification indices provided by the program. The aim was to achieve the best possible representation of treatment levels for each quadruple of manifest variables. This procedure led to the creation of two vectors of numbers serving as constrained loadings (= fixed links). Standardized coefficients corresponding to the fixed links in the covariance structure solution are listed in rows “ne” and “ex” in Figure 2.

In order to examine the relative importance of the experimental and non-experimental components, we compared variances of the two latent variables that were linked to each quadruple of manifest variables. However, since the sizes of the variances depended on the sizes of constrained loadings, large loadings tended to be associated with small variances and small loadings with large variances. To avoid such a confounding of variances with loadings, the loadings were standardized (see Schweizer, 2007b for an example). These comparisons are reported in Table 2.

The Strategy. The main aim of this study was the investigation of the relationships between the latent predictor and criterion variable. This aim could be achieved in two ways. The first approach was the computation of correlations between each individual latent predictor variable and the criterion variable. The second approach was the computation of path coefficients (gamma coefficients in LISREL model) in order to find the most appropriate set of predictors. In this paper, we use both approaches (see Table 3 and Figure 2, respectively).

The data were analyzed with LISREL (Jöreskog & Sörbom, 2001).

Results

Model-data fit

The model-data fit was explored in two ways. First, in what we call the “pure” model, no correlations among error components were allowed. In the second, “adjusted” model we allowed three pairs of error components to correlate among themselves[6]. Table 1 gives the fit indices for these two models.

[6] The pairs of error components were Swaps-accuracy of second treatment level and Swaps-confidence of second treatment level; Triplet-accuracy of fourth treatment level and Triplet-reaction time of fourth treatment level; and also Triplet-accuracy of fourth treatment level and Triplet-confidence of fourth treatment level. Furthermore, two constrained loadings were set free.

Table 1:
Fit Statistics Obtained for the Fixed-Links Models Based on Accuracy, Confidence and Speed Data of Swaps and Triplet Numbers Tests (Stankov, 2000)

Model	χ^2	*df*	RMSEA	SRMR	GFI	CFI	NNFI	AIC	SMC
Pure model	761.68	256	.076 (.070-.082)	.078	.85	.91	.90	899.68	.30
Adjusted model	644.55	253	.067 (.061-.074)	.079	.87	.93	.92	788.55	.30

Note: Confidence interval is included within the parentheses.

The model-data fit of the pure model is acceptable. The ratio of chi-square and degrees of freedom was below 3, and RMSEA was below .08. The fit indices were above or near the borderline for an acceptable fit. The model-data fit of the adjusted model was better than the model-data fit of the pure model. The ratio of chi-square and degrees of freedom was close to 2, and RMSEA was close to .05. Furthermore, almost all the other fit indices indicated an improved model fit. The estimated parameters of the two models turned out to be very close. To avoid clutter, we present only the "pure" model parameters in the remainder of this paper; the main conclusions from the two models are the same.

Constrained experimental and non-experimental parameters

Figure 2 presents standardized coefficients for the "pure" SEM solution. There are two sets of constrained elements in this figure. The row labeled "ne" on the right-hand side provides coefficients for the arrows originating from the "ne" latent variables. In the non-standardized – i.e., covariance matrix – all these coefficients were fixed at one. They all differ from one in this row because of standardization – i.e., variances of the manifest variables differ among themselves – and because an increasing part of each manifest variables' variance is removed by the "ex" component. Underneath the "ne" row are the coefficients for the experimental "ex" latent variables. In the fitted solution based on covariances, these coefficients are constrained to reflect increases corresponding to complexity-enhancing manipulations. They are reflecting variations in task complexity. That is, an increasing proportion of variance for each higher complexity task is due to common variance that this task shares with other tasks in the battery.

Experimental vs. nonexperimental comparison: variances of the latent predictor variables

Since in the fixed-links models the loadings are constrained, the variances have to be estimated. The first column in Table 2 provides these variances.

It is obvious from Table 2 that all variances have reached the level of statistical significance. Thus, both experimental and non-experimental effects are significant.

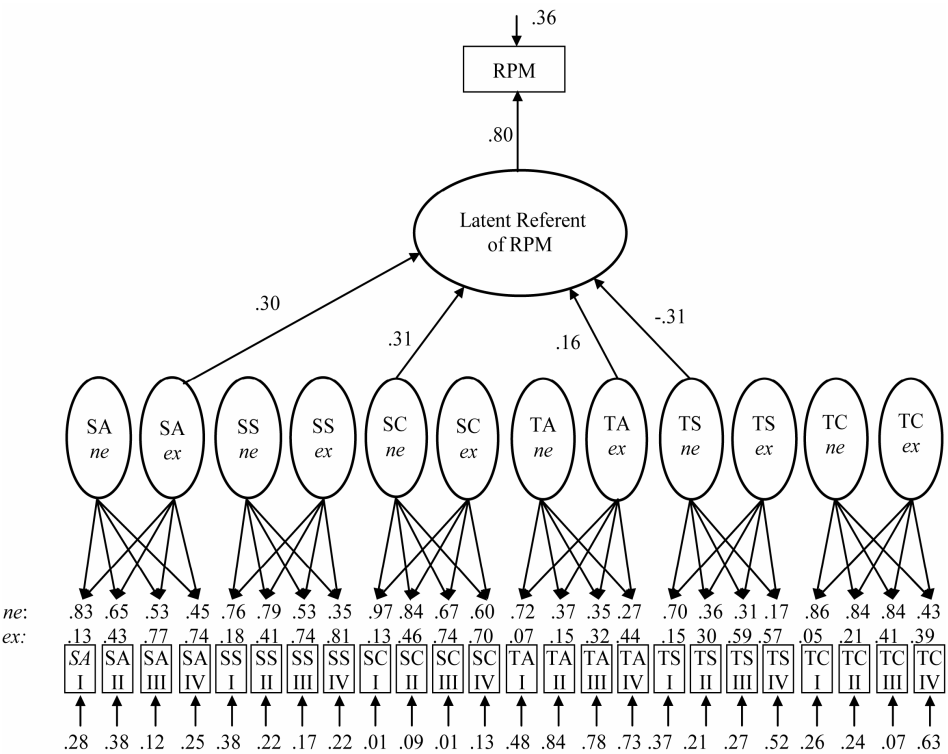

Figure 2:

Structural equation model that relates the substantial set of latent variables representing the experimental and non-experimental parts of the measures of accuracy, speed of processing, and confidence to fluid intelligence with the parameter estimates (The numbers given between the arrow heads and rectangles in the lower part of the Figure are standardized fixed loadings and residual variances.)

Because of standardization (see Schweizer, 2007b), it is possible to compare variances of the two latent variables with links to the same manifest variables (i.e., "ne" and "ex" latent variables). The second column of Table 2 provides values for these comparisons. These are percentages that add up to 100 % for the pairs consisting of experimental and non-experimental latent variables. For example, the percentages of the variances of Swaps accuracy – experimental and Swaps accuracy – non-experimental are 44.2 % and 55.8 %. Clearly, the experimental part of the variance is larger than the non-experimental part. Overall, larger experimental variances are found for Swaps accuracy, Swaps reaction times, and Triplet Numbers accuracy whereas larger non-experimental variances were observed for Swaps confidence, Triplet Numbers speed, and Triplet Numbers confidence. This finding is in general agreement with the hypothesis that complexity-enhancing manipulation will affect accuracy scores from both tests.

Table 2:
Variances of the Latent Variables Representing the Experimental and Non-experimental Parts of the Measures of Accuracy, Speed of Processing, and Confidence

Latent (predictor) variable	Variance	Percentage of Variance
Swaps accuracy - non-experimental	164.72*	44.2
Swaps accuracy - experimental	208.02*	55.8
Swaps speed - non-experimental	10.42*	33.5
Swaps speed - experimental	20.73*	66.5
Swaps confidence - non-experimental	150.69*	52.0
Swaps confidence - experimental	139.32*	48.0
Triplet accuracy - non-experimental	6.24*	52.5
Triplet accuracy - experimental	5.65*	47.5
Triplet speed - non-experimental	16.23*	67.3
Triplet speed - experimental	7.89*	32.7
Triplet confidence - non-experimental	27.71*	73.5
Triplet confidence - experimental	10.01*	26.5

* $p < .05$.

The correlations between latent predictor variables

Table 3 presents correlations between the latent predictor variables and, in the last row, correlations with the Raven's Progressive Matrices (RPM) scores.

Altogether, there are 60 non-zero correlations between the predictor variables. Correlations between six pairs of variables were fixed at zero since the corresponding latent variables were linked to the same set of manifest variables. These latent variables were expected to account for different parts of variance; fixing them at zero ensures their independence. Correlations above about .11 in Table 3 are significant at the .05 level[7]. The number of correlations among the predictor variables reaching the level of significance in Table 3 is 38 (63.3 %).

Overall, correlations between components of the same test (Swaps or Triplet Numbers) are higher than correlations between components of different tests. The highest correlations are between non-experimental accuracy and non-experimental confidence (Swaps: .81, Triplet Numbers: .86) and also between experimental accuracy and experimental confidence (Swaps: .63, Triplet Numbers: .86). The correlations for the pairs of experimental accuracy and experimental speed are also high (Swaps: .67, Triplet Numbers: .64). Moderate correlations between the two tests were found for non-experimental accuracy (.48) and non-experimental speed (.59). Finally, correlations between experimental accuracy and experimental confidence were high (.63 for Swaps and .89 for Triplet Numbers). The pattern of correlations suggests that there are relatively strong relationships between accuracy and confidence and between accuracy and speed.

[7] Each correlation coefficient in Table 3 has its own significance level. The value of .11 that is reported in the text as a cut-off number is only an approximation.

Table 3:
Correlations Between the Latent Variables Representing the Experimental and Non-experimental Parts of the Measures of Accuracy, of Speed of Processing, and of Confidence and Fluid Intelligence (Adjusted Model)

Code of variable	SA *ne*	SA *ex*	SS *ne*	SS *ex*	SC *ne*	SC *ex*	TA *ne*	TA *ex*	TS *ne*	TS *ex*	TC *ne*	TC *ex*
SA *ne*	1.00											
SA *ex*	- -	1.00										
SS *ne*	.10	-.38	1.00									
SS *ex*	.40	.67	- -	1.00								
SC *ne*	.81	.01	.19	.29	1.00							
SC *ex*	-.01	.63	-.27	.50	- -	1.00						
TA *ne*	.48	-.01	.16	.20	.33	.09	1.00					
TA *ex*	.25	.29	-.09	.37	.02	.00	- -	1.00				
TS *ne*	-.08	-.19	.59	-.16	.01	-.01	.14	-.15	1.00			
TS *ex*	.13	.17	.01	.40	.00	.02	.07	.64	- -	1.00		
TC *ne*	.18	.03	.03	.08	.28	.18	.86	-.35	.00	.01	1.00	
TC *ex*	.12	.31	.09	.36	.14	.18	-.35	.89	.03	.04	- -	1.00
RPM	.29	.36	-.22	.36	.29	.17	.05	.27	-.33	.14	.03	.25

Note. SA *ne* = Swaps – accuracy – non-experimental component; SA *ex* = Swaps – accuracy – experimental component; SS *ne* = Swaps – speed – non-experimental component; SS *ex* = Swaps – speed – experimental component; SC *ne* = Swaps – confidence – non-experimental component; SC *ex* = Swaps – confidence – experimental component; TA *ne* = Triplet numbers – accuracy – non-experimental component; TA *ex* = Triplet numbers - accuracy – experimental component; TS *ne* = Triplet numbers – speed – non-experimental component; TS *ex* = Triplet numbers – speed – experimental component; TC *ne* = Triplet numbers – confidence – non-experimental component; TC *ex* = Triplet numbers – confidence – experimental component. RPM = Raven's Progressive Matrices test.

Correlations Between latent predictors and criterion variable (RPM)

The last row in Table 3 presents correlations between all latent predictor variables and a criterion variable (RPM) representing fluid intelligence. Almost all correlations in the last row are significant. There are only two exceptions: the correlation with the non-experimental component of the Triplet Numbers confidence and with non-experimental component of Triplet Numbers accuracy. Almost all correlations are positive. There are only two exceptions: the non-experimental components of speed correlated negatively with fluid intelligence. The range of the correlations (absolute values) is considerable (lower limit: .03, upper limit: .36).

The prediction of RPM test scores (Fluid Intelligence)

Path coefficients (i.e., gamma coefficients in LISREL notation) between latent variables and RPM are also presented in Figure 2. Only four path coefficients reach the .05 level of significance. Two of these path coefficients relate experimental accuracy components to fluid intelligence. These are the path coefficients of the experimental components of Swaps accuracy (.30) and of Triplet Numbers accuracy (.16). This is in accordance with the expectations – accuracy scores from the experimental tasks show increasing correlations with the accuracy scores from the RPM. At the same time, non-experimental components of accuracy scores do not show increasing correlations with RPM.

The other two significant path coefficients relate non-experimental components to fluid intelligence. They are path coefficients of the non-experimental components of Swaps confidence (.31) and of Triplet Numbers speed (-.31). As expected, path coefficients of the non-experimental components of Triplet Numbers speed were the only path coefficients that are negative.

The small number of substantial path coefficients is not a surprising result since the correlations in Table 3 made it obvious that there is a degree of co-linearity among the latent predictor variables. It is obvious from the squared multiple correlation presented in the last column of Table 1 that these four predictors lead to a sizeable multiple correlation of about .55.

Summary and discussion

The analyses reported in this paper extend our knowledge about the role of complexity manipulations in Raven's Progressive Matrices test and, by extension, of fluid intelligence. It seems meaningful to conceive of the complexity of a particular task as consisting of two parts – a part that is sensitive to complexity-enhancing manipulations and a part that is constant across different levels of complexity-enhancing manipulations. This distinction is made in the analyses of this paper and also in the analyses reported by Stankov (2000). From both sets of analyses it is clear that accuracy measures are sensitive to complexity manipulations on both Swaps and Triplet Numbers tasks, and that experimental rather than non-experimental component of accuracy scores is affecting fluid intelligence. This finding is in agreement with findings based on the manipulation of the demands on working memory (Carpenter, Just, & Shell, 1990; Stankov & Myors, 1990). On this basis alone, we can say that the increase in the number of steps to complete the task, which is a salient feature of the Swaps test, or the increase in the number of components in the Triplet Numbers task are important aspects, perhaps the essence, of fluid intelligence as captured by the RPM.

Although the variance of speed and confidence measures can also be divided into the same two parts – i.e., experimental and non-experimental – the experimental part of neither measure shows a relationship to fluid intelligence. Therefore, speed and confidence are not the essence of RPM and fluid intelligence in the same way as accuracy scores are. But they cannot be ignored either, since non-experimental components of Swaps confidence and Triplet Numbers speed also affect fluid intelligence. These last two findings will need to be replicated. Significant path coefficients for the two non-experimental latent variables imply that fluid intelligence captured by RPM test is broader in this study than as it is usually conceived – it is not limited to accuracy scores only, but also to speed and confidence from the latent predictor variables. As mentioned in the introduction, the presence of the significant paths from the non-experimental latent variables to complexity-enhancing manipulations and to RPM suggests that it may be meaningful to distinguish between two conceptualizations of complexity: experimental and task-related.

Replication of this finding is in order for the following two reasons. First, there is ample evidence in the literature that fluid intelligence defines a factor that differs from speed (see Carroll, 1993) and confidence (Kleitman & Stankov, 2007; Stankov, 2000). Second, it is not readily apparent why the other two non-experimental latent variables – Swaps speed and Triplet Numbers confidence – did not have significant path coefficients as well.

In this paper we limit ourselves to the typical measurement of fluid intelligence in terms of accuracy scores from a well-known measure of Gf. There are many other options that we have not addressed here. For example, measures of Gf such as Raven's Progressive Matrices, can also be used to assess speed and confidence. Thus we may have three rectangles – accuracy, speed, confidence – at the top of Figure 1. What happens to "ne" and "ex" path coefficients in this condition? Some careful theorizing about the nature of intelligence that can guide the choice of models to be fitted is clearly in order.

We believe that an improved understanding of abilities at different levels in a hierarchy (primary or broad abilities) can be achieved by a combination of the experimental and correlational approaches we have employed in this paper. One can ask, for example, what are the manipulations that lead to changes in loadings on the mental speed or confidence factors? A systematic program of research along similar lines can be of theoretical interest because it will lead to a more precise definition of human abilities. Inevitably, this would generate practical benefits because of the possibility for systematic creation of tests that would be suitable for particular ability and developmental levels.

References

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

Anderson, J. R, Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23*, 932-945.

Baddeley, A. D. (1986). *Working memory.* Oxford: Clarendon 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*, 71-92.

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning = working memory ≠ attention. *Intelligence, 33*, 251-272.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404-431.

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

Crawford, J. D. (1991). Intelligence, task complexity and the distinction between automatic and effortful mental processing. In H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement.* Hillsdale, NJ: Lawrence Erlbaum.

Crawford, J. & Stankov, L. (1983). Fluid and crystallized intelligence and primacy/recency measures of short-term memory. *Intelligence, 7,* 227-252

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 30*, 116-127.

Detterman (1986). Human intelligence is a complex system of separate processes. In R. J. Sternberg & D. K. Detterman (Eds.), What is intelligence: Contemporary viewpoints on its nature and definition. (pp. 57-61). Norwood, NJ: Ablex.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*, 803-864.

Horn, J.L. (1965). Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities. Ph.D. Thesis. University of Illinois.

Helmbold, N., & Rammsayer, T. (2006). Timing performance as a predictor of psychometric intelligence as measured by speed and power tests. *Journal of Individual Differences, 27*, 20-37.

Hunt, E. (2000). Let's hear it for crystallized intelligence. *Learning and Individual Differences, 12*, 123-129.

Jöreskog, K.G. & Sörbom, D. (2001). *LISREL* (Version 8.5). *Computer Software.* Chicago: Scientific Software International.

Logan, G. D. (1985). Skill and automaticity: relations, implications, and future directions. *Canadian Journal of Psychology, 39*, 367-386.

Lohman, D. F. (2000). Complex information processing and intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 285-340). New York: Cambridge University Press.

Kleitman, S. and Stankov, L. (2007). Self-confidence and Metacognitive Processes. *Learning and Individual Differences,* 17, 161-173.

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics, 16*, 163-200.

McArdle, J. J. (1988). Growth curve analysis in contemporary psychological research. In J. R. Nesselroade & R. B. Cattell (eds.), *Handbook of multivariate experimental psychology* (2nd ed.) (pp. 561-614). New York: Plenum.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110-133.

Neubauer, A. C, Grabner, R. H., Fink, A., & Neuper, C. (2005). Intelligence and neural efficiency: Further evidence of the influence of task content and sex on the brain--IQ relationship. *Cognitive Brain Research, 25*, 217-225.

Raven, J., Raven, J.C., & Court, J.H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General Overview.* San Antonio, TX: Harcourt Assessment.

Raykov, T. & Stankov, L. (1993) On task complexity and "simplex" correlation matrices. *Australian Journal of Psychology, 45*, 81-87.

Roberts, R. & Stankov, L. (1999) Individual differences in speed of mental processing and human cognitive abilities: Towards a taxonomic model. *Learning and Individual Differences, 11*, 1, 1-120.

Roberts, R. D., Beh, H. C. & Stankov, L. (1988) Hick's law, competing tasks, and intelligence. *Intelligence, 12*, 2, 111-131.

Schweizer, K. (1993). The contribution of access to external information, stimulus complexity, and variability to cognitive abilities. *Personality and Individual Differences, 14,* 87-95.

Schweizer, K. (1996). The speed-accuracy transition due to task complexity. *Intelligence, 22,* 115-128.

Schweizer, K. (1998). Complexity of information processing and the speed-ability relationship. *The Journal of General Psychology, 125,* 89-102.

Schweizer, K. (2001). On the role of mechanisms when the processing complexity is high. *European Psychologist, 6*, 133-143.

Schweizer, K. (2006a). The fixed-links model in combination with the polynomial function as tool for investigating choice reaction time data. *Structural Equation Modeling, 13*, 403-419.
Schweizer, K. (2006b). The fixed-links model for investigating the effects of general and specific sources. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2,* 149-160.
Schweizer, K. (2007a). Investigating the relationship of working memory tasks and fluid intelligence tests by means of the fixed-links model in considering the impurity problem. *Intelligence, 35*, 591-604.
Schweizer, K. (2007b). Investigating experimental effects within the framework of structural equation modeling: an example with effects on both error scores and reaction times. *Structural Equation Modeling* (in press).
Schweizer, K. & Koch, W. (2003). Perceptual processes and cognitive ability. *Intelligence, 31,* 211-235.
Schweizer, K., Moosbrugger, H., & Goldhammer, F. (2005). The structure of the relationship between attention and intelligence. *Intelligence, 33*, 589-611.
Stankov, L. (1983) Attention and intelligence. *Journal of Educational Psychology, 75*, 4, 471-490.
Stankov, L. (1988) Aging, intelligence and attention *Psychology and Aging, 3*, 2, 59-74.
Stankov, L. (1994) The complexity effect Phenomenon is an epiphenomenon of age-related fluid intelligence decline. *Personality and Individual Differences*, *16*, 265-288.
Stankov, L. (1999). Die Beziehung zwischen Komplexität und Intelligenz. (The relationship between complexity and intelligence). In K. Schweizer (Ed.) *Intelligenz und Kognition* (pp. 37-63). Landau: Verlag fur Empirische Pädagogik.
Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence, 28*, 121-143.
Stankov, L. (2003). Complexity in human intelligence. In R. Sternberg, J. Lautrey, & T. Lubart (Eds.) *Models of Intelligence: International Perspective*. Washington, DC: American Psychological Association.
Stankov, L. & Crawford, J. D. (1993). Ingredients of complexity in fluid intelligence. *Learning and Individual Differences 5*, 73-111.
Stankov, L. & Myors, B. (1990). The relationship between working memory and intelligence: Regression and COSAN analyses. *Personality and Individual Differences 11*, 1059-1068.
Stankov, L. & Raykov, T. (1995) Modeling complexity and difficulty in measures of fluid intelligence, *Structural Equation Modeling, 2*, 335-366.
Stankov, L., Danthiir, V., Williams, L., Pallier, G., Roberts, R., Gordon, E. (2006). Intelligence and the tuning-in of brain networks. *Learning and Individual differences, 16*, 116-146.
Thomson, G. H. (1956). *The factorial analysis of human ability.* London, UK: University of London Press Ltd.