Investigating the Item-position Effect in a Longitudinal Data with Special Emphasis on the Information provid-ed by the Variance Parameter

Tengfei Wang[1], **Qiong Zhang**[1] **& Karl Schweizer**[1,2]

[1] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China
[2] Institute of Psychology, Goethe University Frankfurt, Frankfurt a. M., Germany

Abstract:
Although the item-position effect has frequently been observed in intelligence testing among adults, it remains unclear whether there is such an effect in children and how it develops over time. Data on Raven's Standard Progressive Matrices (SPM) were collected from 189 primary school-aged children (10-11 years old) twice, with an interval of one and a half years. The item-position effect of SPM was represented separately from the ability-specific component by means of fixed-links modeling. The variance parameters of the latent variables in the model were thoroughly analyzed. The results indicated that including the item-position factor yielded better model fit to the SPM data collected at both time points. The variances of both the ability-specific and item-position factors were significant, confirming the presence of the item-position effect in addition to ability. The comparison of the scaled estimates of the variance parameters showed that the item-position effect accounted for more variances in SPM scores at Time 2 (72.08%) than Time 1 (48.10%). Furthermore, the correlation between the item-position factors (r = 0 .88) across two time points was much smaller than that between the ability-specific factors (r = 0 .34). Taken together, these results demonstrate a substantial change of the item-position effect underlying SPM as children grow up.

Keywords:
item-position effect, fluid intelligence, Raven's Standard Progressive Matrices, fixed-links modeling

Introduction

The item-position effect is a method effect that was described by Campbell and Mohr (1950) and Mollenkopf (1950) in the same year for the first time. This effect means that there is some degree of dependency of the item statistics on the positions of these items within the sequence of items. For factor-analytic research into this effect an investigation by Knowles (1988) is especially important. Knowles reports that the reliabilities of the items increase from the first to the last items. This increase is observed in considering several different arrangements of the individual items. This observation suggests that the sequence of items creates systematic variation that increases from the first to the last items independently of the positions of the individual items. It reveals that systematic variation can be perceived as a function of the position of the item. It is additional systematic variation besides the variation due to the construct represented by the scale. Assume that σ_i^{C} is the complete systematic variance of the *i*th item ($i = 1, \ldots, p$) that is composed of the part due to the genuine source of responding $\sigma_{\mathrm{genuine}\_i}^{\mathrm{C}}$ and the part due to the source of the item-position effect $\sigma_{\mathrm{IP}\_i}^{\mathrm{C}}$ so that

$$\sigma_i^{\mathrm{C}} = \sigma_{\mathrm{genuine}\_i}^{\mathrm{C}} + \sigma_{\mathrm{IP}\_i}^{\mathrm{C}} \quad (1)$$

In confirmatory factor models systematic variance is represented by the product of three multipliers: the factor loading λ, the model-specific systematic variance σ^{m} and the transpose of the factor loading λ' so that Equation 1 can be re-written as

$$\sigma_i^{\mathrm{C}} = \lambda_{\mathrm{genuine}\_i}\sigma_{\mathrm{genuine}\_i}^{\mathrm{m}}\lambda_{\mathrm{genuine}\_i}' + \lambda_{\mathrm{IP}\_i}\sigma_{\mathrm{IP}\_i}^{\mathrm{m}}\lambda_{\mathrm{IP}\_i}' \quad (2)$$

Now function $\mathrm{f}_{\mathrm{IP}}(\ )$ can be introduced that describes how the size of the systematic variance is influenced by the item position. This function that can be defined differently replaces λ_{IP_i} in Equation 2 for representing the item-position effect. Furthermore, there is function $\mathrm{f}_{\mathrm{genuine}}(\ )$ that describes how the genuine source of the scale reflecting the measured construct contributes to the systematic variance. It replaces $\lambda_{\mathrm{genuine}\_i}$. The replacements lead to the following equation that provides a formal description of the complete systematic variance of item *i*:

$$\sigma_i^{\mathrm{C}} = \mathrm{f}_{\mathrm{genuine}}(i)\sigma_{\mathrm{genuine}\_i}^{\mathrm{m}}\mathrm{f}_{\mathrm{genuine}}(i)' + \mathrm{f}_{\mathbf{P}}(i)\sigma_{\mathrm{IP}\_i}^{\mathrm{m}}\mathrm{f}_{\mathbf{P}}(i)' \quad (3)$$

Knowles' observation provided the basis for the factor-analytic investigation of the effect since factor analysis is designed to capture systematic variation of data. The paper by Knowles is not especially specific regarding the course of the increase. What is apparent is that there is a small monotonic increase. In a recent study, possible representations of increase were systematically investigated (Zeller, Krampen, Reiss, & Schweizer, 2017). The relational pattern used in this study was obtained from a large sample of APM (Raven, Raven, & Court, 1997) data. Structured random data simulated according to this pattern were investigated. The best model fit was observed for a piecewise function that was adapted to the data. However, it did not substantially differ from the course of the effect according to the quadratic function:

$$\mathrm{f}_{\mathbf{P}}(i) = \left(\frac{i-1}{p-1}\right)^2 \quad (4)$$

where i = 2, ..., p, p the number of items and $\mathrm{f}_{\mathrm{IP}}(1) = 0$. The position number is divided by p to keep the values within a limited range. So far it is not clear whether the quadratic function is the best function for data collected by all reasoning scales or whether the effect follows a function that is characteristic for APM only. Furthermore, characteristics of the sample may play a role.

The factor-analytic investigation of the item-position effect started with the investigation of this effect in personality items by Hartig, Hölzel, and Moosbrugger (2007) using a model of measurement that included one latent variable only. Such a model requires the integration of two effects. One effect is the effect that is captured by the scale and the other one is the item-position effect. It is necessary to find constraints that capture both effects simultaneously. The confirmatory factor model by Hartig et al. (2007) stays within the framework of the congeneric model (Jöreskog, 1971) but also differs from it in that the factor loadings are constrained instead of estimated.

Furthermore, there is the two-factor model by Schweizer, Schreiner, and Gold (2009). Two-factor models have become popular for the investigation of multitrait-multimethod data. The two-factor confirmatory factor model for investigating the item-position effect includes two latent variables: one for capturing the effect of the source underlying the scale and the other one for capturing the item-position effect. Following McDonald (1965) we address the factor associated with the source captured by the scale as genuine factor and the other one as item-position factor and signify this by the subscript IP:

$$\mathbf{x} = \lambda_{\mathrm{genuine}}\,\xi_{\mathrm{genuine}} + \lambda_{\mathrm{IP}}\,\xi_{\mathrm{IP}} + \delta \tag{5}$$

where $\mathbf{x}$ is the $p\times1$ vector of manifest variables, $\lambda_{\mathrm{genuine}}$ and λ_{IP} the $p\times1$ vectors of factor loadings, ξ_{genuine} and ξ_{IP} the latent variables and δ the $p\times1$ vector of error variables.

An important characteristic of the investigation of the item-position effect is the constraint of factor loadings. This means that the factor loadings are set equal to numbers. For example, there is reason for constraining the factor loadings on the latent variable capturing the item-position effect according to Equation 4. This means that the factor loadings of the ith item on the item-position factor λ_{IP} is given by

$$\lambda_{\mathbf{P}}(i) = \mathrm{f}_{\mathbf{P}}(i) \tag{6}$$

and the vector of factor loadings by

$$\ddot{\mathbf{e}}_{\mathbb{P}} = \begin{bmatrix} \lambda_{\mathbf{P}}(1) \\ \lambda_{\mathbf{P}}(2) \\ . \\ . \\ \lambda_{\mathbf{P}}(p) \end{bmatrix} \tag{7}$$

The constraint of factor loadings assures that the latent variable captures what it is expected to capture. In contrast, in the case of free factor loadings the factor accounts for as much variance as is possible. This can mean that other effects are partly or completely accommodated but this stays unknown. Furthermore, the constraint of the factor loadings on a factor creates a strong hypothesis. We refer to a hypothesis as strong hypothesis if it can be easily rejected. A strong hypothesis depends on one or a few parameters only whereas weak hypotheses depend on a larger number of parameters.

Such a hypothesis is merged with the corresponding model of measurement so that the investigation of the model of measurement implies the investigation of the hypothesis (Schweizer, 2008) If this hypothesis does not apply, model misfit can be expected. Not all the factor loadings need to be constrained if a specific hypothesis is to be investigated. There is the possibility to use a hybrid model that includes constraints for the factor representing the hypothesis and free factor loadings on other factors (e.g., Schweizer, Reiß, Ren, Wang, & Troche, 2019).

In models with fixed factor loadings the influences of the sources captured by the latent variables find its expression in the variance parameter that is estimated. The variance parameter is part of the model of the covariance matrix (Jöreskog, 1970). This $p \times p$ matrix Σ is defined as

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \quad (8)$$

where Λ is the $p \times q$ matrix of factor loadings, Φ the $q \times q$ matrix of the variances and covariances of the latent variables and $\boldsymbol{\Theta}$ the $p \times p$ diagonal matrix of error variances.

The estimate of the variance parameter indicates the amount of variance that is explained by the corresponding latent variable (Schweizer, Troche, & DiStefano, 2019). For achieving comparability of the variances of different latent variables, appropriate scaling is necessary. There is one scaling method that is frequently considered in designing models of measurement: the marker-variable method. It does not allow the comparison of variances because the result depends on the selected marker so that different results are achieved for different markers. We prefer a criterion-based method that enables comparability of the estimated variances (Schweizer, 2011). The EV scaling method produces estimates that are comparable when estimated within the framework of the same model of measurement. A characteristic of the modified version of this scaling method (Schweizer et al., 2019) is that the estimated variance ϕ_{EV} corresponds to the sum of squared factor loadings under the condition that the factor loadings are estimated with the variance parameter of the model set equal to one:

$$\phi_{EV} = \boldsymbol{\lambda}_S{}'\boldsymbol{\lambda}_S \quad (9)$$

where λ_S is the $p \times 1$ vector of factor loadings and the subscript S signifies the specific condition leading to the non-standardized estimates of the factor loadings. The estimated variance ϕ_{EV} represents systematic variation in the sense of complete systematic variation as, for example, $\sigma^{C}_{genuine\_i}$ and $\sigma^{C}_{IP\_i}$ Furthermore, under this specific condition variance estimates correspond to eigen values if everything else is kept constant as, for example, the estimation method (Schweizer, Troche, & Reiss, 2017).

So far the variance parameter has not played a major role in investigations by confirmatory factor analysis since the focus was on model fit. However, the variance parameter of the confirmatory factor model can provide additional valuable information. It can be considered as a statistic that can be used for the same purpose as the regressions weight in multiple regression analysis (Schweizer et al., 2019). If there are several predictors, i.e., latent variables, the scaled variance parameters signify the relative influences of the predictors on the criterion variable.

The research work presented in this paper focuses on the position effect in a fluid intelligence measure (i.e., Raven's Standard Progressive Matrices) among children. Specifically, we aim to investigate: (1) whether there is the item-position effect in a fluid intelligence measure for children when their cognitive abilities are still under development, (2) whether the variance explained by the item-position effect changes as children grow up, and (3) whether the item-position effects observed at different time points are related to each other in longitudinal data. In order to answer these questions, fixed-links models, in which factor loadings are fixed according to the characteristics of the hypothesized construct, are employed. In these models, two sources of fluid intelligence, including the ability-specific and item-position effects, are represented by two latent variables. The statistical investigation of the variance parameter associated with the item-position effect can signify whether the data show this effect or not. The scaling of the estimates of variance parameters (Schweizer et al., 2019) establishes the precondition for investigating the relative importance of the item-position effect compared to ability. Higher percentages of the scaled variance for the item-position effect can signify a larger role played by the item-position effect. Furthermore, the substantial or negligible relationship between the item-position effects estimated at the two time points signifies either the stability or variability of the item-position effect, respectively.

Method

Participants

Participants were 189 primary school students in 5th grade, who were followed through 6th grade. There were 118 males and 71 females with a mean age of 10.71 years (SD = .29) at time point one and a means age of 12.21 years (SD = .29) at time point two. Participants received a gift for their participation.

Measures and procedure

Raven's Standard Progressive Matrices (SPM, Raven, Raven, & Court, 1998). SPM was composed of five sets (A to E), each including 12 items. The items within a set were arranged in an ascending order of difficulty. Each item consisted of a matrix in which one part was missing. Participants were asked to complete a matrix by selecting the most appropriate geometric figure from 6 (sets A and B) or 8 (sets C, D, and E) alternatives. Responses to each item were recorded as binary data. A correct response was coded as 1 whereas an incorrect response as 0. Participants were given 15 min to complete all items.

Participants were tested in classroom groups which ranged in number from 30 to 40. The instruction was given by a trained master student in psychology. During testing the participants were supervised by the class teacher and two other master students. Initially (Time 1), they completed the 30 even-numbered items of SPM in the first week of November 2017. One and a half years later (Time 2), the children were retested on the 30 odd-numbered items of SPM in the last week of May 2019.

Statistical analysis

Following previous studies that have captured the item-position effect underlying Raven's Matrices (e.g., Lozano, 2015; Ren et al., 2012, 2014), 10 composites were calculated by adding the scores of three neighboring items of SPM. These composites were used as manifest variables in the models.

Several CFA models were constructed for investigating the underlying structure of SPM. First, there were a couple of two-factor models including two latent variables that were not allowed to correlate with each other. These models included a latent variable representing the core of fluid intelligence. It was denoted the ability-specific latent variable and showed factor loadings of equal size. The other latent variable represented the item-position effect identified by either linearly or quadratically increasing factor loadings giving rise to two different models (Ren et al., 2014). It was referred to as the item-position latent variable. Second, two types of one-factor models comprising a single latent variable denoted fluid reasoning ability (with either freely estimated or fixed factor loadings) were specified for model comparisons. Eliminating the position-specific factor from the two-factor model leaves the one-factor model with constrained factor loadings. We consider a better model fit for the two-factor model than for the one-factor model as indication of the existence of the item-position effect.

Next, the scaled variance parameters of the latent variables were computed (Schweizer et al., 2019). This enabled comparisons among the estimates of the variance parameters, and therefore, interpretations regarding the relative importance of latent variables. Fourth, multiple group analysis was conducted to investigate whether there were substantial differences between the variances of the ability-specific and item-position factors of SPM across two time points.

Furthermore, the ability-specific and item-position components of SPM at two time points were linked to each other so that the stability of the two components over time could be examined. We also examined whether the correlation between the ability-specific components across time was statistically different from that between the item-position components.

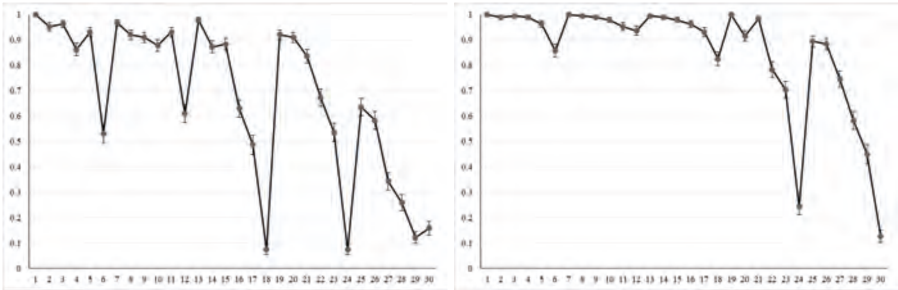
Maximum likelihood method was used to estimate the parameters by means of LISREL 8.8 (Jöreskog & Sörbom, 2006). The fit statistics were evaluated using the criteria recommended by DiStefano (2016) and Kline (2005). The model fit was considered good (or acceptable) if normed $\chi 2$ (= $\chi 2/df$) ≤ 2 (3), RMSEA ≤ .05 (.08), SRMR ≤ .05 (.10), and CFI ≥ .95 (.90). In addition, comparisons between the non-nested models were performed by means of the Akaike's information criterion (AIC). A model with a smaller AIC indicated a better fit.

Results

Descriptive statistics

The mean sum scores of correct responses in SPM were 20.43 (SD = 3.09) at Time 1, and 25.64 (SD = 2.42) at Time 2. Paired sample t-test was conducted to examine the difference in SPM scores between Times 1 and 2. The results revealed that children performed significantly better at Time 2 than Time 1 ($t = 22.65, p < .001$), suggesting a substantial development of fluid intelligence as children grew up. There was only a medium correlation between the sum scores of SPM at Times 1 and 2 ($r = .36, p < .001$). Furthermore, accuracy on individual items of SPM showed different patterns at Times 1 and 2. As Figure 1 illustrates, while the accuracy of items within each set (every six items) showed a decreasing trend at both time points, the participants performed much better on the first three sets at Time 2 than Time 1.

Figure 1 Means and standard errors for each item of SPM at times 1 (left) and 2 (right)



The item-position effect in SPM across two time points

To examine whether there was the suspected item-position effect underlying SPM, a series of two-factor models including both the ability-specific and item-position components were constructed. Table 1 presents the fit results for the first wave of data. When all 10 composites were included in the CFA model as manifest variables, the ability-position models showed acceptable fit results according to the normed χ^2, RMSEA, and SRMR, but no acceptable results for CFI. The bad model fit might be due to the ceiling effect underlying first several composites. That is, the items included in those composites were solved correctly by most participants, and thus showed extremely small variances. Therefore, we retested the ability-position models by eliminating either the first composite or the first two composites. After doing so, the ability-position models still described the data poorly according to CFI. Since the first as well as the third composites showed the smallest variances, these two composites were eliminated. Proceeding with the reduced set of composites, the model with quadratically increasing factor loadings showed acceptable fit on all fit statistics, while the model with linearly increasing loadings was not. Based on the same eight composites, the one-factor models with either freely estimated or constant factor loadings were constructed. The model fit for these two models was not acceptable and substantially worse than the corresponding two-factor model. Taken together, the ability-position model with quadratically increasing loadings provided the best description of the underlying structure of SPM. Furthermore, the variances of the latent variables in this model reached statistical significance (ability-specific: $Z = 3.68$, $p < .001$; item-position: $Z = 3.52$, $p < .001$), suggesting that both the ability-specific and item-position components represented important sources of variance.

The same models were constructed for the second wave of data. Table 2 shows the fit statistics for these models. Results indicated that the ability-position models with 10, 9, or 8 (3-10) composites showed no acceptable CFI results. Again, after eliminating the first and third composites, both models with linearly and quadratically increasing factor loadings turned out to be acceptable. Further evidence was provided by comparing the one-factor and two-fac-

Table 1 Fit Statistics of the Measurement Models of SPM for the First Wave of Data (N = 189).

Model	χ^2	df	χ^2/df	RMSEA	SRMR	CFI	AIC
Ability-position (composites 1-10)							
Linear increase	73.44	43	1.71	.061	.079	.821	97.44
Quadratic increase	70.72	43	1.64	.059	.080	.837	94.72
Ability-position (composites 2-10)							
Linear increase	61.27	34	1.80	.065	.079	.840	83.27
Quadratic increase	57.25	34	1.68	.060	.078	.865	79.25
Ability-position (composites 3-10)							
Linear increase	49.79	26	1.92	.070	.079	.853	69.79
Quadratic increase	47.06	26	1.81	.066	.078	.871	67.06
Ability-position (composites 2, 4-10)							
Linear increase	38.57	26	1.48	.051	.070	.895	58.57
Quadratic increase	37.68	26	1.45	.049	.070	.904	57.68
One-factor (composites 2, 4-10)							
Freely estimated loadings	37.55	20	1.88	.068	.064	.855	69.55
Constant loadings	50.45	27	1.87	.068	.085	.799	68.45

tor models. The corresponding one-factor model with constant factor loadings was not acceptable. Although the one-factor model with freely estimated factor loadings showed good fit, its AIC was slightly larger than the ability-position model with linearly increasing loadings. Taken together, this ability-position model provided the best account of the SPM data collected at Time 2. Furthermore, the variances of the latent variables in this model were also significant (ability-specific: $Z = 3.68$, $p < .001$; item-position: $Z = 3.52$, $p < .001$), suggesting that both the ability-specific and item-position components were crucial in accounting for the variances of SPM. Overall, the item-position component was identified in both SPM datasets.

A further question was whether the variances of the ability-specific and item-position components changed as children grew up. Two approaches were used to answer this question. First, we computed the scaled variances of the latent variables in the two-factor model with best model fit (cf. Schweizer et al., 2019). As for the first wave of data, it revealed values of 0.436 for the ability-specific factor and 0.404 for the item-position factor. The ability-spe-

Table 2 Fit Statistics of the Measurement Models of SPM for the Second Wave of Data (N = 189).

Model	χ^2	df	χ^2/df	RMSEA	SRMR	CFI	AIC
Ability-position (composites 1-10)							
Linear increase	62.46	43	1.45	.049	.075	.870	86.46
Quadratic increase	70.21	43	1.63	.058	.082	.832	94.22
Ability-position (composites 2-10)							
Linear increase	55.18	34	1.62	.058	.079	.871	77.18
Quadratic increase	60.42	34	1.78	.064	.084	.846	82.42
Ability-position (composites 3-10)							
Linear increase	47.91	26	1.84	.067	.082	.860	67.91
Quadratic increase	53.99	26	2.08	.076	.091	.822	73.99
Ability-position (composites 2, 4-10)							
Linear increase	33.64	26	1.41	.040	.071	.948	53.64
Quadratic increase	39.24	26	1.51	.052	.078	.945	59.24
One-factor (composites 2, 4-10)							
Freely estimated loadings	21.79	20	1.09	.022	.047	.987	53.79
Constant loadings	53.51	27	1.98	.072	.099	.846	71.51

cific and the item-position latent variables accounted for 51.90% and 48.10% of latent variance, respectively. Regarding the second wave of data, the scaled variances of the two latent variables were .208 and .537, indicating that the ability-specific and item-position latent variables accounted for 27.92% and 72.08% of latent variance, respectively. This result revealed that the item-position effect accounted for more variances of SPM as children grew up.

Second, multiple-group analysis was conducted to examine whether the variances of the two latent variables were invariant or not. In a first step, we constructed a model with no restrictions placed on parameters among the two waves of data. This model served as a baseline against which the fit of a more restricted model was compared. The fit statistics of this model showed an acceptable fit to the observed data, χ^2 (52) = 71.32, RMSEA = .044, SRMR = .071, CFI = .931, AIC = 111.32. Then, this model was modified by constraining the variances of the ability-specific latent variable to be equal across two time points. The fit of this restricted model turned out to be good, χ^2 (53) = 71.52, RMSEA = .043, SRMR = .071, CFI = .933, AIC = 109.52. The restriction did not worsen the model fit, as indicated by a non-significant model difference, $\Delta\chi^2$ (1) = .20, $p > .05$. However, after the variance of the item-position latent variable was constrained to be equal, the model became unacceptable, χ^2 (53) = 90.56, RMSEA = .061, SRMR = .098, CFI = .871, AIC = 128.56, and the restriction deteriorated the model fit, as indicated by a significant model difference, $\Delta\chi^2$ (1) = 19.24, $p < .001$. This result suggested that the item-position effect substantially changed as fluid intelligence developed.

Relationship between the ability-specific and position-effect components across time

Based on the measurement models used for investigating SPM at Times 1 and 2, we combined them into a comprehensive model, in which the two ability-specific latent variables and the two item-position latent variables were linked to each other (see Figure 2). This model showed an acceptable model fit, χ^2 (114) = 132.96, RMSEA = .030, SRMR = .072, CFI = .941, AIC = 176.96. There was a significant correlation between the ability-specific components across two time points (r = .88, t = 3.46, $p < .001$) whereas the correlation between the item-position components was not significant (r = .34, t = 1.68, $p > .05$). Assuming that the correlations might be affected by the change in the variances of the latent variables over time, the correlations were estimated individually but not simultaneously. The correlation between the ability-specific components was .99 (t = 4.56, $p < .001$), while the correlation between the item-position components was .61 (t = 3.53, $p < .001$). Furthermore, constraining the two correlations to be equal worsened the model fit, $\Delta\chi^2$ (1) = 19.56, $p < .001$. Taken together, these results suggested that as age increased, the ability component remained relatively stable but the item-position component altered substantially.

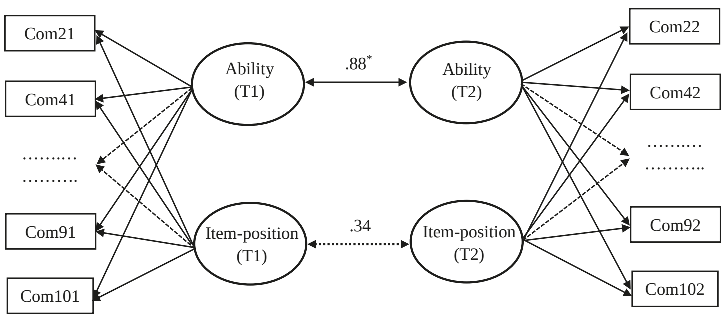


Figure 2 The relationship between the ability and the item-position components of SPM across two time points (Comi1 = the i-th composite score of SPM at time point 1, Comi2 = the i-th composite score of SPM at time point 2, T1 = Time 1, T2 = Time 2, *p < .05).

Discussion

Starting from the question whether the item-position effect can be observed in the fluid intelligence measure for children and how this effect develops as children grow up, longitudinal data collected at two time pointes were investigated by means of fixed-links modeling. The impact of the item-position latent variable on the fit statistics suggests that modeling the item-position effect is essential for achieving good model fit. The comparison of the scaled variances of latent variables showed that the item-position effect accounted for larger proportions of variance in SPM scores at Time 2 than Time 1. Furthermore, while the ability-specific latent variables across two time points were substantially related, the correlation between the item-position effect latent variables was insignificant, suggesting a remarkable change of the item-position effect in contrast to the stability of ability over time.

The superior model fit of the ability-position model indicates the presence of the item-position effect in intelligence testing on children. This result replicates the outcome of the first attempt to examine the item-position effect in children (Sun, Schweizer, & Ren, 2019) and confirms that SPM scores are influenced by the position effect besides ability. Sun et al. (2019) firstly investigated whether there was an age difference in the emergence of the item-position effect in fluid intelligence testing. They compared the ability-position models among primary school age children (7-8 years old) and secondary school age adolescents (12-13 years old). While the ability-position model showed good fit, the variance of the item-position factor was only significant in the older group but not in the younger group. Since the ability to learn and deliberately use complex rules has been proposed as main source of the item-position effect (Ren et al., 2014; Carlstedt, Gustafsson, & Ullstadius, 2000), they ascribe the deficiency of the item-position effect in younger children to the insufficient development of their learning abilities. When their learning ability increases, the item-position effect emerges as a source of systematic variation captured in fluid intelligence testing. Our study lowers the emergence age of the item-position effect from 12-13 years old to 10-11 years old and firstly confirmed the presence of the item-position effect in intelligence testing by the longitudinal data.

The relationship between the item-position effect and learning suggests that working memory may play a major role in the developmental changes regarding the item-position effect. Learning has been found to be closely related to working memory (e.g., Ren, Schweizer, Wang, & Gong, 2017; Wang, Ren, Altmeyer, & Schweizer, 2013; Wang, Ren, & Schweizer, 2015; Wang, Ren, & Schweizer, 2017; Wang, Ren, & Schweizer, 2019; Zeller, Wang, Reiss, & Schweizer, 2017). For instance, Wang et al. (2015) revealed that the ability to hold information temporarily and the shifting process play essential roles in acquiring new rules for categorization. According to developmental research, working memory undergoes rapid development during primary school years (Gathercole, Pickering, Ambridge, & Wearing, 2004; Xu et al., 2014). Therefore, the increase of working memory may underlie the emergency of the item-position effect in intelligence testing, which should be investigated in future research.

The scaling of the estimates of the variance parameters provides further information regarding the relative importance of the item-position effect compared with ability. Without scaling, the variances are not comparable. Scaled variance results indicate that nearly half of the variances in SPM scores are explained by the item-position effect. This result suggests that the ability to learn and spontaneously use abstract rules in the following items of SPM is as equally important as the ability in completing intelligence measures. Furthermore, the increasing importance of the item-position effect together with the weak correlation between the item-position factors at Times 1 and 2 suggest that the development of fluid intelligence might be driven mainly by the development of learning ability. In contrast, ability remains relatively stable over a relatively short period (i.e., one and a half years).

Our results have implications for the measurement of fluid intelligence. The current study together with previous studies focusing on adults suggests that fluid intelligence measures stimulate heterogeneous processes that cannot be explained by a single latent variable in the CFA model. It is consistent with the observation that there is a high degree of complexity with the underlying processes which are necessary for completing the items of intelligence measures (Schweizer, 1998; Stankov, 2000). Thus, more complex models that consider the item-position effect seem to be essential for investigating fluid intelligence and its relationship with the other cognitive abilities appropriately.

Methodologically, the current study demonstrates the values of using variance parameters to test a series of substantive research questions. The fixed-links models enable the isolation of method effects, e.g., the item-position effect and speed effect (Schweizer, Reiß, et al., 2019), from genuine construct. With statistical tests of the variance parameters of the latent variables, the significance of the item-position factor can be verified. With the scaled estimates of the variance parameters, the relative importance of the ability-specific and item-position components can be compared either within a single age group or across different age groups.

References

Carlstedt, B., Gustafsson, J. E., & Ullstadius, E. (2000). Item sequencing effects on the measurement of fluid intelligence. *Intelligence, 28*, 145–160. doi:10.1016/S0160-2896(00)00034-9

Campbell, D. T., & Mohr, P. J. (1950). The effect of ordinal position upon responses to items in a check list. *Journal of Applied Psychology, 34*, 62-67.

DiStefano, C. (2016). Examining fit with structural equation models. In K. Schweizer, & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 166-193). Göttingen, Germany: Hogrefe.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology, 40*, 177. doi: 10.1037/0012-1649.40.2.177

Hartig, J., Hölzel, B., & Moosbrugger, H. (2007). A confirmatory analysis of item reliability trends (CAIRT): Differentiating true score and error variance in the analysis of item context effects. *Multivariate Behavioral Research*, *42*(1), 157–183.

Jöreskog, K. G. (1970). A general method for analysis of covariance structure. *Biometrika, 57*, 239-257. doi: 10.2307/2334833

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109-133. doi: 10.1007/BF02291393

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, *55*(2), 312–320. doi:10.1037/0022-3514.55.2.312

Lozano, J. H. (2015). Are impulsivity and intelligence truly related constructs? Evidence based on the fixed-links model. *Personality and Individual Differences, 85*, 192–198. doi:10.1016/j.paid.2015.04.049

McDonald, R. P. (1965). Difficulty factors and nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology, 18*, 11-23. doi:10.1111/j.2044-8317.1965.tb00690.x

Mollenkopf, W. G. (1950). An experimental study of the effects on item-analysis data of changing item placement and test time limit. *Psychometrika*, *15*(3), 291–315. doi:10.1007/BF02289044

Raven, J. C., Raven, J., & Court, J. H. (1997). *Raven's progressive matrices and vocabulary scales*. Edinburgh: J.C. Raven Ltd.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales. Section 3, The standard progressive matrices*. Oxford, UK: Oxford Psychologists Press.

Ren, X., Goldhammer, F., Moosbrugger, H., & Schweizer, K. (2012). How does attention relate to the ability-specific and position-specific components of reasoning measured by APM?. *Learning and Individual Differences, 22*, 1-7. Doi: 10.1016/j.lindif.2011.09.009

Ren, X., Schweizer, K., Wang, T., Chu, P., & Gong, Q. (2017). On the relationship between executive functions of working memory and components derived from fluid intelligence measures. *Acta Psychologica, 180*, 79-87. doi: 10.1016/j.actpsy.2017.09.002

Ren, X., Wang, T., Altmeyer, M., & Schweizer, K. (2014). A learning-based account of fluid intelligence from the perspective of the position effect. *Learning and Individual Differences, 31*, 30–35. doi:10.1016/j.lindif.2014.01.002

Schweizer, K. (1998). Complexity of information processing and the speed-ability relationship. *The Journal of General Psychology, 125,* 89-102. doi:10.1080/00221309809595578

Schweizer, K. (2008). Investigating experimental effects within the framework of structural equation modeling: an example with effects on both error scores and reaction times. *Structural Equation Modeling, 15*, 327-345. doi: 10.1080/10705510801922621

Schweizer, K. (2011). Scaling variances of latent variables by standardizing loadings: applications to working memory and the position effect. *Multivariate Behavioral Research, 46*, 938-955. doi: 10.1080/00273171.2011.625312

Schweizer, K., Reiß, S., Ren, X., Wang, T., & Troche, S. (2019). Speed effect analysis using the CFA framework. *Frontiers in Psychology* (Section Quantitative Psychology and Measurement)*, 10*, ArtID 239. doi: 10.3389/fpsyg.2019.00239

Schweizer, K., Schreiner, M., & Gold, A. (2009). The confirmatory investigation of APM items with loadings as a function of the position and easiness of items: A two-dimensional model of APM. *Psychology Science Quarterly*, *51*, 47–64.

Schweizer, K., Troche, S., & DiStefano, C. (2019). Scaling the variance of a latent variable while assuring constancy of the model. *Frontiers in Psychology* (Section Quantitative Psychology and Measurement), *10*, ArtID 887. doi: 10.3389/fpsyg.2019.00887

Schweizer, K., Troche, S., & Reiß, S. (2017). Can variances of latent variables be scaled in such a way that they correspond to eigenvalues? *International Journal of Statistics and Probability, 6*, 35-42. doi: 10.5539/ijsp.v6n6p35

Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence, 28*, 121–143. doi:10.1016/S0160-2896(99)00033-1

Sun, S., Schweizer, K., & Ren, X. (2019). Item-position effect in Raven's Matrices: A developmental perspective, *Journal of Cognition and Development, 20*, 370-379, doi: 10.1080/15248372.2019.1581205

Wang, T., Ren, X., Altmeyer, M., & Schweizer, K. (2013). An account of the relationship between fluid intelligence and complex learning in considering storage capacity and executive attention. *Intelligence, 41*, 537-545. doi: 10.1016/j.intell.2013.07.008

Wang, T., Ren, X., & Schweizer, K. (2015). The contribution of temporary storage and executive processes to category learning. *Acta psychologica, 160*, 88–94. doi: 10.1016/j.actpsy.2015.07.003

Wang, T., Ren, X., & Schweizer, K. (2017). Learning and retrieval processes predict fluid intelligence over and above working memory. *Intelligence, 61*, 29-36. doi: 10.1016/j.intell.2016.12.005

Wang, T., Schweizer, K., & Ren, X. (2019). Executive Control in Learning: Evidence for the Dissociation of Rule Learning and Associative Learning. *Advances in Cognitive Psychology, 15*, 41–51. doi: 10.5709/acp-0255-8

Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PloS one, 8*, e77770. doi: 10.1371/journal.pone.0077770

Zeller, F., Wang, T., Reiss, S., & Schweizer, K. (2017). Does the modality of measures influence the relationship among working memory, learning and fluid intelligence? *Personality and Individual Differences, 105*, 275–279. doi: 10.1016/j.paid.2016.10.013

Zeller, F., Krampen, D., Reiss, S., & Schweizer, K. (2017). Do adaptive representations of the item-position effect in APM improve model fit? a simulation study. *Educational and Psychological Measurement, 77*, 743-765. doi: 10.1177/0013164416654946

Author Note:

This research was supported by the Zhejiang Provincial Natural Science Foundation of China (No. LQ18C090001).

Corresponding author:

Qiong Zhang
Department of Psychology and Behavioral Sciences, Zhejiang University
Tianmushan Road 148
310028 Hangzhou, China
zhangqiongzgh@zju.edu.cn