Psychological Test and Assessment Modeling, Volume 55, 2013 (1), 62-78

The position effect in tests with a time limit: the consideration of interruption and working speed

Karl Schweizer[1] *& Xuezhu Ren*[2]

Abstract

The position effect is a possible source of impairment of the structural validity of a test concerning model fit. In the case of tests with a time limit there is even a complication of the situation because of a decreasing number of participants completing the last few items of the test. Therefore, it is assumed that the appropriate representation of the position effect must additionally consider interruption due to the time limit and the effect of working speed. Interruption can be represented by the same latent variable as the position effect whereas the contribution of working speed requires another one. Confirmatory factor models including a representation of the position effect as a linear, quadratic or logarithmic increase were compared with models additionally considering interruption as a logistic decrease or simply as immediate interruption. Furthermore, there were models additionally considering working speed. In the sample of 305 participants the investigation of probability-based covariances made apparent that the modeling of interruption and also working speed substantially improved model fit. The best-fitting model was characterized by a linearly increasing representation of the position effect combined with a logistic decrease in the more difficult items and a contribution due to working speed.

Key words: position effect, confirmatory factor analysis, tau-equivalent model, method effect

[1] *Correspondence concerning this article should be addressed to:* Karl Schweizer, PhD, Department of Psychology, Goethe University Frankfurt, Mertonstr 17, 60054 Frankfurt a. M., Germany; email: K.Schweizer@psych.uni-frankfurt.de

[2] Goethe University Frankfurt, Frankfurt a. M., Germany

The investigation of the position effect started in the 50s (Campbell & Mohr, 1950). Since that time a position effect has been observed in the items of many personality and ability measures. The position effect describes the dependency of the responses to the items on the responses to items that have been processed immediately before. The position effect is especially obvious in the change of item reliability that was found to increase from the first to last items of a scale (Knowles, 1988). Since this effect is observable whenever items representing the same ability or trait must be processed successively, it is considered by some researchers as a source of method variance (Spector, 2006; Spector & Brannick, 2010). What so far has not been considered in this research are the consequences of a time limit to test application. Since a position effect can only be expected to occur when test takers actively process items, a time limit may impair or even eliminate the position effect and individual differences in working speed may become apparent instead.

In the late 80s the position effect came into the focus of IRT research. Since that time there have accumulated a number of studies demonstrating that it is possible to identify a position effect by means of IRT models (e.g., Embretson, 1991; Gittler & Wild, 1989; Hohensinn, Kubinger, Reif, Holocher-Ertl, Khorramdel, & Frebort, 2008; Hohensinn, Kubinger, Reif, Schleich, & Khorramdel, 2011; Kubinger, 2003; Kubinger, Formann, & Farkas, 1991; Verguts & De Boeck, 2000). The models were mostly either multidimensional Rasch models or linear logistic test models. In the multidimensional Rasch models the ability or trait on one hand and the position effect on the other hand are represented independently of each other whereas in the linear logistic test models one dimension serves the simultaneous representation of the ability or trait and the position effect. A specificity of the linear logistic test models (LLTM) is the modeling of the position effect by appropriately selected numbers (e.g., Kubinger, 2003). Within the IRT approach of investigating the position effect a specific source of the effect was proposed: learning as the result of becoming familiar with completing the items (Embretson, 1991; Verguts & De Boeck, 2000).

An even more recent development is the investigation of the position effect in the framework of confirmatory factor analysis (CFA) (Hartig, Hölzl, & Moosbrugger, 2007; Ren, Goldhammer, Moosbrugger, & Schweizer, 2012; Schweizer, 2012a; Schweizer, Schreiner, & Gold, 2009; Schweizer, Troche, & Rammsayer, 2011). Some of the applied factor-analytic methods included variants of the tau-equivalent model of measurement (Graham, 2006; Lord & Novick, 1968, p. 50) for representing the position effect. The advantage of this representation is that the course of the position effect needs to be described rather precisely. As a consequence, it is possible to consider different courses and to compare them with each other.

Since confirmatory factor analysis (CFA) serves the investigation of the dimensionality of a covariance matrix (Jöreskog, 1970), using the factor-analytic approach for investigating the position effect means focusing on the need for a second latent variable besides the latent variable for representing the ability or trait of interest. It is necessary to show that the ability respectively trait latent variable is insufficient for accounting for the observed covariances alone. In this characteristic the CFA approach fundamentally differs from the LLTM that highlights the comparison of different orderings of items.

The available results suggest the representation of the course of the position effect by an increasing function. However, there are differences concerning the details. In some cases a quadratically increasing curve proved to be most appropriate, (e.g., Schweizer, Schreiner, & Gold, 2009) whereas in others the linearly increasing curve was the favorable one (e.g., Schweizer, Troche, & Rammsayer, 2011). Furthermore, there was even an increase in model fit due to the omission of the first few and last few items of a test because of ceiling and floor effects on those items (Schweizer, 2012a). Obviously, a general way of representing the position effect has so far not been identified if there is any.

In order to rule out the possibility that the representation of the position effect erroneously reflects the item contents leading to different degrees of difficulty, there is an adjustment of the representation of the variances and covariances that takes the different degrees of difficulty into account. This adjustment is in line with the generalized linear model (McCullagh & Nelder, 1985; Nelder & Wedderburn, 1972) that relates variables following different distributions and is appropriate for considering the effects of the different degrees of difficulty on variances and covariances. It enables the consideration of the weighted version of the tau-equivalent model of measurement in investigating probability-based covariances obtained from binary data (Schweizer, 2012b). Furthermore, a general justification of investigating binary data by means of factor-analytic methods is provided by Takane and de Leeuwe (1987).

A possible reason for the variety of representations of the position effect is the difference between power and speed tests. In completing the items of power tests each test taker has the opportunity to invest as much time as necessary into completing the items and, thus, to reach the best result possible for the own person. In contrast, in completing speed tests almost all the test takers have to stop completing items before they reach their best result possible. However, stopping early is not the problem with speed tests in itself. The real problem is that the test takers may differ according to their working speed and, as a consequence, differ in the degree of approaching their best result possible. Consequently, there is reason for expecting that the position effect is observable in all the items of a power test but may be restricted in the items of a speed test. In tests with a time limit working speed may become apparent as another source of performance.

This paper serves the study of the representation of the position effect in data obtained by means of a speed test assessing the reasoning ability. Various confirmatory factor models for representing the hypothesized effects besides the position effect are proposed and compared with each other. There are models with the suspected reduction of the position effect near to the end of the sequence of items because of the limitation of processing time. Furthermore, some models additionally reflect the hypothesis that there will be an independent contribution of working speed, which becomes apparent in the section of a sequence of items where the position effect is vanishing.

The work presented in this paper continues research work of the CFA tradition in order to make a method available that can be expected to be a bit less demanding on the sample size than corresponding IRT models. Another reason is the separation of the source of the position effect from the assessed ability for investigating aspects of validity. However, in the light of the large commonality of the traditions leading Mellenbergh (1994) and others to argue that CFA models are members of the family of generalized IRT models there is no reason for overemphasizing the difference between the traditions.

The framework for the representation of the position effect

The tau-equivalent model of measurement (Graham, 2006) provides the outset for the representation of the position effect. According to Bollen (1989) such a model of measurement is composed of a testable set of equations, one for each manifest variable. Since it is cumbersome to address the details in using vectors, in the following the characteristics of the model of measurement are described with respect to one of these equations only. According to the tau-equivalent model the ith manifest random variable Y_i (i = 1, …, p) where i identifies the variable taken from the set of p manifest random variables is composed of a true part and an error part. Furthermore, the true part is considered as the latent intercept μ and the person-specific product composed of the factor loading λ_τ and the latent random variable η that represents the person attribute of interest. Moreover, there is the error random variable ε_i (i = 1, …, p). The parts of this model are added up so that

$$Y_i = \mu_i + \lambda_\tau \eta + \varepsilon_i \, . \tag{1}$$

Please note that the notation by Graham (2006) omits the indication of the participant that is frequently considered in model equations (Rasch, Kubinger, & Yanagida, 2011, p. 28).

The investigation of the position effect in considering the time limit demands three modifications of Equation 1: (1) the representation of the position effect, (2) the representation of the effect of working speed and (3) a link function for relating the distribution of the latent variables to the distribution of the manifest variables. The representation of the position effect is achieved by another product composed of the factor loading λ_i (i = 1, …, p), where i identifies the variable taken from the set of p manifest random variables, and the latent random variable η_p that represents the impact of the position effect. Since the theoretical considerations concerning the position effect, as they are presented in the introductory section, suggest a source of responding that reflects working speed, it is reasonable to consider a further product of a factor loading that is λ_i^s (i = 1, …, p) and a latent random variable for representing working speed that is the supplementary variable η_s. Adding these products to Equation 1 and considering weight w_i (i = 1, …, p) serving as a link function gives

$$Y_i = \mu_i + w_i \lambda_\tau \eta + w_i \lambda_i \eta_p + w_i \lambda_i^s \eta_s + \varepsilon_i \tag{2}$$

where η, η_p and η_s represent latent variables. The weight is described in more detail in the method section. Since in this model the factor loadings are constrained, the variances of the latent variables are estimated instead. Combinations of fixed factor loadings and free variances have served well in the fixed-links model approach (Schweizer, 2008, 2009).

Functions for describing the position effect

The various reports of the position effect suggest its description by a gradually increasing curve. However, the exact course of the increase is not known and may even vary depending on the underlying processes and the condition of testing. Therefore, it is necessary to consider different types of increase in an attempt of identifying the most appropriate representation. Three functions *f* are considered for this purpose in this paper.

At first, we discuss a model that represents the position effect by a linear function. This function is suggested by the principle of economy. Since the linear function can produce rather large numbers, it is reasonable to define the function f_l in such a way that the predicted numbers do not surmount one:

$$f_l(i) = (i-1)/(p-1) \tag{3}$$

where i ($i = 1, \ldots, p$) corresponds to the position of the individual item in the sequence of p items of the measure.

Second, we consider a function that can represent the case of an effect characterized by a gradually increasing slope. The increase may be minor in the beginning and speed up subsequently. For example, if complex learning is underlying the position effect, it can be argued that some small achievements regarding different aspects of cognitive information processing must come together before a considerable improvement of performance is finally possible. This possibility can be represented by the quadratically increasing function f_q:

$$f_q(i) = (i-1)^2/(p-1)^2. \tag{4}$$

Third, we take into consideration an increasing function with a slope that is gradually decreasing since there is the possibility that the effect is large in the beginning in the sense of a steep slope followed by a decreasing degree of steepness. Such a course is reasonable in the case of an upper limit, for example because of the limited range of the scale. The initially steep slope is large but is gradually reduced in order to stay below this upper limit. Such a decrease in slope can be represented by the function f_{ln}:

$$f_{ln}(i) = \ln(i)/\ln(p). \tag{5}$$

This function is addressed as logarithmic function. A major characteristic of each one of the three functions is the variation between zero and one.

Before concluding this section on functions it needs to be stated how the functions are to be integrated into Equation 2. The functions define the loading λ_i on the latent variable representing the impact of the position effect:

$$\lambda_i = f(i). \tag{6}$$

The interruption of the effect

The representations of the position effect according to Equations 3 to 5 imply the assumption that there is no specific limitation to this effect. However, as already indicated in the introductory section, speed tests may restrict the effect to a subset of items. The restriction due to testing time means that the position effect can simply not occur or only to a minor degree in the items arranged near to the end of the sequence of items. There is not enough time for completing these items properly. It appears to be reasonable to consider different ways to manage the interruption of the position effect.

The first and easiest way of representing interruption is stopping the representation of the position effect at a specific point. This means that it is necessary to determine the number j $\in$ {2, …, p-1} that separates the part of the sequence of items with a position effect from the other part. Position effect and interruption are represented by function g such that

$$g(i) = \begin{cases} f(i) & \text{for} \quad i \leq \mathrm{j} \\ 0 & \text{else} \end{cases} \tag{7}$$

where f is the function representing the position effect and i (i = 1, …, p) indicates the position of the item.

The second way assumes that the influence of the position effect is virtually constant in the beginning but decreases considerably near the jth item and fades out afterwards. Such a course is described by function h_j that includes the logistic function:

$$h_j(i) = 1 - \frac{e^{i-j}}{1+e^{i-j}}. \tag{8}$$

Function h_j is close to one in the beginning, shows the strongest decrease near j and is approaching zero in the end. If this specific function is multiplied with another function, in most items positioned in front of the jth item the outcome reflects the other function and approaches zero in the items following the jth item. So it is well suited for multiplication with the function for representing the position effect,

$$g(i) = f(i) \times h_j(i), \tag{9}$$

in order to achieve a smooth passage from the part with the effect to the part without the effect. In each one of these ways subscript j plays a major role. Since the course of the change is not exactly known, it is necessary to search for the number j $\in$ {2, …, p-1} that optimizes model fit. This type of interruption is denoted logistic decrease.

In the case of the additional representation of interruption the factor loading λ_i on the latent variable representing the impact of the position effect is defined as

$$\lambda_i = g(i). \tag{10}$$

The contribution of working speed

In this section the consequences of limiting the time of testing are considered from the perspective of working speed. Because of individual differences in working speed the test takers can be expected to complete different numbers of items before the time limit is reached. These differences in working speed enable some test takers to come closer to their best result possible within the available time span than other test takers. So there is another potential source of true variance that can be assumed to be especially effective in the items of the second half of the sequence of items, and in a way can be considered as a replacement of the position effect.

Since working speed reflects the test taker's performance style but rather not ability, there is reason for assuming that it is independent of the source giving rise to the position effect and being related to learning (Embretson, 1991; Verguts & De Boeck, 2000). The previous reasoning suggests that this additional source of performance is especially effective in the items near to the end of the sequence of items although it seems not to be possible to determine the area of effectivity exactly in beforehand. The function g_s selected for representing this source is the logistic function:

$$g_s(i) = \frac{e^{i-j}}{1+e^{i-j}} \tag{11}$$

where i indicates the position of the actual item and j the position of the item that separates the part of the sequence, which is mainly under the influence of working speed, from the other part. The selection of this function has implications: it is implicitly assumed that there is a subset of items showing a rather constant influence of working speed on performance. It needs to be added that this kind of representation also requires the identification of j in investigating the data.

Since the effect of working speed is assumed to be independent of the position effect, it gives rise to the last true component to Equation 2. This component is composed of the factor loading λ_i^s and the latent random variable η_s that represents differences in working speed. The factor loading is defined as follows:

$$\lambda_i^s = g_s(i) \; . \tag{12}$$

Objectives

Three major objectives were selected for this paper. The first one was the comparison of the basic ways of representing the position effect in a test with a time limit. These were the ways according to the functions (3) to (5). The second objective was the comparison of the ways of representing interruption in a test with a time limit. The functions (7) and (9) served the representation of interruption. The third one was the investigation of the contribution of working speed, as it is represented by Equation 11. These objectives required that the investigation was conducted on ability data that were collected with a limitation to testing time. Since reasoning measures played the most prominent role in the assessment of ability, reasoning data were selected for the study.

Method

Participants

Three hundred and five university students completed the reasoning measure. Whereas the psychology students received course credit for participation, a small amount of money was given to the other students.

Measure

Horn's (1983) LPS Reasoning Scale (LPS-4) was selected for the assessment of the participants' reasoning ability. This Scale included 40 items and the time allotted for completion was 8 minutes. It was an ability test with a restrictive time limit and was part of a test battery designed according to Thurstone's (1938) model of intelligence. This scale has been quite popular probably because it provided an estimate of the reasoning ability in a very economic way. Since the first half of items was very easy, it could be completed by every member of a high ability sample (i.e. a sample drawn from the population of university students) virtually without failure. Because of this characteristic the statistical investigation was restricted to the second half of items consisting of the 21st to 40th items although the responses to all the items were recorded.

Models

All the models for the investigations according to the first and second objectives included two latent and 20 manifest variables. One latent variable served the representation of the reasoning ability and the other one the representation of the position effect and also the interruption. Because of the third objective a third latent variable was additionally included. Furthermore, for having a comparison level a basic model was considered. It was the tau-equivalent model of measurement characterized by one latent variable only for representing the reasoning ability (Graham, 2006; Lord & Novick, 1968, p. 50). This basic model provided the outset for the construction of the models for representing the position effect.

The tau-equivalent model of measurement served the representation of the reasoning ability as part of all the more complex models. The loadings on this latent variable were constrained to equal sizes. This was achieved by setting all the loadings equal to one in combination with setting the variance of the corresponding latent variable free for estimation. The position effect was represented by numbers according to either the linear, quadratic or logarithmic function selected as constraints for the factor loadings on the second latent variable. For investigating the effect of interruption the numbers for representing the position effect were modified in considering the following options: immediate interruption (7) or logistic decrease (9). The numbers for constraining the factor loadings on the third latent variable were achieved by means of Equation 11. Since in the case of the second and third latent variables the loadings were also fixed, the corresponding variances were also set free for estimation. Furthermore, since the models were expected to decompose the variances of the items into a part that was due to ability and other parts, the first latent variable was not allowed to correlate with the other latent variable(s).

Finally, the link function for the computation of the weights of Equation 2 needed to be specified. A function that already worked well in investigating the position effect was selected (Ren et al., 2012; Schweizer, 2012b; Schweizer, Schreiner, & Gold, 2009). This function was based on the probability of the ith item X_i (i = 1, …, p) where the binary events were coded as 0 and 1. Accordingly, the weight w_i (i = 1, …, p) was defined as a function of $\Pr(X_i=1)$ such that

$$w_i = \sqrt{\Pr(X_i = 1)\left[1 - \Pr(X_i = 1)\right]}. \tag{13}$$

Statistical investigation

The statistical investigations were performed by means of the statistical software LISREL (Jöreskog & Sörbom, 2006). The fit statistics recommended by Kline (2005) served the evaluation of the models: normed χ^2 (= χ^2/df), RMSEA, SRMR, CFI, and also the limits proposed by Kline. Normed χ^2 and RMSEA were considered as the more important statistics in the evaluation of the models of this study. An acceptable degree of fit required a normed χ^2 smaller than 3 and a RMSEA smaller than .08 whereas the corresponding limits for a good degree of fit were 2 and .05 respectively. The limits for SRMR and CFI were .08 and .90 in corresponding order (Hu & Bentler, 1999). Furthermore, AIC was considered. Since the models were not nested with a few exceptions, the comparisons between models were mainly performed by means of AIC and CFI. Special emphasis was given to CFI since a CFI difference of .01 and larger was found to indicate a substantial difference (Cheung & Rensvold, 2002).

Probability-based covariances were investigated. Probability-based covariances are computed in the following way: in the first step we computed the probability of the selected response (e.g., the correct response) for each item. These probabilities were transformed into probability-based covariances in the second step. Let X_i and X_j be two binary items with corresponding probabilities $\Pr(X_i)$ and $\Pr(X_j)$ and 1 be the coding for the selected response in either case. Then the probability-based covariance $\mathrm{cov}(X_i, X_j)$ is given by

$$\mathrm{cov}(X_i, X_j) = \Pr(X_i = 1 \wedge X_j = 1) - \Pr(X_i = 1)\Pr(X_j = 1).$$

Deviations of item distributions from normality were not considered as a problem since the factor loadings were not estimated but predicted by means of constraints. The variances of the two or three latent variables were estimated instead. These variances referred to the whole set of 20 items showing a large range of arithmetic means. Because of the large range of these means with an overall mean of about .5 these variances could be assumed to be unbiased.

Results

Data description

The arithmetic means of the items (coded 1 for a correct response and 0 otherwise) varied between .04 and .94 and the standard deviations between .20 and .50. The mean of the scores obtained by summing up the 20 items was 11.63 and the standard deviation 3.74. The consistency of these items according to Cronbach's (1951) Alpha was 0.80.

The fit results for the models without a representation of the position effect

In order to be able to demonstrate that considering the position effect is an advantage, the weighted version of the tau-equivalent model of confirmatory factor analysis was applied to the reasoning data. A not so good degree of model fit (χ^2 = 487.96 (df = 189), χ^2/df = 2.58, RMSEA = .071, SRMR = .10, CFI = .70, AIC = 520.96) was observed for this model. It was only acceptable according to the normed χ^2 (χ^2/df) and RMSEA results.

The fit results for the models with a representation of the position effect

The fit results for the models with a representation of the position effect according to the functions (3) to (5) are presented in Table 1.

The fit results for the models assuming linear, quadratic or logarithmic effects are given in the first, second and third rows in corresponding order. The normed χ^2 (χ^2/df) and RMSEA results indicated an acceptable degree of model fit whereas the SRMR and CFI results signified a bad model fit. The model constructed according to the quadratic function showed the overall best degree of model fit. Furthermore, according to the CFI results there was virtually no difference between the models.

However, these models showed an advantage when compared with the original tau-equivalent model. The χ^2 were reduced by between 38 and 63, and the models with the

Table 1:
Fit Statistics of the Models According to the Three Functions for Representing the Position Effect (n = 305)

Type of representation	χ^2	df	χ^2/df	RMSEA	SRMR	CFI	AIC
Linear increase	437.08	188	2.32	.066	.096	.75	481.08
Quadratic increase	424.71	188	2.26	.064	.099	.75	468.71
Logarithmic increase	449.49	188	2.39	.068	.094	.74	493.49

representation of the position effect were clearly superior to the tau-equivalent model according to the AIC and CFI results. So even the model showing the worst degree of model fit substantially differed from the original tau-equivalent model (χ^2_{diff}=38.47, *df*=1, *p*<.05). The outcome of this chi-square difference test clearly justified the consideration of the second latent variable.

The fit results for the models considering the interruption of the effect

Since no one of the three representations of the position effect led to a really favorable result, each one of these representations was combined with each one of the two ways of representing interruption. The fit results for the six combinations of functions are presented in Table 2. These results were obtained for models that assumed a critical item position (j) of 14 (i.e., the 34th item in the whole sequence of items) in applying Equations 7 and 9. This critical position was selected because it minimized the χ^2.

The first and second rows give the results for the models assuming a linear position effect and also interruption, the third and fourth rows for the models assuming a quadratic position effect in combination with interruption, and the fifth and sixth rows for the models assuming a logarithmic position effect and interruption. The normed χ^2 (χ^2/df) and RMSEA results indicated an acceptable degree of model fit whereas the results for SRMR and CFI were unacceptable. The improvement in model fit due to the consideration of interruption was especially obvious in the χ^2: the improvement varied between 7 and 53. According to the AIC and CFI results the models assuming a linear increase combined with either a logistic interruption or an immediate interruption of the increase did substantially better than the other models.

Table 2:
Fit Statistics of the Models for Representing the Position Effect Combined with the Two Ways of Representing Interruption (n = 305) (j = *14*)

Way of considering interruption	χ^2	df	χ^2/df	RMSEA	SRMR	CFI	AIC
Linear increase							
Logistic decrease	387.17	188	2.06	.059	.091	.79	431.17
Immediate interruption	392.84	188	2.09	.060	.087	.79	436.84
Quadratic increase							
Logistic decrease	417.93	188	2.22	.063	.083	.77	461.93
Immediate interruption	403.36	188	2.14	.061	.090	.77	447.36
Logarithmic increase							
Logistic decrease	409.89	188	2.18	.062	.089	.77	453.89
Immediate interruption	396.07	188	2.11	.060	.095	.77	440.07

Furthermore, the results for these models were compared with the results for the tau-equivalent model. Differences of 95 and 100 χ^2 between best-fitting models and the original tau-equivalent model indicated a considerable improvement. These differences signified an improvement by between 19 and 20 percent.

The fit results for the models considering working speed

Next, the contribution of working speed was investigated. Working speed was considered in combination with all the models of the previous investigations. The fit results are presented in Table 3.

This Table shows the same structure as Table 2. The results were quite favorable: all the normed χ^2 (χ^2/df) and RMSEA results indicated a good degree of model fit. A good SRMR was found for all the models including logistic decrease. In contrast, the CFI results were still not good but came close to the boundary between acceptable and unacceptable results.

Furthermore, all the models were superior to the tau-equivalent model that served as comparison level. The differences in χ^2 between these and the original tau-equivalent models varied between 182 and 205. Also the AIC and CFI statistics signified a considerable improvement resulting from the consideration of working speed. Furthermore, the differences between the chi-squares of the best fitting model of Table 2 and the chi-squares of the models of Table 3 varied between 83 and 104. Even the smallest one of these differences was substantial according to the chi-square difference test (χ^2_{diff}=83.12, df=2, p<.05). So there was also a convincing justification for the consideration of the third latent variable.

Table 3:
Fit Statistics of the Models Including the Representation of Working Speed in Addition to the Representation of the Position Effect and Interruption (n = 305) (j = *15*)

Way of considering interruption	χ^2	df	χ^2/df	RMSEA	SRMR	CFI	AIC
Latent variables of working speed and linear increase							
Logistic decrease	282.88	186	1.52	.041	.077	.88	330.88
Immediate interruption	294.13	186	1.58	.044	.080	.87	342.13
Latent variables of working speed and quadratic increase							
Logistic decrease	304.05	186	1.63	.046	.080	.86	352.05
Immediate interruption	302.70	186	1.63	.045	.083	.86	350.70
Latent variables of working speed and logarithmic increase							
Logistic decrease	297.88	186	1.60	.044	.079	.87	345.88
Immediate interruption	305.60	186	1.64	.046	.084	.86	353.60

The characteristics of the best-fitting model

The model with a linear representation of the position effect in combination with logistic decrease and a representation of working speed was superior to all the other models according to normed χ^2 (χ^2/df), RMSEA, SRMR, CFI and AIC. The CFI results even indicated a substantial difference between this model and all the other models. Figure 1 enables the evaluation of the completely standardized loadings obtained from the constraints of the factor loadings on the three latent variables.

The curves are not completely smooth since the completely standardized factor loadings reflect the variation of the variances of the items additionally. The curves representing the position and working-speed effects intersect. It appears that the fading of the position effect gives way to the representation of working speed.

Furthermore, the scaled variances were computed for the latent variables (see Schweizer, 2011a) in order to be able to compare these sources according to their importance for performance. The variance of the ability-specific latent variable was .0104 ($t = 3.64$, $p < .01$), of the position-specific latent variable .0186 ($t = 6.59$, $p < .01$) and of the working-speed latent variable .0093 ($t = 6.37$, $p < .05$). Since there were zero loadings associated with the curve for working speed, the zero loadings were excluded (Schweizer, 2011b). This correction led to a variance of .0132 ($t = 6.37$, $p < .05$) for the working-speed latent variable. Taking the correlation of the second and third latent variable of .24 into consideration, this means that 25.5 percent of the variance at the latent level was due to ability, 43.57 percent due to the source of the position effect and 30.93 percent due to working speed. Apparently, position effect and working speed accounted for more variance than the ability-specific source of performance.

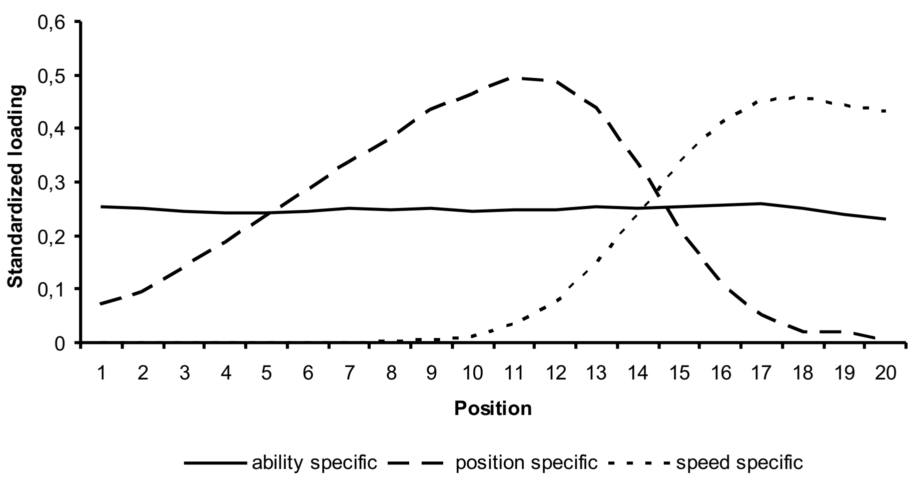

Figure 1:
Loading curves based on the completely standardized factor loadings constrained for representing ability, position effect and working-speed effect

Discussion

The position effect appears to be a typical method effect since it results from the arrangement of items (Knowles, 1988). Method effects must be taken seriously since Campbell and Fiske (1959) demonstrated that the true respectively systematic part of measurement does not only represent the ability or trait of interest but also characteristics of the measure. Furthermore, there is a lot of evidence of such method effects observed by means of confirmatory factor analysis of multitrait-multimethod data (e.g., Kenny & Kashy, 1992; Marsh, 1989; Marsh & Bailey, 1991; Widaman, 1985). Moreover, there are specific method effects like the item wording effect that influences model fit negatively and recently has received a lot of attention. Equal numbers of positively and negatively worded items were shown to impair model fit considerably (DiStefano & Motl, 2006; Vautier, Steyer, Jmel, & Raufaste, 2005). In this context the position effect appears to be just another method effect that reflects another characteristic of the observational situation: the test taker has to complete a series of rather similar items in a specific series.

In the evaluation of the outcomes of this study it is convenient to start with the comparison of the models based on the tau-equivalent model of measurement (Graham, 2006; Lord & Novick, 1968, p. 50) from the perspective of AIC. The consideration of the position effect led to an improvement of the model fit by about 8 percent. Adding the representation of the interruption increased the improvement to about 14 percent. Considering additionally the working speed raised the improvement to about 33 percent. Apparently, the working speed is the major source of the improvement of model fit when the data originate from a speed test. In sum, no one of the considered sources of responding seems to be dispensable in an attempt of achieving an acceptable or even good degree of model fit.

A shortcoming of the described approach of investigating data achieved by speed tests is the necessity to consider a number of models that differ according to the number representing the vanishing of the position effect that results from the limitation to the time for completing the items. The replacement of this set of models by one model including a procedure for determining the number representing the vanishing of the effect is another desirable aim. Further research is necessary for contriving such a procedure. Another problem is that there is no separation of errors and omissions. Such a separation is useful if the participants check only items which they seriously try to solve. However, if they check difficult items at random before the time runs out in order to increase the score by random hits, the separation of errors and omissions is flawed. Because of the limitation to the time of testing it is rather likely that some test takers proceeded in that way. So with respect to the test selected for the study it appeared not to be reasonable to separate errors and omissions.

Finally, the reasonableness of the new findings needs to be considered in the light of previous findings. Are the results of this study in agreement with the results of other studies? Unfortunately, IRT results cannot be considered since the assumed course of the position effect is not clear in the corresponding models. However, there are a few other studies including the application of reasoning measures that can be regarded as speed tests and of CFA. In these studies the representation of the position effect was restricted to a subset of items in order to achieve a sufficiently good degree of model fit (e.g., Schweizer, Troche, & Rammsayer, 2011). The restriction to a subset of items presumably excluded items which would otherwise heavily load on the latent variable representing working speed. Other studies reporting a good model fit without considering working speed included data obtained by power tests (e.g., Schweizer, Schreiner, & Gold, 2009). So there is reason for assuming acceptability.

Reference

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

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.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling, 13*, 440-464.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 56*, 495-515.

Fischer, G. H. (1972). *Conditional maximum-likelihood estimation of item parameters for a linear logistic test model* (Research Bulletin 9). Vienna: University of Vienna, Psychological Institute.

Gittler, G., & Wild, B. (1989). Der Einsatz des LLTM bei der Konstruktion des Itempools für das adaptive Testen [The use of LLTM for the construction of an item pool for adaptive testing]. In K. D. Kubinger (Ed.), *Moderne Testtheorie – Ein Abriß samt neuesten Beiträgen* (S. 115-139). München: PVU.

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement, 66*, 930-944.

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, 157-183.

Hohensinn, C., Kubinger, K. D. Reif, M., Holocher-Ertl, S., Khorramdel, L., & Frebort, M., (2008). Examining item-position effects in large-scale assessment using the Linear Logistic Test Model. *Psychology Science Quarterly, 50*, 391-402.

Hohensinn, C., Kubinger, K. D., Reif, M., Schleich, E., & Khorramdel, L. (2011). Analysing item position effects due to test booklet design within large-scale assessment. *Educational Research and Evaluation, 17*, 497-509.

Horn, W. (1983). *Performance-Test-System*. 2nd ed. Göttingen: Hogrefe.

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

Jöreskog, K.G., & Sörbom, D. (2006). *LISREL 8.80*. Lincolnwood, IL: Scientific Software International Inc.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112*, 165-172.

Kline, R. B. (2005). *Principles and practices 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*, 312-320.

Kubinger, K. D. (2003). On artificial results due to using factor analysis for dichotomous variables. *Psychology Science, 45*, 106-110.

Kubinger, K. D. (2009). Applications of the Linear Logistic Test Model in psychometric research. *Educational and Psychological Measurement, 69*, 232-244.

Kubinger, K. D., Formann, A. K., & Farkas, M. G. (1991). Psychometric shortcomings of Raven's Standard Progressive Matrices SPM) in particular for computerized testing. *European Review of Applied Psychology, 41*, 295-300.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MASS: Addison-Wesley.

Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: many problems and a few solutions. *Applied Psychological Measurement*, 13, 335-361.

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement*, 15, 47-70.

McCullagh, P., & Nelder, J. A. (1985). *Generalized linear models*. London: Chapman and Hall.

Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychological Bulletin, 115*, 300-307.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, A, 135*, 370-384.

Rasch, D., Kubinger, K. D., & Yanagida, T. (2011). *Statistics in Psychology. Using R and SPSS*. Chichester: John Wiley & Sons.

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.

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.

Schweizer, K. (2009). Fixed-links models for investigating experimental effects combined with processing strategies in repeated measures designs: a cognitive task as example. *British Journal of Mathematical and Statistical Psychology, 62*, 217-232.

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

Schweizer, K. (2011b). Probability-based and measurement-related hypotheses with full restriction for investigations by means of confirmatory factor analysis: an example from cognitive psychology. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 7*, 157-164.

Schweizer, K. (2012a). The position effect in reasoning items considered from the CFA perspective. *International Journal of Educational and Psychological Assessment, 11*, 44-58.

Schweizer, K. (2012b). A weighted version of the tau-equivalent model of measurement for items with ordered response categories. *International Journal of Statistics and Probability, 1*, 151-163. doi: 10.5539/ijsp.v1n2p151

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., & Rammsayer, T. (2011). On the Special Relationship between Fluid and General Intelligence: New Evidence Obtained by Considering the Position Effect. *Personality and Individual Differences, 50,* 1249-1254.

Spector, P. E. (2006). Method variance in organizational research: truth or urban legend? *Organizational Research Methods, 9*, 221-232.

Spector, P. E., & Brannick, T. M. (2010). Common method issues: an introduction to the feature topic in organizational research methods. *Organizational Research Methods, 13*, 403-406.

Takane, Y., & de Leeuwe, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393-408.

Thurstone, L. L. (1938). *Primary and mental abilities*. Chicago, IL: The University of Chicago Press.

Vautier, S., Steyer, R., Jmel, S. & Raufaste, E. (2005). Imperfect or perfect dynamic bipolarity? The case of antonymous affective judgements. *Structural Equation Modeling, 12*, 391-410.

Verguts, T., & De Boeck, P. (2000). A Rasch model for detecting learning while solving an intelligence test. *Applied Psychological Measurement, 24*, 151-162.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 10*, 1-22.