Inductive Reasoning and its Underlying Structure: Support for Difficulty and Item Position Effects

Karl Schweizer[a,b], Stefan Troche[c], Thomas Rammsayer[c], and Florian Zeller[a]

[a] Institute of Psychology , Goethe University Frankfurt, Frankfurt, Germany

[b] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China

[c] Department of Psychology, University of Bern, Bern, Switzerland

ABSTRACT

This paper reports an investigation of the influence of method effects on the measurement of reasoning and of the relationships of these effects to basic cognitive processes. For this purpose, the variation due to the item-position and difficulty effects was separated from the variation due to the measured latent source of inductive reasoning. Data were collected by means of inductive reasoning items and cognitive tasks measuring working memory (WM) updating, rule learning, and automatization. Confirmatory factor analysis models served the decomposition of the variation of inductive reasoning data into a purified version of inductive reasoning, item-position, and difficulty components. The investigation of the relationships of corresponding latent variables and basic cognitive processes revealed two major associations: (a) the purified version of reasoning correlated with WM updating and (b) the item-position effect correlated with variants of learning. These results could be interpreted as signifying a two-dimensional structure of reasoning associated with executive functioning and learning processes.

KEYWORDS

automatization
inductive reasoning
difficulty effect
item-position effect
rule learning
working memory updating

INTRODUCTION

Although there seems to be no generally accepted definition of inductive reasoning, most researchers agree that inductive reasoning is the ability to detect similarities and/or dissimilarities in patterns of stimuli (Molnár et al., 2013; Klauer & Phye, 2008). Inductive reasoning is of particular interest in research on individual differences in psychometric intelligence because of its close association with the *g* factor of intelligence (Gustafsson, 1984; Schweizer et al., 2011). Therefore, it frequently serves as a proxy of general intelligence in intelligence research (cf. Carroll, 1993; Jensen, 1998; Raven, 1989; Sternberg & Gardner, 1983), but it is also used in psychological practice to obtain an index of intelligence in verbally impaired participants (Roth & Herzberg, 2008; Urbina, 2011).

A major characteristic of inductive reasoning is its assessment by larger sets of similar items requiring the detection of regularities or irregularities, demanding the uncovering and application of construction rules, or the formation of analogies (cf. Cattell, 1961; Formann, & Piswanger, 1979; Raven et al., 1997). However, the response to an item of such a scale is not only determined by inductive reasoning alone but also subject to various systematic and random influences. Of particular interest for the present study are two effects that show some likelihood of being stimulated alongside the measurement of reasoning, namely,

Corresponding author: Karl Schweizer, Institute of Psychology , Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt a. M., Germany

Email: k.schweizer@psych.uni-frankfurt.de

the item-position effect (Carlstedt et al., 2000) and the *difficulty effect* (e.g., Ferguson, 1941, Gibson, 1960), also referred to as method effects (Maul, 2013). The current study sought to isolate components of reasoning scores associated with the difficulty and item-position effects and to check whether they are related to cognitive processes measured by cognitive tasks or simply reflect negligible characteristics of measurement. As there is typically an overlap in the conditions leading to these effects in assessment, the patterns of relationships to cognitive processes will signify whether these effects correspond or differ.

The item-position effect refers to the dependency of statistical features of an item on the position of the item within the sequence of all items constituting the scale (Campbell & Mohr, 1950; Mollenkopf, 1950). Especially important for differential research is Knowles' (1988) observation that item reliability increases from the first to the last items of a scale indicating increasing systematic (or true) variance. More recent studies separated the item-position effect from inductive reasoning ability by using confirmatory factor analyses (CFAs) and extracting two latent variables from participants' responses instead of one (Schweizer, 2012; Troche et al., 2016). By doing so, the item-position effect was captured by an additional factor with increasing factor loadings from the first to the last item of the scale. It reflected the increasing influence of the item-position effect, while inductive reasoning was represented by another latent variable with equal-sized factor loadings of the items. Good model fit for this two-factor solution signified the appropriateness of the model as well as impurity in measurement.

The most favored explanation of the item-position effect ascribes it to learning processes that occur while completing a number of similar items (e.g., Carlstedt et al., 2000; Embretson, 1991; Verguts & De Boeck, 2000). Empirical support of the learning hypothesis of the item-position effect was provided by specifying learning as rule learning and automatization rather than associative learning (Ren et al., 2014; Schweizer et al., 2019). Furthermore, Ren et al. (2017) provided evidence for the notion that the item-position effect contributes to the functional relationship between working memory (WM) updating and performance on an inductive reasoning scale. This finding is of particular interest as it demonstrates that an uncontrolled item-position effect not only affects the factorial validity of an inductive reasoning scale but may also lead to an overestimation of the correlational association between inductive reasoning and cognitive processes such as WM updating.

The difficulty effect is expressed in an additional factor with factor loadings that reflect the difficulty levels of the investigated items, which are derived from the probabilities of providing a correct response. It has been proposed to originate from the variability and the wide range of item difficulties (Ferguson 1941; McDonald, 1965; McDonald & Ahlawat, 1974). In addition, the presence of a subset of items showing similar extremely high difficulty levels seems to play an important role in the generation of the difficulty effect. Such similarities have been suggested to create systematic variation that may necessitate the consideration of the difficulty factor in addition to the main factor in factor analytical studies (Bandalos & Gerster, 2016). A difficulty factor showing this characteristic was found in the assessment of inductive reasoning (Zeller, Reiss et al. 2017). Thus, due to impure measurement, test scores of inductive reasoning may not solely represent inductive reasoning but also a bias due to an effect of the difficulty levels of items. In contrast to the item-position effect, no study seems to exist on the cognitive processes underlying the difficulty effect. Moreover, convincing interpretations of cognitive correlates of the item-position effect and the difficulty effect are complicated by the fact that these two method effects heavily overlap even in a well-constructed inductive reasoning test where item difficulties increase with item positions. Consequently, unless item-position and difficulty effects are not statistically or otherwise dissociated from each other, it is impossible (or at least very difficult) to decide whether observed correlations between the item-position effect and specific cognitive processes, such as WM updating, are attributable to item position or item difficulty.

The described effects impair the quality of measurement unless they are eliminated. The elimination has to occur statistically, otherwise items of the same difficulty would be required for avoiding the difficulty effect and large breaks separating the applications of the individual items would be required for avoiding the item-position effect. The statistical elimination can be achieved by decomposing the variation of data into components reflecting the construct that is to be measured, the item-position effect and the difficulty effect. Fixed-links models (Schweizer, 2006, 2008) are well suited for this purpose. Factor loadings fixed to reflect the impact of the to-be-captured influences that unfold during the timespan of measurement characterize such models. The fixation of factor loadings is compensated by free variance parameters of the corresponding latent variables. If appropriately scaled, the estimates of variance parameters reflect the amount of variance captured by the corresponding latent variables. These variables become available for investigating the relationships to other variables representing constructs of interest for examining the validity of the scale.

The fixations occur according to the following rationales: the construct of interest as influence (e.g., inductive reasoning) is usually thought of as a source that contributes to all items. Therefore, it is usually represented by equal-sized factor loadings. The item-position effect as influence is described as an increasing trend (Knowles, 1988). It is captured by linearly and quadratically increasing series of numbers (usually ranging from zero to one, Schweizer, 2012; Zeller et al., 2017). Finally, there is the difficulty effect as influence. A characteristic feature of the corresponding factor is correspondence of patterns of factor loadings and difficulty levels (Guilford, 1941). Accordingly, the factor loadings have to be selected to reflect the difficulty levels (Schweizer & Troche, 2018).

In the present study, item-position and difficulty effects were separated from each other on the basis of a specific item arrangement. For this purpose, 18 items of Raven's Advanced Progressive Matrices (Raven et al., 1997) were presented in pseudorandom order. This arrangement of items assured that there was no overlap of item position and item difficulty. Furthermore, item-position and difficulty effects were separated from the inductive reasoning score. What remains after the removal of the effects is a "purified version of inductive reasoning

ability," which we referred to as "purified inductive reasoning." In addition, participants performed a battery of cognitive tasks previously shown to be related to the item-position effect. After the statistical separation of the effects using the modeling approach, the main objective of the present study was to find out how purified inductive reasoning and the two method effects related to WM updating, rule learning, and automatization. In particular, we examined whether the two method effects were the result of systematic variation but without substantive cognitive meaning or whether the two method effects reflected specific cognitive processes that contributed to performance when completing a reasoning scale.

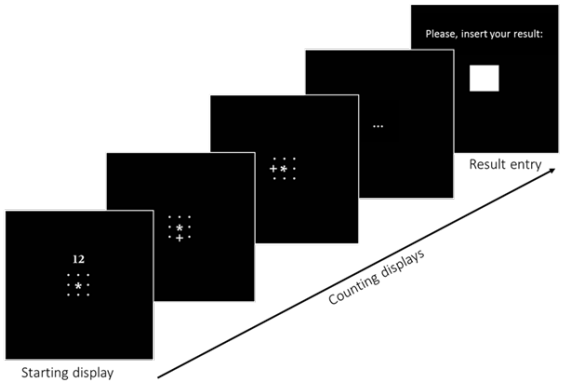

FIGURE 1.
Sample trial from positive updating from WM updating task.

METHOD

The 287 participants (92 males, 173 females, 22 not self-reporting their gender) were students at Goethe University Frankfurt in Germany ranging in age from 17 to 54 years (M_{age} ± SD = 22.8 ± 4.2 years). The participants received a financial reward or course credit. All participants reported normal hearing and normal or corrected-to-normal sight. Before being enrolled in the study, all participants were informed about the study protocol and gave their written informed consent. The reasoning data of a subsample served as illustration in a paper describing the mathematical foundation for investigating the item-position and difficulty effects. Here, the focus is on the cognitive underpinning of the two effects.

Materials

ADVANCED PROGRESSIVE MATRICES (APM)

Inductive reasoning was assessed by a short version of Raven's APM (Raven et al., 1997) introduced by Mackintosh and Bennett (2005). This version comprised 18 items. Each item consisted of a 3 × 3 matrix of geometric forms arranged according to logical rules. One cell of the matrix was free. Participants had to choose one out of eight alternative forms fitting the free cell of the matrix in accordance with the underlying rule. Items were presented in a pseudo-random order so that each participant completed the set of items in the same order but the items were not arranged according to increasing item difficulty, that is, a random order of items was established before the start of data collection and kept constant until the end of data collection. Testing time was 20 min.

ASSESSMENT OF WM UPDATING

The sign counting task was used to measure WM updating as a basic executive function that was proposed by Miyake et al. (2000). The task contained 18 trials for WM updating. Each trial consisted of a starting display and a varying number (6, 8, 10, or 12) of counting displays successively presented for 400 ms on a computer screen (see Figure 1). The starting display showed a two-digit starting number (e.g., 12) and, below this number, a fixation asterisk surrounded by nine dots. On the following counting displays, the starting number was not presented and on each display one of the dots was replaced by a plus sign. Participants' task was to add 1 to the starting number for each presented plus sign. After the last counting display of a trial was presented, the result of the counting had to be entered. As a measure of performance, the mean number of correct responses was computed for the first, second, third, and final subsets of the updating trials.

ASSESSMENT OF COMPLEX RULE LEARNING

Following Schweizer and Koch (2001), this task consisted of an initial computer-based learning phase and a subsequent testing phase. During the learning phase, participants were supposed to learn two simple rules and three complex rules. A capital letter (e.g., "G") was assigned to each rule. For each of the two simple rules, four displays were presented on the computer screen (see Figure 2). Each display contained a row of three "o" and/or "x" letters (e.g., "xxx", "xoo", "oxx", "ooo" with the joint rule that the third letter is the same as the second one). Participants could switch between the displays by pressing the up and down keys of the computer keyboard to compare the different displays and to detect the underlying rule. The same procedure was used for the complex rules. There were eight displays with four "o" and/or "x" letters. Participants were instructed to detect the rule underlying the displays of each block associated with a capital letter (e.g., "G") so that they would be able to add the last letter if this letter was omitted in the testing phase. A time limit of 10 min was set for the learning phase.

In the testing phase, participants were given a test sheet with 20 incomplete rows of "o" and/or "x" letters. For each row, one of the previously learnt rules (e.g., "G") was specified to be applied. Responses were given without time limitation. As a dependent variable, the number of correct answers was computed for each rule.

ASSESSMENT OF AUTOMATIZATION

The assessment of automatization occurred indirectly since automatization in the first place meant a change of the way in processing. Such a change could occur in combination with different cognitive operations stimulated in completing a cognitive task in virtually the same way. It was indirect because the focus was not on the mean performance but on the change of performance. Therefore, the decomposition of variation observed in repeated stimulations of the same

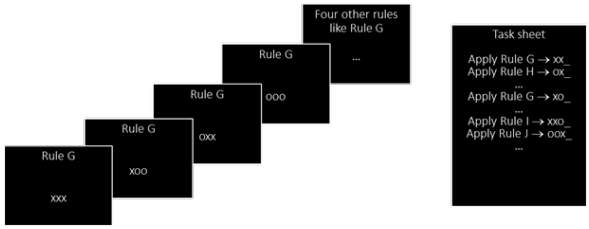

FIGURE 2.

Illustration of the rule learning task with four displays to learn Rule G (a simple rule). The other four rules also were learnt by four displays (simple rules) or eight displays (complex rules), respectively. Afterwards, a task test served the check of the learning.

cognitive operations and the isolation of what represents the expected change were important parts of the assessment of automatization.

For the assessment of automatization, a simple concentration task was used, as was suggested by Ren, Schweizer et al. (2013). This task required the repeated check of whether a simple stimulus was a target. Each time the same perceptual and attentional processes were called up, so that automatization in performance could occur across the course of checking the individual stimuli of the task. The task comprised four conditions. In each condition, two screens, with five rows of 20 digits each, were presented for 30 seconds (see Figure 3). The digits varied in pseudorandom order. Above and underneath the digits, one to four dashes were presented. The digit 9 accompanied by two dashes served as target stimulus. The participants' task was to correctly identify all target stimuli by means of a mouse click, unless the target stimulus was immediately preceded by the digit 5 (ignore stimulus). The number of ignore trials (i.e., the digit 5 preceding a 9 with two dashes) was systematically increased from five for each screen in the first condition to 10, 15, and 20 on each screen of the second, third, and fourth condition, respectively. Regardless of the increasing number of ignore trials, there were 22 valid targets for each screen across all task conditions. Presentations of the screens were separated by a 5- s break.

For each task condition, a performance score was computed as the difference between the number of correctly identified targets minus the number of false responses (i.e., responses to distractors or to ignore trials). These scores provided the outset for the statistical decomposition that is reported in the next section.

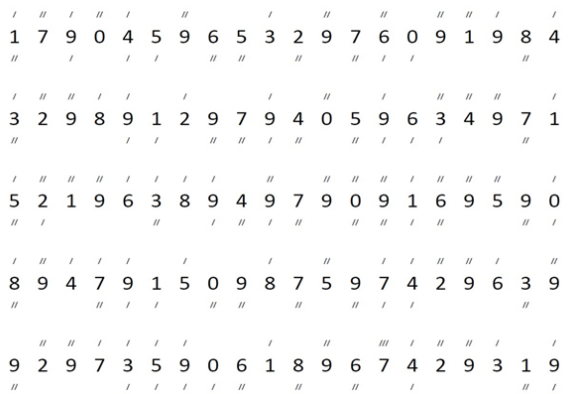

FIGURE 3.

Sample display from the concentration task used for collecting data for deriving automatization scores.

Statistical Analysis

At first, we prepared the three-factor confirmatory factor model with latent variables representing purified inductive reasoning, the item-position effect, and the difficulty effect and with APM items as manifest variables. Figure 4 illustrates this model (after the removal of parameters that did not reach the level of significance). Because of the need for brevity, the three latent variables are referred to as reasoning, position and difficulty latent variables in corresponding order. Analyses were based on probability-based covariances because of the binary APM data (Schweizer et al., 2015).

To assure that the latent variables represented reasoning, the item-position effect, and the difficulty effect, the factor loadings were constrained. All factor loadings on the inductive reasoning latent variable showed the same size and were set to 1. The numbers for constraining the factor loadings on the item-position effect latent variable were computed by means of the quadratic function and adjusted so that the largest loading was 1 (Zeller et al., 2017). The constraints for the factor loadings on the difficulty latent variable were achieved by subtracting the probabilities of a correct response from 1 so that the factor loadings were a direct function of the items' literal difficulties. Because of the binomial distribution of the reasoning data, a link transformation of the constraints was conducted as suggested by Schweizer et al. (2015). The variance parameters of the latent variables were set free for estimation and were required to yield statistical significance.

Working memory updating was represented by a latent variable extracted from four parcels of scores obtained from successive trials. Factor loadings were freely estimated and variance parameters were set to 1. Figure 5A illustrates this measurement model.

From the five scores of rule learning, two latent variables were derived to represent learning of simple rules and learning of complex rules, respectively. This separation of types of learning was necessary due to low correlations between performance on the easy and complex blocks. All factor loadings on these latent variables were set equal to 1 (see Figure 5B).

A characteristic of the task for the assessment of automatization was that the demands on selective attention were systematically increased across the four task conditions. To represent this increase in selective attention, a latent variable with fixed factor loadings (1, 2, 3, and 4) for the performance scores was extracted. The frequently repeated processing of very similar stimuli was supposed to induce automatization of information processing from the first to the fourth condition. Such an automatization process should result in a decrease of variance and, thus, was modeled by a latent variable with a decreasing course of factor loadings of the performance scores (1, 1/2, 1/3, and 1/4). A third latent variable captured individual differences in auxiliary processes, such as perceptual encoding or motor-related processes, which should be unaffected by task manipulations. Therefore, factor loadings on this third latent variable were kept constant (1, 1, 1, 1; see Figure 5C).

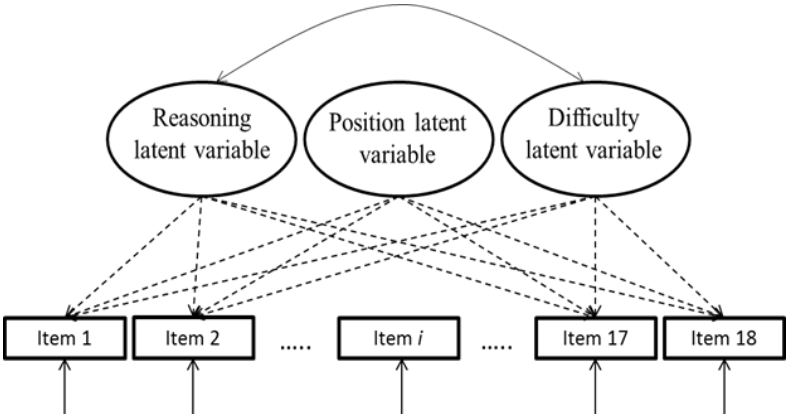

FIGURE 4.

Illustration of the APM measurement models with the purified-inductive reasoning latent variable ("Reasoning latent variable"), the item-position effect latent variable ("Position latent variable") and the difficulty effect latent variable ("Difficulty latent variable"). Solid shafts of arrows indicate free parameters and dashed shafts fixed parameters.

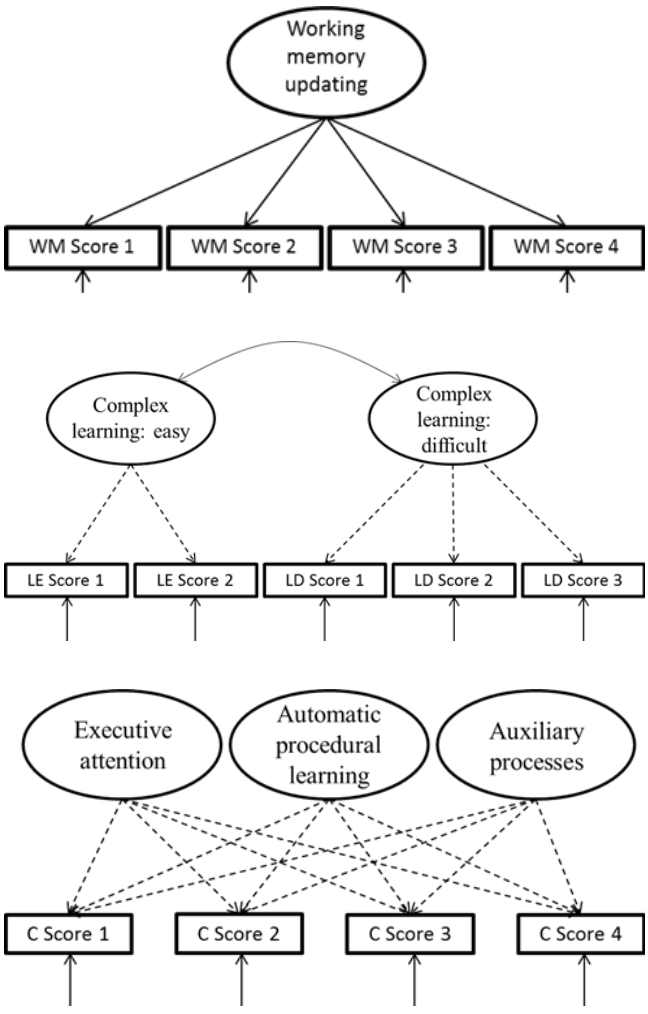

FIGURE 5.

Graphical representations of the measurement models regarding WM updating (top panel), rule learning (middle panel), and automatization (bottom panel). Solid shafts of arrows indicate free parameters and dashed shafts fixed parameters.

Finally, the relationships of the latent variables derived from APM data on the one hand and the latent variables representing selected cognitive processes on the other were investigated by means of structural equation modeling. The investigations of model fit were conducted separately for each combination of the three latent variables extracted from the APM items and the latent variable(s) of a cognitive task. The reason for this provision was that in modeling (unlike in basic statistics), estimated correlations were not stable over different configurations of parameters that were to be estimated. The parameters of a model were estimated simultaneously during the process of maximization of model fit. As a consequence, different configurations of parameters were likely to lead to more or less differing estimates for the relationship between the same two latent variables. Because of this instability of estimates, separated investigations were conducted to keep the number of influences on the estimation of the parameters of interest low. We proceeded as follows: first, we designed a model for each combination of APM and a cognitive task, which only comprised the necessary latent and manifest variables. These were combinations of purified-inductive reasoning, the item-position and difficulty latent variables, and of the latent variables representing selected cognitive processes associated with one cognitive task. Next, for each of these combinations, model fit was estimated. Subsequently, we estimated each parameter of interest, for example, the correlation between purified-inductive reasoning and WM updating, separately from the other parameters of interest, in this case, the correlations of the item-position latent variable and WM updating as well as of the difficulty latent variable and WM updating.

The parameters and model fit were estimated using maximum-likelihood estimation by means of the LISREL software package (Jöreskog & Sörbom, 2006). There are a number of fit indices with overlapping properties for investigating model fit. Usually, only a few of them are considered in the evaluation of model fit, as is recommended (see Schweizer, 2010). Root mean squared error of approximation (RMSEA; Steiger & Lind, 1980) is a badness-of-fit index. Standardized root mean square residual (SRMR) is an index that summarizes the residuals of investigating the difference between a model matrix and an observed covariance matrix. Comparative fit index (CFI; Bentler, 1990) compares the to-be-investigated model with the independence model with respect to data at hand. The following criteria for the fit indices were applied (see DiStefano, 2016; Hu & Bentler, 1999): RMSEA ≤ .06, SRMR ≤ .08, CFI ≥ .95, Akaike Information Criterion (AIC) that is in use for comparing models. Furthermore, the variance parameters of the latent variables were scaled to achieve estimates that could be compared with each other (Schweizer & Troche, 2019).

RESULTS

Position and Difficulty Latent Variables

The results of investigating APM data reported in this section were based on 284 participants since three participants provided random responses. The three-factor confirmatory factor model with correlations among the reasoning and position and difficulty latent variables showed an overall good model fit, χ^2 = 186.5 (*df* = 147), RMSEA = 0.031, SRMR = 0.062, CFI = 0.957, and AIC = 234.5. However, the check of the parameters revealed that the correlations between the position latent variable and the other two latent variables were not significant. After the removal of the insignificant correlations, the correlation between the reasoning and difficulty latent variables remained significant. Furthermore, the variance estimate of the position latent variable was significant irrespective of whether correlations with the reasoning or difficulty latent variables were allowed or not, supporting the assumption that the item-position factor is an independent latent variable. However, the significance of the difficulty factor depended heavily on the possibility to correlate with the reasoning latent variable. After the adjustment of the parameters, the final model showed the following fit characteristics: χ^2 = 195.4 (149), RMSEA = 0.033, SRMR = 0.066, CFI = 0.949, AIC = 239.4, indicating that the data were represented well by the model. We refer to this model as the APM model. In the following analyses, this model was used to assess the relations between the item-position and difficulty effects, on the one hand, and WM updating, rule learning, and automatization of information processing, on the other.

Relationship with WM Updating

As one participant did not complete the task, the following analyses were based on 283 participants only. To achieve sufficient degree of freedom for the measurement model with one latent variable, the error variables were assumed to correspond. This model described the data well, $\chi^2(2)$ = 1.57, RMSEA = 0.000, SRMR = 0.029, CFI = 1.00, and was added to the APM Model. The core structure of this model is illustrated by Figure 6, which also includes the estimates of the standardized correlations. The model fit of the complete model was good, χ^2 (202) = 263.5, RMSEA = 0.026, SRMR = 0.061, CFI = 0.960. The reasoning latent variable was significantly correlated with WM updating, r = .25, t = 3.76, p < .05. Furthermore, the difficulty latent variable correlated with WM updating, r = .29, t = 3.71, p < .05, but not the item-position latent variable, r = −.22, t = −1.35, p = .178.

Relationship with Automatization

In the automatization task, performance of five participants was considerably worse compared to the other participants, so they were excluded from further analyses. Three latent variables were derived from the automatization task to represent automatization of cognitive processing, executive attention, and auxiliary processes. The measurement model described the data well with the exception of RMSEA, χ^2 (3) = 11.19, RMSEA = 0.099, SRMR = 0.038, CFI = 0.987. Furthermore, combining this model with APM Model led to an acceptable to good model fit, χ^2 (223) = 378.3, RMSEA = 0.051, SRMR = 0.079, CFI = 0.924. As illustrated in Figure 7, the item-position latent variable correlated positively and significantly with the latent variable reflecting automatization of cognitive processing, r = .44, t = 2.93, p < .05. Furthermore, there were small but substantial correlations of au-

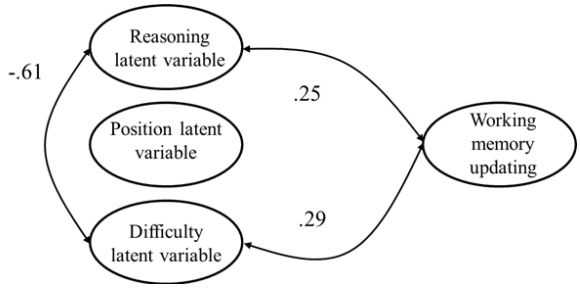

FIGURE 6.

Illustration of the model used for investigating the relationship of the APM latent variables on one hand and working memory updating on the other hand with standardized parameter estimates. The ellipses symbolize latent variables. The ellipses to the left are derived from the APM items and the ellipses to the right from the scores of the WM updating task. The double-headed arrows symbolize correlations.

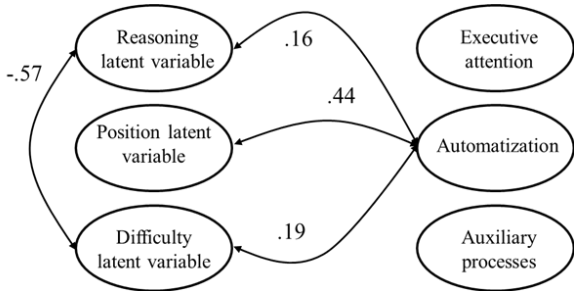

FIGURE 7.

Illustration of the model used for investigating the relationship of the APM latent variables on one hand and automatization on the other hand with standardized parameter estimates. The ellipses symbolize latent variables. The ellipses to the left are derived from the APM items and the ellipses to the right from the scores of the concentration task used for capturing automatization. The double-headed arrows symbolize correlations.

tomatization with the reasoning latent variable, $r = .16$, $t = 2.30$, $p > .05$, and the difficulty latent variable, $r = .19$, $t = 2.12$, $p > .05$.

Relationship with Rule Learning

As described above, two latent variables were derived from the rule-learning task to represent easy and difficult rule learning. The model fit of the measurement model was good, χ^2 (6) = 8.16, RMSEA = 0.036, SRMR = 0.039, CFI = 0.984. Combining this measurement model with the APM Model with reasoning, position, and difficulty latent variables led to a model with good RMSEA and SRMR, and acceptable CFI, χ^2 (240) = 293.6, RMSEA = 0.029, SRMR = 0.063, CFI = 0.945. A large correlation of the position latent variable and learning of difficult rules was observed, $r = .42$, $t = 2.18$, $p < .05$ (see Figure 8). Furthermore, the reasoning latent variable showed significant correlations with learning of easy rules, $r = .22$, $t = 3.41$, $p < .05$, as well as difficult rules, $r = .21$, $t = 3.32$, $p < .05$. Moreover, there was a weak but statistically significant correlation between the difficulty latent variable and easy rule learning, $r = .17$, $t = 2.21$, $p < .05$.

DISCUSSION

In the empirical sciences, the validity of research results heavily depends on the validity of the data. This applies to all kinds of experimental and differential research, including intelligence research. Validity is even considered as "the most important concept in psychometrics" (Sireci, 2007, p. 477). Possible impairments of the validity of psychometric data are method effects (Maul, 2013). But method effects are not just impairments in the sense of error variation. Instead, they are impairments in the sense of systematic variation of data that is unrelated to the construct to be measured. Systematic variation means that there is a source that influences responding in a systematic way.

The results of the investigation of the relationships between the components of inductive reasoning and a selection of cognitive processes provided further evidence of latent sources underlying method effects. The list of stronger correlational effects, that is, correlations suggesting five or more percent of common variance, extends to combination of the purified version of reasoning and working memory updating (.25), of difficulty effect and working memory updating (.29), of position effect and automatization (.44), and also of position effect and rule learning (.42). The list of weaker correlational effects comprises the combinations of the purified version of reasoning and automatization (.16), of the purified version of reasoning and learning easy and also difficult rules (.22 and .21), of difficulty effect and automatization (.19), and of difficulty effect and learning easy rules (.17). To note, in distinguishing between stronger and weaker correlational effects, we followed the suggestion to highlight results promising the higher degree of replicability (Aarts et al., 2015).

The results regarding the difficulty effect are interesting in three different ways. First, the attainment of (marginally) good model fit is due to the contribution of the difficulty latent variable. This latent variable improves model fit and at the same time provides a reason for the otherwise poor model fit. The likely reason is a set of a few items showing extreme difficulty levels (Bandalos & Gerstner, 2016) that create systematic variation which is not reachable for a general latent variable

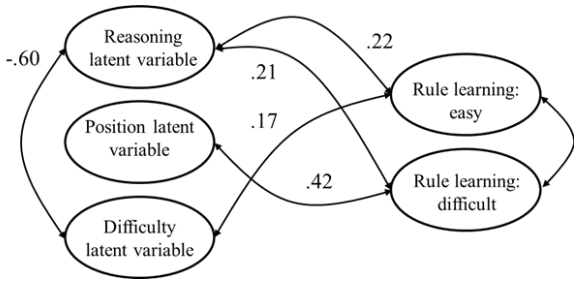

FIGURE 8.

Illustration of the model used for investigating the relationship of the APM latent variables on one hand and rule learning on the other hand with standardized parameter estimates. The ellipses symbolize latent variables. The ellipses to the left are derived from the APM items and the ellipses to the right from the scores of the rule learning task. The double-headed arrows symbolize correlations.

(i.e., the reasoning latent variable). Second, after the rearrangement of the items for separating the representations of the difficulty and position effects proved to be efficient, it is clear that these effects and the corresponding latent variables have an existence of their own. They do not even correlate with each other. Third, there is an indication of a close (negative) relationship of the reasoning and difficulty latent variables. They even seemed to complement each other in the sense that the difficulty latent variable accounts for variation that is out of reach for the main reasoning latent variable.

Next, we provide an explanation for the not-so-large correlations observed between the reasoning latent variable and the cognitive processes investigated in the present study. The strongest correlations include the position latent variable instead of the reasoning latent variable. This outcome might be the result of the study design that was not so much focused on the purified version of inductive reasoning ability but on the difficulty and item-position effects. This means that we selected cognitive tasks that already proved to be especially strong in their relationship to the item-position effect. But, despite this, it is the reasoning latent variable that correlated with all three cognitive processes whereas the position latent variable only correlated with two of them. The relationship with WM updating was unique, and it was in line with previous research regarding executive functioning (Ren et al., 2013; Wang et al., 2020). The two correlations for the position effect refer to a common source that is learning, and is in line with the most widely accepted hypothesis of the item-position effect that learning is the underlying cognitive mechanism (e.g., Carlstedt et al., 2000; Embretson, 1991; Verguts & De Boeck, 2000).

Lastly, we propose general interpretations of the results regarding inductive reasoning. There are two options: one is highlighting the difference between the two major types of sources of performance in completing the APM; these are substantive and method sources. This option suggests that the measure is impure and that removing the item-position effect from the observed scores yields purified inductive reasoning. This option is supported by the empirical fact that the item-position effect is not restricted to measures of reasoning but can also be found in attitude data (Knowles, 1988) and other data types. The other option is to assume a hierarchical structure of inductive reasoning. Whereas the lower level includes pure inductive reasoning and learning as components, the upper level consists in general inductive reasoning. If we accept that adaptation to novelty (Cattell, 1963; Sternberg, 1984) is a main characteristic of fluid reasoning, the second option is the most appropriate one. Furthermore, the observation that scales of fluid reasoning predict success in educational settings especially well is in line with this interpretation (Ren et al., 2015). Further research may show which one of these options provide a fruitful basis for further developments and the better account for empirical observations.

Possible limitations to our findings include the influence of other cognitive effects, such as the carryover effect, wherein switching from completing one cognitive task to completing another cognitive task leads to fatigue that could accumulate over the course of the experiment. These types of limitations are especially likely to impact the average results whereas the rank orders among the participants may be modified but to a lesser degree.

ACKNOWLEDGEMENTS

This work was supported by Deutsche Forschungsgemeinschaft, Kennedyallee 40, 53175 Bonn, Germany [grant number SCHW 402/20-1].

REFERENCES

Aarts, A., Anderson, J. E., Anderson, C. J., & Attridge, P. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). doi: 10.1126/science.aac4716

Bandalos, D. L., & Gerstner, J. J. (2016). Using factor analysis in test construction. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction* (pp. 26–51). Hogrefe.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. doi: 10.1037/0033-2909.107.2.238

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. doi: 10.1037/h0061818

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

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

Cattell, R.B. (1961). *The Culture Free Intelligence Test, Scale 3*. Institute for Personality and Ability Testing.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *Journal of Educational Psychology, 54*, 1–22. doi: 10.1037/h0046743

DiStefano, C. (2016). Examining fit with structural equation models. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction* (pp. 166–193). Hogrefe.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 56*, 495–515. doi: 10.1007/BF02294487

Ferguson, G. A. (1941). The factorial interpretation of test difficulty. *Psychometrika, 6*, 323–329. doi:10.1007/BF02288588

Formann, A. K., & Piswanger, K. (1979). *Wiener Matrizen-Test*. Beltz.

Gibson, W. A. (1960). Nonlinear factors in two dimensions. *Psychometrika, 25*, 381–392. doi: 10.1007/BF02289755

Guilford, J. P. (1941). The difficulty of a test and its factor composition. *Psychometrika, 6*, 67–77. doi: 10.1007/BF02292175

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179–203.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi: 10.1080/10705519909540118

Jensen, A. R. (1998). *The g factor. The science of mental ability*. Praeger.

Jöreskog, K. G., & Sörbom, D. (2006). LISREL 8.80. Lincolnwood, IL:

Scientific Software International Inc.

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

Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of Educational Research, 78*, 85–123.

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.

Mackintosh, N. J., & Bennett, E. S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence, 33*, 663-674. doi: 10.1016/j.intell.2006.06.003

Maul, A. (2013). Method effects and the meaning of measurement. *Frontiers in Psychology, 4*:169. doi: 10.3389/fpsyg.2013.00169

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

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27*, 82–99. doi: 10.1111/j.2044-8317.1974.tb00530.x

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100. doi: 10.1006/cogp.1999.0734

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

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity, 9*, 35–45.

Neubauer, A. C. (1990). Coping with novelty and automatization of information processing: An empirical test of Sternberg's two-facet subtheory of intelligence. *Personality and Individual Differences, 11*, 1045–1052. doi: 10.1016/0191-8869(90)90132-B

Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement, 26*, 1–16. doi: 10.1111/j.1745-3984.1989.tb00314.x

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

Ren, X., Altmeyer, M., Reiss, S., &. Schweizer, K. (2013). Process-based account for the effects of perceptual and executive attention on fluid intelligence: an integrative approach. *Acta Psychologica, 142*, 195–202. doi: 10.1016/j.actpsy.2012.12.007

Ren, X., Schweizer, K., & Xu, F. (2013). The sources of the relationship between sustained attention and reasoning. *Intelligence, 41*, 51–58. doi: 10.1016/j.intell.2012.10.006

Ren, X., Schweizer, K., Wang, T., & Xu, F. (2015). The prediction of students' academic performance with fluid intelligence in giving special consideration to the contribution of learning. *Advances in Cognitive Psychology, 11*, 97–105. doi: 10.5709/acp-0175-z

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

Roth, M., & Herzberg, P. Y. (2008). Psychodiagnostik in der Praxis: State of the Art. [Psychological diagnostics in practice: State of the art]. *Klinische Diagnostik und Evaluation, 1*, 5-18.

Schweizer, K. (2006). The fixed-links model for investigating the effects of general and specific processes on intelligence. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2*, 149–160. doi: 10.1027/1614-2241.2.4.149

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. (2010). Some guidelines concerning the modelling of traits and abilities in test construction. *European Journal of Psychological Assessment, 26*, 1–2. doi: 10.1027/1015-5759/a000001

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

Schweizer, K. & Koch, W. (2001). A revision of Cattell's investment theory: cognitive properties influencing learning. *Learning and Individual Differences, 13*, 57–82. doi: 10.1016/S1041-6080(02)00062-6

Schweizer, K., Troche, S. J., & Rammsayer, T. H. (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. doi:10.1016/j.paid.2011.02.019

Schweizer, K., Ren, X., & Wang, T. (2015). A comparison of confirmatory factor analysis of binary data on the basis of tetrachoric correlations and of probability-based covariances: A simulation study. In R. E. Millsap, D. M. Bolt, L. A. van der Ark, & W.-C. Wang (Eds.), *Quantitative psychology research* (pp. 273–292). Springer.

Schweizer, K., & Troche, S. (2018). Is the factor observed in investigations of the item-position effect actually the difficulty factor? *Educational and Psychological Measurement, 78*, 46–69. doi: 10.1177/0013164416670711

Schweizer, K., & Troche, S. (2019). The EV scaling method for variances of latent variables. *Methodology, 15*, 175–184. doi: 10.1027/1614-2241/a000179

Schweizer, K., Zeller, F., & Reiß, S. (2019). Higher-order processing and change-to-automaticity as explanations of the item-position effect in reasoning tests. *Acta Psychologica*. doi: 10.1016/j.actpsy.2019.102991

Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher, 36*, 477–481. doi: 10.3102/0013189X07311609

Steiger, J. H., & Lind, J. C. (1980, May). Statistically based tests for the number of common factors. In *Annual meeting of the Psychometric Society, Iowa City, IA (Vol. 758)*.

Sternberg, R. J. (1984). Toward a triarchic theory of human intelligence. *The Behavioral and Brain Sciences, 7*, 269–315. doi: 10.1016/0160-2896(81)90021-0

Sternberg, R. J., & Gardner, M. K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General, 112*, 80–116. doi: 10.1037/0096-3445.112.1.80

Troche, S., Wagner, F. L., Schweizer, K., & Rammsayer, T. H. (2016). The structural validity of the Cultural Fair Test under consideration of the item-position effect. *European Journal of Psychological Assessment*. doi:10.1027/1015-5759/a000384

Urbina, S. (2011). Tests of intelligence. In R.J. Sternberg & S.B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 20–38). Cambridge UniversityPress.

Verguts, T., & De Boeck, P. (2000). A Rasch model for detecting learning while solving an intelligence test. *Applied Psychological Measurement, 24*, 151–162. doi:10.1177/01466210022031589

Wang, T., Li, C., Wei, W., & Schweizer, K. (2020). An investigation on how inhibition in cognitive processing contributes to fluid reasoning. *Advances in Cognitive Psychology, 16*, 176–185. doi: 10.5709/acp-0295-7.

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

Zeller, F., Reiss, S., & Schweizer, K. (2017). Is the item-position effect in achievement measures induced by increasing item difficulty? *Structural Equation Modeling, 24*, 745–754. doi: 10.1080/10705511.2017.1306706

RECEIVED 09.11.2020 | ACCEPTED 27.09.2021