The development of inductive reasoning under consideration of the effect due to test speededness

Natalie Borter[1], Annik E. Völke[1] & Stefan J. Troche[1]

[1] Institute of Psychology, University of Bern, Bern, Switzerland

Abstract:
Measures of inductive reasoning are frequently used as proxy of a child's cognitive development. Unfortunately, a reasoning scale might be affected by speededness introduced by limited testing time. As a result, the scale might be heterogeneous and its correlation with age is hard to interpret. Here we investigated the development of inductive reasoning when a possible bias by the effect of speededness is controlled for. In 250 children, ranging in age from 8;0 to 12;8 years, inductive reasoning assessed with the Culture Fair Test 20-R (CFT 20-R) increased with age. The effect of speededness was identified in all four CFT 20-R subtest and was also related to age indicating increasing processing speed with higher age. After controlling for the effect of speededness, the relation between age and inductive reasoning was still observed but substantially decreased. Consequences of these results for the description of inductive reasoning data obtained with time-limited tests and for developmental studies on the interplay between age, inductive reasoning and speed of information processing are discussed.scussed.

Keywords:
inductive reasoning, cognitive development, speededness, culture-fair test, confirmatory factor analysis

Introduction

Inductive reasoning can be defined as the detection of similarities and differences between characteristics and relations of objects (Klauer & Phye, 2008). In intelligence research, inductive reasoning is of particular interest as an integral part of fluid intelligence. Due to the close relationship between fluid intelligence and the higher-order *g* factor of intelligence (Gustafsson, 1984), inductive reasoning tests such as Raven's Matrices (Raven, 1962) or Cattell's Culture Fair Test (CFT; Cattell & Cattell, 1963) are commonly used to estimate individuals' general intelligence.

Inductive reasoning is well known to increase from early childhood to young adulthood. This was reported in early studies on fluid intelligence (e.g., Horn & Cattell, 1967) but also in more recent studies (e.g., Engel de Abreu, Conway, & Gathercole, 2010; Fry & Hale, 2000; Molnár, Greiff, & Csapó, 2013). Csapó (1997) showed this strong development in a broad variety of reasoning tests and it seemed to be independent of the specificity of the content to be processed. The developmental course of inductive reasoning, as a proxy of general intelligence, is frequently compared with and related to the developmental course of other cognitive processes. Such investigations contributed to the dissociation of fluid and crystallized intelligence (Horn & Cattell, 1967) but also elucidated the interplay of the development of inductive reasoning and the development of other cognitive abilities such as processing speed (Kail, 2000) or working memory capacity (Fry & Hale, 2000; Swanson, 2008).

In recent years, a growing number of studies on the factorial validity of inductive reasoning tests challenged the assumption of a single construct underlying the covariance of the items or, in other words, the assumption of item homogeneity (Estrada, Roman, Abad, & Colom, 2017; Ren, Schweizer, Wang, & Fen, 2015; Schweizer, 2011; Schweizer, Schreiner, & Gold, 2009; Schweizer, Troche, & Rammsayer, 2018; Troche, Wagner, Schweizer, & Rammsayer, 2016). This is quite surprising since the items of inductive reasoning scales are quite similar with regard to the content to be processed and the rules to be identified in order to solve the items.

Item heterogeneity reduces the interpretability of an individual test score as an estimator of the "true" inductive reasoning ability. Furthermore, correlations between scores obtained by means of a heterogeneous reasoning scale and any other variable are difficult to interpret because it is not clear whether the correlation is based on variation in inductive reasoning or variation in another construct affecting the inductive reasoning scores. Speededness of a test might be a possible reason for the heterogeneity of inductive reasoning scales when these scales are administered with a time limit (Lu & Sireci, 2007).

According to Lu and Sireci (2007), a "test is viewed as "speeded" when examinees' scores are determined by the amount of items attempted as well as the accuracy of responses" (p. 30). In other words, test speededness corresponds to the degree a test is affected by its time limit because participants have not enough time to attempt all items or guess on items due to the limited time (Estrada et al., 2017; Wilhelm & Schulze, 2002). Test speededness would be a minor problem as long as test performance of each participant would be equally affected by the time limit (Schweizer, Reiss, & Troche, 2019). Due to individual differences in processing speed, however, test speededness probably affects slower participants more strongly than faster participants. Consequently, individual differences in processing speed can be expected to result in a systematic influence on (primarily) the items at the end of the scale (i.e. not-reached items). Thus, a factor might become extractible from these items, which depends on the extent of limited testing time (i.e. test speededness) as well as on individual differences in processing speed (Estrada et al., 2017; Schweizer, Reiss, et al., 2019). To note, while speededness is a characteristic of a time-limited test, the effect of speededness refers to individual differences in processing the test items within this time limit. Thus, this effect of speededness reflects individual differences in speed of test-taking or, more generally, processing speed.

Estrada et al. (2017) proposed using bifactor modeling to control for the effect of speededness on inductive reasoning. In their bifactor model, they extracted one latent variable from the common variance of all items of an inductive reasoning scale to represent individual differences in inductive reasoning. Concurrently, they derived a second latent variable from the items at the end of the scale, which were not reached by at least one participant. Due to the increasing number of omissions in these items, Estrada et al. (2017) suggested this latent variable to reflect the effect of speededness. The structure of such a bifactor model is depicted in Figure 1. The data description was improved by Estrada et al.'s bifactor model compared to a congeneric model. An unambiguous interpretation of the latent variable in terms of an effect of speededness, however, was exacerbated because also other similarities of the items at the end of the scale (e.g. difficulty or item position) were probably captured by this latent variable.

Schweizer, Reiss, et al. (2019) used a similar bifactor modelling approach (also see Schweizer & Ren, 2013). In contrast to Estrada et al. (2017), however, they proceeded from the explicit assumption that individual differences in (latent) processing speed are normally distributed. Such a normally distributed source of the effect of speededness should lead to an approximately cumulative distribution of omissions. Consequently, in Schweizer, Reiss, et al.'s (2019) conceptualization of a bifactor model, the factor loadings of the latent variable representing the effect of speededness were fixed to follow the course of the increasing logistic function to trace the cumulative distribution of omissions. All factor loadings on the latent variable representing inductive reasoning were fixed to the same value with the assumption of homogeneity. Schweizer, Reiß, Ren, Wang and Troche (2019) demonstrated that this procedure with fixed factor loadings was better suited to depict the effect of speededness and to interpret the latent variable less ambiguously compared to the free estimation as suggested by Estrada et al. (2017). Furthermore, similar to the study by Estrada et al. (2017), Schweizer, Reiss, et al. (2019) demonstrated that the speededness effect accounted for a substantial portion of the variance and covariance of the inductive reasoning items. Therefore, its explicit consideration in the measurement model of inductive reasoning improved the quality of data description substantially.

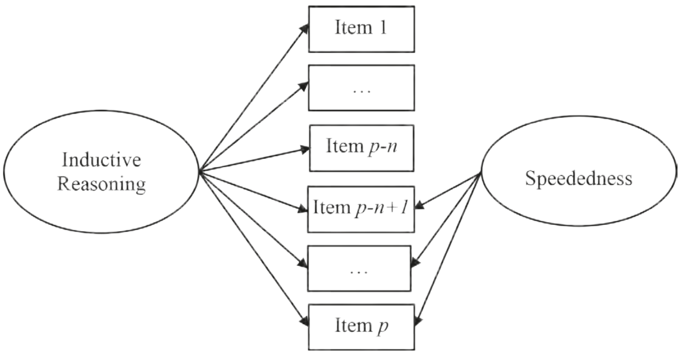

Figure 1
Bifactor model with one latent variable representing inductive reasoning and one latent variable representing the effect of speededness. Note: *p* = the number of items in the test; *n* = the number of items in the test affected by speededness.

A possible confound of inductive reasoning and processing speed might be a challenge primarily for developmental studies since not only inductive reasoning but also processing speed increases with chronological age (Kail, 1991). Kail (2000) found the development of processing speed to be domain general across a wide variety of cognitive processes so that he proposed a global mechanism underlying the development of processing speed, which might be directly connected to the development of the central nervous system. From this point of view, uncontrolled confounding effects of processing speed and inductive reasoning hamper the interpretation of an age-related increase of test scores in time-limited inductive reasoning scales since it is not clear whether or to which extent this increase is caused by an increase of inductive reasoning or by an increase of processing speed.

The first goal of the present study was to probe whether the description of inductive reasoning data of 250 children ranging in age from 8 to 13 years obtained with the Culture Fair Test 20-R (CFT 20-R) as a commonly used and speeded measure of inductive reasoning can be improved by the consideration of the effect of speededness. For this purpose, we used confirmatory factor analysis to compare the data description by the congeneric model, the essentially tau equivalent model, and a bifactor model, which considered the effect of speededness. More specifically, in the congeneric model, only one latent variable was assumed to describe the data well when factor loadings on all items were freely estimated. With the essentially tau equivalent model the assumption was made that each item was an equally good estimation of the latent variable so that all factor loadings were fixed to the same value. Proceeding from the assumption that an effect of speededness led to item heterogeneity, the bifactor model followed the approach proposed by Schweizer, Reiss, et al. (2019). Thus, one latent variable loaded equally on all items while a second latent variable loaded only on items, which were not reached by all participants. The course of factor loadings of this second latent variable followed the logistic function in order to connect this latent variable as close as possible to individual differences in processing speed (see Schweizer, Reiß, et al., 2019, as well as the method section for more details).

The second goal was to investigate the correlation between chronological age and inductive reasoning. Presupposing that an effect of speededness could be identified in the data of the CFT 20-R, it was of particular interest whether the correlational relationship between inductive reasoning ability and age would change considerably when statistically controlled for the effect of speededness.

Method

Participants. The sample consisted of 250 children from public schools in Switzerland ranging in age from 8;0 to 12;8 years (M=9;2 years; SD=0;4 years). The sample contained 141 girls (56.4%) and 109 boys (43.6%). Parents gave their written consent for their children's participation. The local ethics committee approved the study protocol.

Culture Fair Test (CFT) 20-R. As a measure of inductive reasoning, we used the short version of the German CFT 20-R (Weiß, 2006). The CFT 20-R is composed of four subtests: *Series, Classifications, Matrices,* and *Topologies.* The first three subtests contain 15 items, Topologies contains 11 items.

The level of an individual's intelligence is estimated across all four subtests with an internal consistency of .92 (Weiß, 2006).

In *Series*, each item is composed of a progressive series of three figures and five response alternatives. The participants are required to choose the alternative, which continues the progressive change of the series. Each item of *Classifications* consists of five figures. The participants' task is to select one figure, which differs from the other four figures regarding a specific feature such as horizontal vs. vertical orientation. The items of *Matrices* contain a 2x2 matrix or a 3x3 matrix. One cell of each matrix is empty and participants have to choose the correct figure out of five alternatives, which would fit the cell according to the rules underlying the matrix. In *Topologies*, a reference configuration of geometrical figures (e.g., an arc and a circle) is given for each item. One or more dots are presented in this reference configuration. Participants choose one out of five alternative configurations of geometrical figures, in which the dot(s) could be placed with the same topological relationship to all parts of the configuration as in the reference configuration (e.g., below an arc and within a circle).

As recommended in the test manual, standardized testing time was four minutes for Series and Classifications, respectively, and three minutes for Matrices and Topologies, respectively. Each testing session lasted about 30 minutes. Testing took place in small groups of two to five participants.

Confirmatory factor analyses. Confirmatory factor analyses (CFA) were conducted using R version 3.2.0 and the lavaan package (Rosseel, 2012). The children's responses on each item were coded as 1 for correct responses or 0 for incorrect responses or omissions. From the resulting binary data, variances and probability-based covariances between the items served as input to CFA (Schweizer, 2013). For the *congeneric models*, one latent variable was derived from the items of each subtest and the factor loadings (λ) were freely estimated. In a second step, essentially *tau-equivalent models* were computed with one latent variable derived from the items of each subtest. In contrast to the congeneric models, all factor loadings were fixed to the same value ("1") assuming that only one ability influences the processing of all items in the same way.

For the bifactor model, a second latent variable, independent of the first latent variable and representing the effect of test speededness was added to the tau-equivalent model. The factor loadings on this latent variable were fixed in an increasing manner from the first item affected by speededness to the last item. The increasing trajectory was assumed to follow a logistic function (Schweizer & Ren, 2013):

$$\lambda(i) = \frac{e^{i-j}}{1+e^{i-j}}$$

where i indicates the position of an item within all items not-reached by at least one participant, while j is a constant that may be selected to optimize model fit (Schweizer et al., 2019).

As in previous research, we registered the last attempted item for each participant. Missings after this item were classified as "not-reached" responses.

In addition to this first bifactor model, a second bifactor model was tested. When the majority of test takers do not reach the items at the end of a scale, these items are only insufficient indicators of individual differences in inductive reasoning. Accordingly, inductive reasoning might not be represented by all items in the same way.

Therefore, instead of fixing all loadings on the latent variable representing inductive reasoning to the same value, a fading out was modelled in this second bifactor model (see Zeller, Reiss, & Schweizer, 2019). For this purpose, the fixations were multiplied by the following function:

$$\lambda(i) = 1 - \frac{e^{i-k}}{1 + e^{i-k}}$$

where *i* indicates the position of the item within all items not-reached by at least one participant, and *k* is again a constant that may be selected to optimize model fit.

To take into account differences between data and model regarding scale level and distribution (Schweizer, 2012), each fixated factor loading was weighted by the standard deviation (SD_i) of the corresponding item *i,* where p_i is the probability of a correct response:

$$SD_i = \sqrt{p_i \cdot (1 - p_i)}$$

For model evaluation, variances of the latent variables with fixed factor loadings were scaled according to the criterion-based method proposed by Schweizer (2011) with the average of squared factor loadings equaling one. These variances were tested for statistical significance, in a first step, to examine whether the respective latent variables would have psychological meaning. In a second step, models were judged as being acceptable when $\chi^2/df < 3$ (Schermelleh-Engel, Moosbrugger & Müller, 2003), root mean square error of approximation (RMSEA) below 0.06 (Hu & Bentler, 1999), and standardized root mean square residual (SRMR) ≤ .08 (Hu & Bentler, 1999). Finally, the Akaike Information Criteria (AIC) was used to compare models. A lower AIC indicates a better model fit under consideration of the model's complexity.

Usually, the evaluation of CFA models also includes the comparative fit index (CFI). The CFI compares the model fit of the tested model with a baseline model, in which all covariances are assumed to be zero (Bentler, 1990). As a rule of thumb, Kenny (2015) suggested that an RMSEA less than 0.158 in the *baseline model* (indicating a quite good baseline model) leads to a non-informative CFI. All baseline models investigated in the present study had an RMSEA below 0.158 (*Series*: RMSEA = .108; *Classification*: RMSEA = .072; *Matrices:* RMSEA = .149; *Topologies*: RMSEA = .099). Hence, CFI was not used for model evaluation in the present study.

Results

Descriptive statistics of the children's raw scores in the four subtests and the corresponding IQ scores are given in Table 1. The mean IQ was close to the population mean of 100 and the standard deviation was only marginally restricted compared to the population standard deviation of 15. As also presented in Table 1, raw scores of all four subtests as well as the sum score increased significantly with age. Due to the age-related standardization, the IQ score did not correlate significantly with age.

More than 99.5% of the participants correctly solved the second item of Series, the first item of Classification, the first four items of Matrices and failed to solve the last item of Topologies. These items were excluded from further analyses as their variance was too restricted.

In the bifactor models, the parameter *j* of the logistic function describing the course of factor loadings on the latent variable rep-

Table 1 Mean (M), Standard Deviation (SD) and Range of Standardized IQ Scores, Sum Scores Across all Four CFT20-R Subtests and Raw Scores in the Subtests Series, Classifications, Matrices, and Topologies in 250 Children as well as Pearson Correlations between Intelligence Scores and Age.

	M	SD	Min	Max	Correlation with age
IQ score	103.62	13.14	70	135	.02
Sum score	30.24	6.53	13	44	.45***
Series	9.81	2.47	0	15	.31***
Classification	7.16	2.36	1	13	.32***
Matrices	8.78	2.42	3	15	.34***
Topologies	4.49	1.95	0	10	.31***

***p < .001.

Table 2 Fit Indices for Congeneric, Essentially Tau-Equivalent and Speededness-Effect Models on CFT20-R Subtests as well as Variances (φ) of the Latent Variables with Constant (Reasoning) and Increasing (Speededness) Factor Loadings (N=250).

Subtest	Model	χ^2	df	*p*
Series	Congeneric	130.82	77	<.001
	Essentially tau-equivalent	178.38	90	<.001
	Speededness without fading out	178.38	89	<.001
	Speededness with fading out	136.93	89	<.001
Classifications	Congeneric	107.47	77	.012
	Essentially tau-equivalent	120.72	90	.017
	Speededness without fading out	119.30	89	.018
	Speededness with fading out	109.67	89	.068
Matrices	Congeneric	102.67	44	<.001
	Essentially tau-equivalent	161.30	54	<.001
	Speededness without fading out	154.78	53	<.001
	Speededness with fading out	81.11	53	.008
Topologies	Congeneric	63.46	35	.002
	Essentially tau-equivalent	83.36	44	<.001
	Speededness without fading out	78.53	43	.001
	Speededness with fading out	69.22	43	.007

*p < .05; **p < .01; ***p < .001

resenting the speededness effect was set to 1.5 for all four subtests. The parameter *k* of the function describing the fading out of inductive reasoning due to the shrinkage of variance in the last items was set to 6 for the subtests Series and Classification and to 5 for the subtests Matrices and Topologies.

As can be taken from Table 2, the congeneric model described the data of subtest *Series* better than the essentially tau-equivalent model and the bifactor model, when no fading out of inductive reasoning was assumed. However, when such a fading out was considered, the resulting fit of the bifactor model was good according to χ^2/df, RMSEA, and SRMR and even better than the model fit of the congeneric model as indicated by the AIC comparison. Accordingly, even with a penalty for lower parsimony, the bifactor model with fading out described the data better than the more parsimonious congeneric model. Variances of both latent variables yielded statistical significance in this model (see Table 2).

The essentially tau-equivalent model better described items of *Classifications* than the congeneric model as indicated by the lower AIC. The consideration of the effect due to speededness improved the data description only when a fading out was added. Both variances of the bifactor model with fading out were statistically significant (see Table 2). The model fitted the data well according to χ^2/df, RMSEA, and SRMR and its AIC was lower than the AIC of the other models.

For *Matrices*, the congeneric model led to better model fit than the essentially tau-equivalent model according to the AIC comparison. As for Series and Classifications, the bifactor model yielded a better model fit only when the assumption was

χ^2/df	RMSEA	SRMR	AIC	$\phi_{Reasoning}$	$\phi_{Speededness}$
1.70	.053	.07	3110.27	-	-
1.98	.063	.09	3131.82	.122***	-
2.00	.063	.09	3133.82	.122***	.001
1.54	.046	.08	3092.38	.107***	.068***
1.40	.040	.06	3914.40	-	-
1.34	.037	.07	3901.65	.077***	-
1.34	.037	.07	3902.24	.069***	.018
1.23	.030	.06	3892.60	.056***	.058***
2.33	.073	.07	2729.96	-	-
2.99	.089	.11	2768.59	.165***	-
2.99	.089	.12	2764.06	.169***	.054*
1.53	.046	.08	2690.39	.179***	.098***
1.81	.057	.06	2797.96	-	-
1.89	.060	.08	2799.86	.097***	-
1.83	.057	.07	2797.03	.052*	.057*
1.61	.049	.07	2787.73	.046**	.092***

made that the influence of inductive reasoning fades out across the last items of the scale. For the bifactor model with fading out, χ^2/df and RMSEA indicated a good model fit and SRMR an acceptable model fit. Moreover, the fit of this bifactor model was better than the fit of the other models according to the AIC comparison. Variances of both latent variables were statistically significant.

A similar pattern of results as for the previous three subtests was found for *Topologies*. The congeneric model fitted the data better than the essentially tau-equivalent model as indicated by the lower AIC. The bifactor model described the data only better when inductive reasoning was allowed to fade out with the last items. In this case, however, the bifactor model yielded a good model fit according to χ^2/df, RMSEA, SRMR and described the data better than the other models according to the AIC comparison. Both latent variances were statistically significant.

The combination of the four measurement models in one structural equation model with correlations between the four latent variables representing inductive reasoning and correlations between the four latent variables representing the effect of speededness, led to a good model fit, $\chi^2(1156) = 1355.87$, $\chi^2/df = 1.173$, RMSEA = .026, SRMR = .069, AIC = 12302.772. The four latent variables representing inductive reasoning correlated significantly with each other ($.341 < r < .631$). The same was true for the latent variables representing the effect of speededness ($.374 < r < .701$).

Proceeding from this pattern of correlations, we assumed that second-order latent variables representing inductive reasoning and the effect of speededness, respectively, can be derived from the first-order latent variables. The fit of this model, presented in Figure 2, was good, $\chi^2(1160) = 1362.81$, $\chi^2/df = 1.175$, RMSEA = .026, SRMR = .069, AIC = 12301.715. The model fit did not differ significantly from the model with correlated first-order latent variables, $\Delta\chi^2(4) = 6.94$, $p = .14$. Accordingly, the first-order latent variables extracted from the four subscales were successfully summarized by a second-order latent variable representing inductive reasoning and another second-order latent variable representing the effect of speededness.

In a next step, the correlational relationships between age, on the one hand, and inductive reasoning as well as the effect of speededness, on the other hand, were investigated. For this purpose, age was added to the second-order latent-variable model. The model fit was still good, $\chi^2(1207) = 1418.39$, $\chi^2/df = 1.175$, RMSEA = .026, SRMR = .069, AIC = 13012.484. The correlation between reasoning and age was $r = .304$, $p < .001$, while the correlation between the speededness effect and age was $r = .496$, $p < .001$. Fixing the correlation between age and reasoning to the value obtained for the correlation between age and the effect of speededness resulted in a worse model fit, $\Delta\chi^2(1) = 8.68$, $p < .01$, indicating that age was more strongly related to the effect of speededness than to inductive reasoning.

As can be seen from Table 1, the correlation between age and CFT 20-R sum scores (uncontrolled for the effect of speededness) was $r = .45$, $p < .001$. Fixing the correlation between inductive reasoning controlled for the effect of speededness and age to the value obtained for CFT 20-R sum scores significantly worsened the model fit, $\Delta\chi^2(1) = 5.14$, $p < .05$. Thus, as expected the correlation between inductive reasoning and age was overestimated when the effect of speededness was not controlled for.

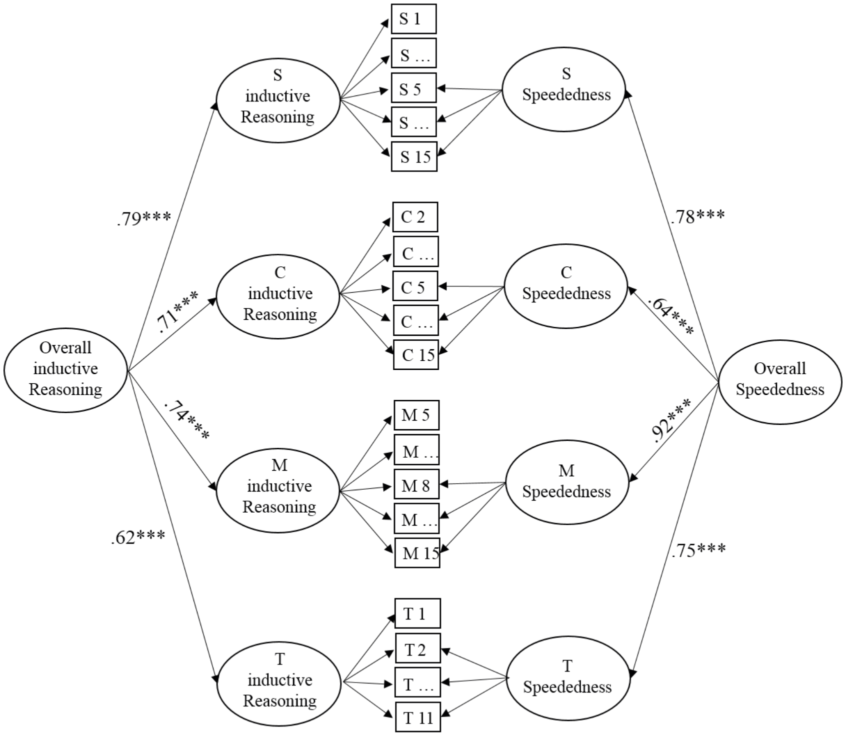

Figure 2 Measurement model on the four CFT 20-R subtests. For each subtest, one latent variable with constant factor loadings (and a fading-out effect at the end) represents inductive reasoning and one latent variable with logistically increasing factor loadings represents the effect of speededness. From both the four inductive-reasoning and the four speededness latent variables second-order latent variables were extracted to represent inductive reasoning and the effect of speededness across the four CFT 20-R subtests. Error variables omitted.
***$p < .001$

In a last step, it was tested how much the CFT 20-R sum score was associated with inductive reasoning and with the effect of speededness and whether the effect of speededness was related to the number of not-reached items. The CFT 20-R sum score was highly related to the latent variable representing inductive reasoning, $r = .76$, $p < .001$, and in a similar way to the latent variable representing the effect of speededness, $r = .67$, $p < .001$. As expected, the number of not-reached responses was highly related to the effect of speededness, $r = -.67$, $p < .001$, but not to inductive reasoning, $r = .05$, $p = .44$.

Discussion

The first research question of the present study was whether the description of 250 children's CFT 20-R data could be improved by the consideration of the effect of speededness. Such an effect could be clearly identified in all four CFT 20-R subscales. When this effect was explicitly represented in a bifactor model, the data description for all four subscales was substantially better than the description by the one-dimensional congeneric model or the essentially tau-equivalent model of measurement. Inductive reasoning, however, was better described by a latent variable with decreasing factor loadings on the last items than by constant factor loadings across all items. This contradicted the assumption of item homogeneity after controlling for the effect of speededness.

The second research question addressed the relationship between chronological age and inductive reasoning and its changes when the measure of inductive reasoning was statistically controlled for the effect of speededness. The present results showed that the correlational relationship between inductive reasoning and age was substantially reduced when the effect of speededness was controlled for because the effect of speededness was also related to age. Thus, the effect of speededness in time-limited tests leads to an overestimation of the relationship between inductive reasoning ability and chronological age of children in the age range from eight to about 13 years.

The assumption that the latent variable with increasing factor loadings represented the effect of speededness was corroborated by the fact that it was closely related to the number of not-reached items. Since this latent variable was not extracted exclusively from omissions, however, it also covered the increasing number of incorrect responses because of speeded and superficial performance as well as guessing on the last items due to limited testing time. Thus, such a latent variable is probably better able to depict the effect of speededness than a latent variable extracted exclusively from omissions or a manifest sumscore of omissions. Both these alternatives would fail to consider that test speededness might not only lead to omissions at the end of scale but alternatively to incorrect responses due to superficial processing and guessing.

The presence of an effect of speededness in combination with the finding that inductive reasoning seems to fade out across the last items emphasizes that (for a given time limit and a given sample of participants) the last items of a scale are less informative for the estimation of the inductive reasoning ability. The CFT 20-R aims to assess fluid intelligence in children and adolescents from 6 to 17 years as well as in young and older adults. It might be assumed that older adolescents or young adults would have reached more items due to more highly developed inductive reasoning and higher processing speed. Consequently, it should be expected that the logistic function of increasing factor loadings of the latent variable representing the effect of speededness would be shifted to later items. Concurrently, the fading-out effect of factor loadings of the latent variable representing inductive reasoning would be postponed to the end of the scale or even completely disappear in such samples. From this point of view, the factorial structure of a time-limited test depends on the extent of the time limit but also on the inductive reasoning ability as well as the processing speed of the sample under investigation.

Limiting the testing time might be necessary for practical reasons (primarily in the context of research). In this case, however, the obtained data should be interpreted carefully and consider the effect of speededness. In the present sample, the CFT 20-R sumscore correlated similarly highly with the second-order latent variable representing the effect of speededness as with the second-order latent variable representing inductive reasoning. Thus, the impact of the speededness effect on the estimation of children's IQ was similarly strong as the impact of inductive reasoning. For research issues and testing of groups of participants, the present study results demonstrate how the effect of speededness can be statistically controlled for. Individual assessments of inductive reasoning in a practical context, however, might benefit from using untimed tests. Adaptive testing might be of particular interest for this purpose to assess inductive reasoning without time limit and, concurrently, reduce the testing time by using a lower number of highly informative items (Kubinger & Holocher-Ertl, 2014).

Considering the effect of speededness in the assessment of inductive reasoning ability is not only an issue of data description but may be of particular importance for research on inductive reasoning per se and its development. The well-known correlational relationship between inductive reasoning and chronological age (Csapó, 1997; Molnár et al., 2013) was clearly overestimated in the present study when the effect of speededness was not considered. Moreover, the relation between age and the effect of speededness was stronger than the relation between age and inductive reasoning. This relationship between the effect of speededness and age is consistent with Kail's (2000) notion that the development of processing speed is not limited to specific processes but rather domain general. Apparently, also in the sense of test-taking speed, processing speed shows a strong developmental course – at least in the age range from 8 to about 13 years in the present study. Nevertheless, inductive reasoning still significantly increased with increasing age after being controlled for the effect of speededness. This was expectable since this relationship has been established also with untimed tests of inductive reasoning (e.g. Fry & Hale, 1996). There is, however, some evidence that this relationship is particularly large when time-limited tests are used (Nettelbeck & Burns, 2010) – probably due to uncontrolled effects of speededness.

Furthermore, many models on cognitive development assume that the development of processing speed is an important source for the development of fluid intelligence (Demetriou et al., 2014; Fry & Hale, 1996, 2000; Kail, 2000). Most of the studies investigating the relationship between processing speed and fluid intelligence used inductive reasoning scales. In the case of time-limited test administration, however, the correlational relationship between processing speed and inductive reasoning can be expected to be overestimated due to the effect of speededness in the measure of inductive reasoning (Wilhelm & Schulze, 2002). With the present approach, it might be easier for future studies to use time-limited tests of inductive reasoning, which might be beneficial for the testing of large samples, and, concurrently, to statistically control for the effect of speededness in order to avoid an overestimation of the relation between inductive reasoning and processing speed.

To sum up, each of the four subscales of the CFT 20-R was best described by a bifactor model, which considered the increasing influence of the effect of speededness on the items' variance and, concurrently, allowed for a decreasing effect of inductive reasoning on the last items of the subtests. The resulting latent variables could be successfully combined to two second-order latent variables representing inductive reasoning ability and the effect of speededness, respectively. Age was closely related to both inductive reasoning ability as well as to the effect of speededness indicating that the relationship between chronological age and inductive reasoning is overestimated when the measure of inductive reasoning is blurred by the effect of speededness.

References

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

Cattell, R. B., & Cattell, A. K. S. (1963). *Culture fair intelligence test*. Champaign, IL: Institute for Personality and Ability Testing.

Csapó, B. (1997). The development of inductive reasoning: Crosssectional assessments in an educational context. *International Journal of Behavioral Development, 20*, 609–626. doi: 10.1080/016502597385081

Demetriou, A., Spanoudis, G., Shayer, M., van der Ven, S., Brydges, C. R., Kroesbergen, E., . . . Swanson, H. L. (2014). Relations between speed, working memory, and intelligence from preschool to adulthood: Structural equation modeling of 14 studies. *Intelligence, 46*, 107-121. doi: 10.1016/j.intell.2014.05.013

Engel de Abreu, P. M. J., Conway, A. R. A., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence, 38*, 552-561. doi: 10.1016/j.intell.2010.07.003

Estrada, E., Roman, F. J., Abad, F. J., & Colom, R. (2017). Separating power and speed components of standardized intelligence measures. *Intelligence, 61*, 159-168. doi: http://dx.doi.org/10.1016/j.intell.2017.02.002

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*, 237-241. doi: 10.1111/j.1467-9280.1996.tb00366.x

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology, 54*, 1-34. doi: 10.1016/S0301-0511(00)00051-X

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179-203. doi: 10.1016/0160-2896(84)90008-4

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*, 107-129. doi: 10.1016/0001-6918(67)90011-X

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

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109*, 490-501. doi: 10.1037/0033-2909.109.3.490

Kail, R. (2000). Speed of information processing: Developmental change and links to intelligence. *Journal of School Psychology, 38*, 51-61. doi: 10.1016/S0022-4405(99)00036-9

Kenny, D. A. (2015). Measuring model fit. Retrieved December 17, 2019, from http://www.columbia.edu/~mrl23/research.html

Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of Educational Research, 78*, 85–123. doi: 10.3102/0034654307313402

Kubinger, K. D., & Holocher-Ertl, S. (2014). *Adaptives Intelligenz Diagnostikum 3*. Göttingen, Germany: Beltz Test Gesellschaft.

Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement, 26*, 29-37. doi: 10.1111/j.1745-3992.2007.00106.x

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. doi: 10.1016/j.tsc.2013.03.002

Nettelbeck, T., & Burns, N. R. (2010). Processing speed, working memory and reasoning ability from childhood to old age *Personality and Individual Differences, 48*, 379-384. doi: 10.1016/j.paid.2009.10.032

Raven, J. C. (1962). *Advanced progressive matrices*. London: Lewis & Co.

Ren, X., Schweizer, K., Wang, T., & Fen, X. (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

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36. doi: https://doi.org/10.18637/jss.v048.i02.arXiv:arXiv:1501.0228.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23-74.

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

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

Schweizer, K. (2013). A threshold-free approach to the study of the structure of binary data. *International Journal of Statistics and Probability, 2*, 67-75. doi: 10.5539/ijsp.v2n2p67

Schweizer, K., Reiß, S., Ren, X., Wang, T., & Troche, S. J. (2019). Speed effect analysis using the CFA framwork. *Frontiers in Psychology, 10*, 239. Retrieved, doi:10.3389/fpsyg.2019.00239

Schweizer, K., Reiss, S., & Troche, S. (2019). Does the effect of a time limit for testing impair structural investigations by means of confirmatory factor models? *Educational and Psychological Measurement, 79*, 40-64. doi: 10.1177/0013164418770824

Schweizer, K., & Ren, X. (2013). The position effect in tests with a time limit: the consideration of interruption and working speed. *Psychological Test and Assessment Modeling, 55*, 62-78.

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. J., & Rammsayer, T. H. (2018). Does processing speed exert an influence on the special relationship of fluid and general intelligence? *Personality and Individual Differences, 131*, 57-60. doi: https://doi.org/10.1016/j.paid.2018.04.022

Swanson, H. L. (2008). Working memory and intelligence in children: What develops? *Journal of Educational Psychology, 100*, 581-602. doi: 10.1037/0022-0663.100.3.581

Troche, S. J., Wagner, F. L., Schweizer, K., & Rammsayer, T. H. (2016). On the structural validity of the Culture Fair Test. *European Journal of Psychological Assessment*. doi: 10.1027/1015-5759/a000384

Weiß, R. H. (2006). *CFT 20-R - Grundintelligenztest Skala 2*. Göttingen: Hogrefe.

Wilhelm, O., & Schulze, R. (2002). The relation of speeded and unspeeded reasoning with mental speed. *Intelligence, 30*, 537–554. doi: https://doi.org/10.1016/S0160–2896(02)00086–7.

Zeller, F., Reiss, S., & Schweizer, K. (2019). Does speededness is collecting reasoning data lead to a speed factor? . *European Journal of Psychological Assessment, Advanced online publication*. doi: 10.1027/1015-5759/a000498

Corresponding author:
Natalie Borter
Institute of Psychology
University of Bern
Fabrikstrasse 8
CH-3012 Bern
Switzerland
natalie.borter@psy.unibe.ch
Phone: +41 (0) 31 631 33 58