Psychological Test and Assessment Modeling, 2023.65:354-387

DOI: https://doi.org/10.2440/001-0004

Conceptualization of the Reasoning-Test "Numerical Topologies"

Klaus D. Kubinger & Nina Heuberger

University of Vienna

Abstract: With reference to Kubinger (2023a) and his six categories of reasoning tests this paper suggests a test conceptualization as concerns the *fluid* facet and *numerical* contents. Basically it meets the concept of a (mathematical) topology, dealing with the (de-) formation of geometric objects in metric spaces. Each item contains three two-dimensional tableaus where several figures (numbers) are arranged in a structural way with respect to their (absolute and with reference to their mutual relative) location and orientation. In the right tableau, at a specific position a figure (number) is missing which has to be deduced (and finally filled in) by discovering the underlying regularities and logically compelling connections of the figures (numbers) over all three tableaus. A first draft of the test *Numerical Topologies* with two times 32 items has been psychometrically analyzed. Actually, the Rasch model holds *a-posteriori*, that is after deleting a few items – retrospectively their misfits can be elucidated by construction flaws. In the second instant by means of LLTM (linear logistic test model) it could be proven that the hypothesized elementary cognitive operation components and item generating rules, respectively, do indeed explain the items' difficulties in a sufficient manner.

Keywords: reasoning, item generating rules, elementary cognitive operations, Rasch model, LLTM

Author Note

Prof. Klaus D. Kubinger, PhD, MSc. c/o University of Vienna, Faculty of Psychology. klaus.kubinger@univie.ac.at

Introduction

Kubinger (2023a) systemizes with reference to *Raymond B. Cattell* and *Adolf O. Jäger* potential reasoning tests in a two times three classification of *fluid* vs. *crystallized* facets and *lexical* vs. *numerical* vs. *figural* contents – defining reasoning as the “ability to realize regularities and logically compelling connections in order to put in appropriate use” (Kubinger, 2019, p. 244; translation by the authors).

As concerns the combination of the *fluid* facet with *numerical* contents, however, there is no test at a practitioner’s disposal. That is not surprising as handling numbers is traditionally grounded in *crystallized intelligence*. Nevertheless, even if numbers are intended to function only as different figures and therefore respective tasks intrinsically regards *fluid intelligence* the question arises whether these tasks refer to the same aptitude as tasks with figures of hardly any educational meaning. The phenomenon of arithmophobia (numerophobia; fear of numbers) raises alone doubt. Such a reasoning test seems indeed necessary, especially when the ability is requiered to arrange numbers (as figures) in some two-dimensional plane as for instance on a screen of common (computer) applications: Many computer-based administration tasks referring to fill out a form require corresponding actions – apart for instance from installing some electrical/electronical distribution box or the like.

In this paper we are now giving a test concept for measuring *fluid-numerical* reasoning.

Method

The basic idea of this test concerns (mathematical) topology, that is in simple terms, the (de-) formation of geometric objects in metric spaces. Each item contains three two-dimensional tableaus where several figures (numbers) are arranged in a structural way with respect to their (absolute and with reference to their mutual relative) location and orientation. In some ways similar to the traditional test principle of figure or number sequencing the task is to find that figure (number) in the third tableau which is according to both the other tableaus at a specific position logically missing. The answer is to be filled in by pencil which means there is a free response format – for the time being, in the form of a paper-pencil test. Figure 1 shows introduction item b) of a first draft of the test *Numerical Topologies*. In the first two tableaus from the left, in the middle (considered vertically as well as horizontally) the number “5” occurs; consequently, at the same position which is signalized by a rectangular frame in the third, the right tableau, “5” is missing – all other presented symbols are of no relevance.

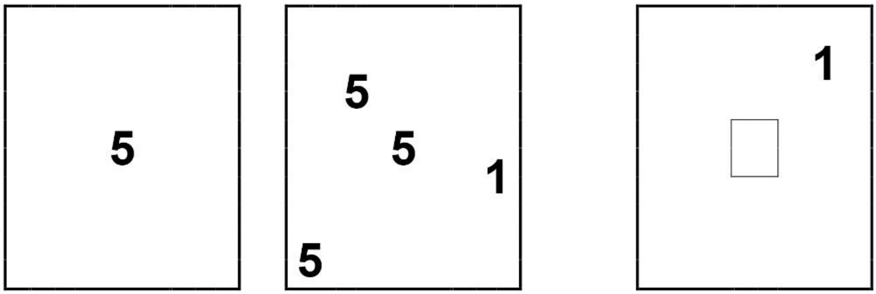


Figure 1
Introduction item b) of the first draft of the test Numerical Topologies. That figure (number) is to fill into the frame which fits the underlying topological structure according to reasoning. The solution is "5".

The items are constructed by using item generating rules; i.e. the underlying logic of the topologies in the three tableaus with respect to the missing figure has been scheduled in advance. Such item generating rules concern basic operations, mathematical operations, spatial relations, and perceptual complexity. Each of those categories refer to several elementary cognitive operations, so-called radicals. These are listed in the following. It should be noted that an item may consist of more than a single radical from the given list. In the first draft of the test not only numbers are presented figures but letters and geometric objects were used as well. Therefore some rules will not apply when the test is actually focussed on numerical topologies. Note, that for the moment there are also items which require arithmetical operations such as addition and multiplication. Of course, such items contradict the conceptualization of a fluid intelligence test.

Basic operations

1) Continuation: the missing figure (number) occurs in both the first tableaus and has to be continued in the third tableau
2) Equal distribution: in each of the first two tableaus there is an (ocassionally different) figure (number) equally often presented; in the third tableau a (maybe new) figure (number) is missing in order to make its occurance there equal frequent according to the other tableaus

Mathematical operations

3) Sequence:
    a. The missing figure (number) continues a logical sequence of figures (numbers) across the tableaus from left to right
    b. The missing figure (number) continues a logical sequence of figures (numbers) across the tableaus from right to left
    c. There is within each tableau a logical sequence of figures (numbers) forwards
    d. There is within each tableau a logical sequence of figures (numbers) backwards
    e. The figures are numbers
    f. The figures are letters
4) Addition: The missing number results as the sum of the third tableau’s numbers, just like it happens in both of the first tableaus
5) Multiplication: The missing number results as the arithmetic product of the third tableau’s numbers, just like it happens in both of the first tableaus

Spacial relations

6) Rotation: The missing figure (number) has to be filled in rotated 90 or 180 degrees
7) Absolute position: The position of the missing figure (number) corresponds exactly with the position of the logically relevant figures (numbers) in both the left tableaus
8) Relative position: The position of the missing figure (number) corresponds in relation to a certain other figure (numbers) in the same way as it is the case in both the first tableaus

Perceptual complexity

9) Irrelevances:
    a. Some combination of figures (numbers) in each tableau pretend a rule relevant for the solution, while in fact they are irrelevant
    b. Specific combinations of figures (numbers) within the three tableaus distract from the logically relevant figures
    c. There are the same figures (numbers) in all the three tableaus only to serve as fillers
10) Amount of figures: The amount of figures (numbers) within each tableau in average (either 3 or 4 to 5 or 6 to 7 or 8 and more)
11) There is some partial relevant rule which applies in both the left tableaus
12) There is some partial relevant rule which applies only in one of the both left tableaus

While the introduction item b) in Figure 1 only refers to radical 1) Continuation and 7) Absolute position, Figure 2 shows item 26 which refers to quite more radicals: These are 3)c. "There is within each tableau a logical sequence of numbers forwards", 3)e. "The figures are numbers", 8) "Relative position: The position of the missing number corresponds in relation to a certain other number in the same way as it is the case in both the first tableaus", 10) "Amount of figures is 5 or 6" [counts: 2], and 11) "There is some partial relevant rule which applies in both the left tableaus" (the numbers "2", "3", and "4" being there). The solution is"4".

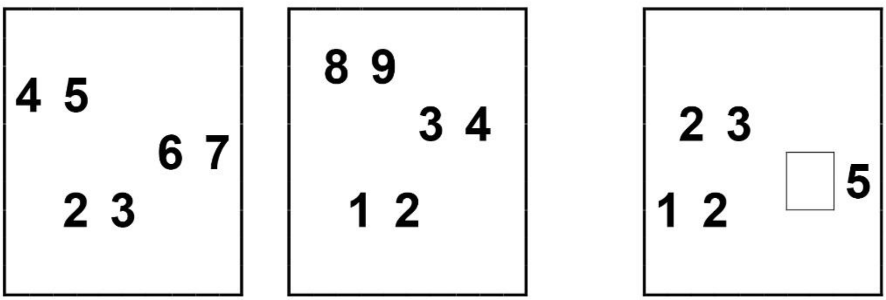


Figure 2
Item 26 of the first draft of the test Numerical Topologies. The solution is "4".

At the end, the first draft of the test has 32 items, each of them evolved in two parallel form versions. For instance item 26, used explicitly in parallel form A (see in Fig. 2), looks in the version for parallel form B as shown in Figure 3. From the test constructor's point of view both versions differ only in a way that is irrelevant to the solution.

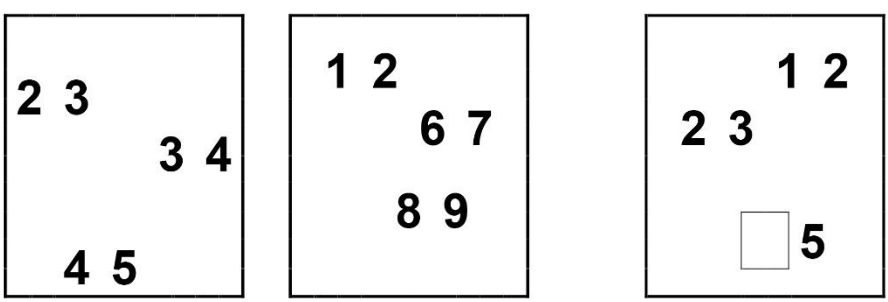


Figure 3
Item 26 of the first draft of the test Numerical Topologies as converted for parallel form B. The solution is "4".

The matrix of items and radicals are given in Table 1. The items themselves, if not already presented or not presented in the following, are listed in the appendix (the remaining introduction items included).

Table 1

The elementary cognitive operations (radicals) used for the item construction of the first draft of the test Numerical Topologies (32 items). In case a certain radical applies for an item, there is entered "1" (if applied twice, then: "2"; as concerns the radical "amount of figures" this is graded with "0", "1", "2", or "3"; for simplicity, empty cells mean: "0").

	Basic operations	Basic operations	Mathematical operations Sequence	Mathematical operations Sequence	Mathematical operations Sequence	Mathematical operations Sequence	Mathematical operations Sequence	Mathematical operations Sequence	Mathematical operations	Mathematical operations	Spacial relations	Spacial relations	Spacial relations	Perceptual complexity Irrelevances	Perceptual complexity Irrelevances	Perceptual complexity Irrelevances	Perceptual complexity	Perceptual complexity	Perceptual complexity
Item No.	Continu-ation	Equal distribution	across the tableaus left to right	across the tableaus left to right	within each tableau forwards	within each tableau backwards	numbers	letters	Addition	Multi-plication	Rotation	Absolute position	Relative position	same in each tableau	specific within each tableau	Filling number s	Amount of figures	partial relevant rule in both left tableaus	partial relevant rule in one left tableau
9-1A	1	1																	
9-1B	1	1																	
9-2A	1		1																
9-2B	1		1																
9-3A			1	1									1				2	1	
9-3B			1	1									1				2	1	
9-4A											1		1		1				
9-4B											1		1		1				
1A		1																	
1B		1																	
2A	1	1														1	1		
2B	1	1														1	1		
3A	1											1			2		1		
3B	1											1			2		1		
4A		1	1					1							1		1		
4B		1	1					1							1		1		
5A		1														1	1		
5B		1														1	1		
6A	1												1	1			1		
6B	1												1	1			1		
7A	1													1	1		1		
7B	1													1	1		1		
8A	1												1		1		2	1	
8B	1												1		1		2	1	
9A			1				1					1		3			2		
9B			1				1					1		3			2		
10A	1													3			1		
10B	1													3			1		
11A						1	1						1				1		
11B						1	1						1				1		
12A	1												1	2			2		
12B	1												1	2			2		

13A					1			1					1				2	1	
13B						1		1					1				2	1	
14A	1														4		2		
14B	1														4		2		
15A	1												1	1			2		
15B	1												1	1			2		
16A										1					1		1		
16B										1					1		1		
17A											1			2			2		
17B											1			2			2		
18A	1													2			1		
18B	1													2			1		
19A					1		1						1				2	1	
19B					1		1						1				2	1	
20A	1												1						1
20B	1												1						1
21A					1		1					1					2		
21B					1		1					1					2		
22A									1			1							
22B									1			1							
23A				1			1						1		4		2		
23B				1			1						1		4		2		
24A					1		1					1					2		
24B					1		1					1					2		
25A										1							2		
25B										1							2		
26A					1		1						1				2	1	
26B					1		1						1				2	1	
27A	1												1				2		1
27B	1												1				2		1
28A											1		1				3	1	
28B											1		1				3	1	

From the psychometric point of view two issues are of interest. First, do such items measure uni-dimensional? For this the Rasch model is a proper means. Second, if data prove that the Rasch model holds (at least *a-posteriori*, after the deletion of very few items), the question is whether the elementary cognitive operations components (radicals) do indeed explain the items' difficulties in a sufficient manner. This can be tested best by the so-called LLTM (linear logistic test model; Fischer, 1973; see also Fischer, 2005, as well as Kubinger, 2008, 2009). This model decomposes the Rasch model's item difficulty parameters σ_i, $i = 1, 2, \ldots k$ by using a linear combination of the respective elementary (cognitive) operation parameters η_j ($j = 1, 2, \ldots p < k$); that is $\sigma_i = \sum_j^p q_{ij}\eta_j$, where q_{ij} weighs these operation parameters according to the hypothesized radicals. Here, the so-called structure matrix $((q_{ij}))$ corresponds to the entries in Table 1. If the elementary operation parameters actually sufficiently explain the items' difficulties this means: First, the test's validity is definitely given because passing the test confirms a testee's ability to apply these cognitive operations components properly. Without such a confirmation, however, it is rather ambiguous which specific ability the test measures. Second, these cognitive operations components constitute in various combinations some item universe, that is they might be used by test constructors to compose new items to achieve a specific item difficulty – this being of big use for adaptive testing (cf. Kubinger, 2016).

For the psychometric analyses, the item pool of two times 32 items were arranged in two times two different test-booklets with 28 items each. There are four very easy items which should not be administered to testees older than nine years, and there are four rather difficult items which could be ignored for testees younger than ten years. Nevertheless, all the versions (parallel forms A and B up to 9 years; parallel forms A and B from 10 years on) match a connected incomplete block design (cf. Rasch, Kubinger, & Yanagida, 2011) as required in order to make parameter estimations possible for the item pool as a whole (cf. e.g. Kubinger, 2008). The tested sample were 423 pupils of school in (Upper) Austria in between the ages of seven and 14 years. After a very detailed instruction with five instruction items the testees had 25 minutes to work through the test. Actually, that time limit proved suitable or even rather generous.

Analyses were executed by the R-package eRm (Mair, Hatzinger, Maier & Rusch, 2015). Testing the Rasch model's validness in accordance with the state of the art (cf. Kubinger, 2005) focusses on Andersen's Likelihood-ratio test (LRT) with several partition criteria of the given total sample into disjoint subsamples (here 1. score: "high-" *vs.* "low-scorers", that is the partition in testees with a high number of solved items *vs.* testees with a low number of solved items; 2. sex: male *vs.* female testees; 3. age: testees up to eleven years *vs.* testees twelve years and up; 4. first language: German *vs.* not German). If any LRT results in significance (comparison-wise type-I-risk $\alpha = .01$ – running four comparisons this means a study-wise type-I-risk of approximately $\alpha = .04 < \alpha = .05$), items should be deleted step by step by repeating the LRT until it results for each partition criterion in non-significance – or non-conformity of the items with the Rasch model must be stated (according to a rule of thumb by Kubinger & Draxler, 2007, the latter is the case, if more than 10% of the items have to be deleted). Deletion of an item is done according to Rasch’s graphical model check, which illustrates the coincidences of the item parameter estimations when based on different subsamples: differences of any item parameter estimation between two subsamples larger than a tenth of the parameters’ range indicates model misfit (cf. again Kubinger, 2005). Given the Rasch model holds, at least *a-posteriori* after deleting only a few items, the LLTM can also be tested by an asymptotically χ^2-distributed Likelihood-ratio test (cf. e.g. Kubinger, 2008): The null-hypothesis states that the observed data (an item is solved or not solved) can be explained as well by LLTM’s elementary (cognitive) operation parameters as by the Rasch model’s item difficulty parameters.

Results

Table 2 summarizes the results of Andersen’s Likelihood-ratio test (LRT) with respect to the four partition criteria mentioned above.

Table 2
The Rasch model tests for two times 32 items of the first draft of the test Numerical Topologies. For the applied four criteria of partitioning the calibration sample the results of the asymptotically χ^2-distributed Andersen’s Likelihood-ratio test statistic (LRT) are given as well as the degrees of freedom (df) and the respective p-value – if any item within a certain subsample is solved either by all testees or by none, that item is not included, as a consequence df is reduced. The results are based on 423 testees.

partition criterion	χ^2	*df*	*p*
score	92.25	58	.003
sex	51.05	62	.838
age	57.37	54	.351
first language	78.77	62	.704

Only the partition criterion score resulted in a significant LRT. The graphical model check in Figure 4 reveals no eminently misfitting item – there the item parameter estimations within the respective two subsamples (“high-“ *vs.* “low-scorers“) are opposed in a Cartesian coordinate system; ideally, only dots lying on a 45° line which meets the origin would result because, given Rasch model’s validness, each item has the same parameter (estimation) in every subsample of testees. In our case, item 27 (in parallel form B) fits worst and therefore has been deleted in the first step (as well as its version in parallel form A, though there it was not conspicuous in the graphical model check). In the next step item 20 (due to parallel form A, while it was not conspicuous in parallel form B), and finally, before no more significant LRT resulted, item 11 (due to parallel form A, while it was not conspicuous in parallel form B) had to be deleted. The LRTs' exact results with respect to the four partition criteria are summarized in Table 3.

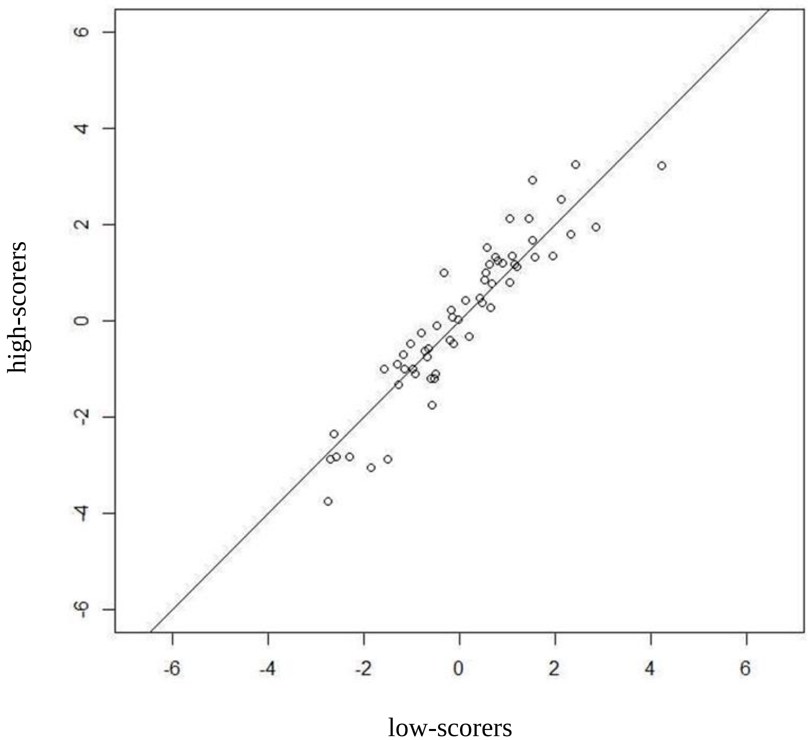


Figure 4
Graphical model check for 59 items of the two times 32 item pool of the first draft of the test Numerical Topologies – item (difficulty) parameter estimations according to the Rasch model as opposed for trainees with a high score (ordinate) and for trainees with a low score (abscissa).

Table 3
The Rasch model tests for the remaining two times 29 items of the first draft of the test Numerical Topologies. For the applied four criteria of partitioning the calibration sample the results of the asymptotically χ^2-distributed Andersen's Likelihood-ratio test statistic (LRT) are given as well as the degrees of freedom (df) and the respective p-value – if any item within a certain subsample is solved either by all testees or by none, that item is not included, as a consequence df is reduced. The results are based on 423 testees.

partition criterion	χ^2	*df*	*p*
score	62.31	51	.133
sex	47.86	57	.800
age	45.19	49	.629
first language	60.21	57	.361

Apart from the realization that item 27 proves to be more difficult (in relation to the other items) for "high-scorers" than for "low-scores" (as indicated by the graphical model check), this item's misfit can be blamed on a construction flaw. The item refers to the radical 12) "There is some partial relevant rule which applies only in one of the both left tableaus", which finally means the item solution is not unequivocal (see Fig. 5). The item originally (e.g. as concerns parallel form A) has been designed as follows: Starting from the third, the right tableau the neighbouring numbers 1 – 8 complete a sequence from the first, the left tableau, i.e. 3 – 8, 5 – 8, 7 – 8; and similarly the numbers 8 – 3 complete the sequence 2 – 3, 4 – 3, 6 – 3. Now, within the right tableau there is the pairing 5 – ? and when looking at the center tableau it becomes apparent that 5 – 2 is missing within the sequence (1 – 2), 4 – 2, 6 – 2, 7 – 2, as a consequence of which "2" is the solution of the item. However, the number "8" also represents a logical answer. This is due to the fact that in the left tableau two sequences of numbers occur, namely 3, 5, 7 and 2, 4, 6, the first of which is always paired with the number "8" and the second always with the number "3". As in the right tableau there is the sequence 1, 3, 5, the paired number of them being "8" twice, the missing number may be "8" as well – ignoring the wrong positioning of the number "8" in one case. Of course, the existence of a two-way solution should be avoided in the future. Actually item 20 faces the same problem; it is the only other item which refers to the radical 12) (see Fig. 6). The originally intended solution is a circle, because there is a vertical line located under a symbol only for that symbol. However, recognizing that in both, the left and the center tableau there are several symbols locationally paired with a vertical line but never another vertical line, a second vertical line as the solution is compelling rational; or completely ignoring the left tableau (which is quite conceivable according to the conceptualization of radical 12)) leads to the answer plus sign. Apart from the fact that item 20 does not use numbers but geometric figures and is therefore not suitable for an aimed-for test *Numerical Topologies*, anyway, this finding would best be taken into consideration in the future.

Figure 5
Item 27 of the first draft of the test Numerical Topologies in both versions (parallel form A on the left side, parallel form B on the right side). The originally intended solution is "2".

Figure 6

Item 20 of the first draft of the test Numerical Topologies in both versions (parallel form A on the left side, parallel form B on the right side). The originally intended solution is "circle".

As concerns item 11, again a construction flaw is responsible for an unequivocal item solution (see Fig. 7). The item originally has been designed by the radical 3) c./d. "There is within each tableau a logical sequence of numbers (forwards/backwards)", which is realized in the left as well as in the center tableau in the direction "forwards" while in the right tableau the direction happens to be "backwards". Hence, instead of the conceptualized solution "4" and "3", respectively, always applying the direction "forwards" will lead to the answer "8" and "7", respectively. While in parallel form A this alternative answer might be less convincing to a lot of people, as there the rectangular frame for the missing number is positioned to the right of the incomplete sequence (and therefore more likely imposes the correct answer, i.e. "4"), in parallel form B the rectangular frame is positioned to the left of the incomplete sequence (and therefore compels a lot of people to the answer "7", which is scored as wrong). At any rate, this item is not fair.

Figure 7
Item 11 of the first draft of the test Numerical Topologies in both versions (parallel form A on the left side, parallel form B on the right side). The originally intended solution is "4" and "3", respectively.

To summarize, six deleted items out of 64 result in an acceptable rate of 9.38%. The item parameters for the remaining items lie between -4.16 and 3.88 which, from experience, means a rather great range.

Coincidentally a second study was done in order to test the Rasch model’s validness for the test *Numerical Topologies* (Gamsjäger, 2012)[1]. Because that sample covered 331 police security trainees aged between 18 and 32 years the four easy items from version up to 9 years were not administered. Andersen’s Likelihood-ratio test with respect to the partition criteria score, sex, and age revealed likewise only for the first criterion significance (χ^2 = 80.43, df = 39, p = .000). In the graphical model check item 11 and item 20 again, and additionally item 10 (in parallel form A) proved to be misfitting. Although item 10 (see Fig. 8) essentially refers to letters but not numbers, it serves as an example of another construction flaw. The item originally has been designed by the radicals 1) “Continuation: the missing figure (number) occurs in both the first tableaus and has to be continued in the third tableau” (i.e. [A –] C, [B –] C, [C –] ? or [C –] E, [D –] E, [E –] ?) and 9) a. “Some combination of figures (numbers) in each tableau pretend a rule relevant for the solution, while in fact they are irrelevant” (i.e. 2 – 9, 2 – 8, 2 – 9 or 3 – 1, 3 – 0, 3 – 1). As a consequence the solutions are “C” and “E”, respectively. However, the rule only pretending to be relevant here is yet somehow relevant, because testees may discover a sequence of alternating figures (i.e. [2 –] 9, [2 –] 8, [2 –] 9 or [3 –] 1, [3 –] 0, [3 –] 1) which might be applied to the sequence A [– C], B [– C], [C –] ? with the answer “A” or the sequence C [–E], D [– E], [E –] ? with the answer “C” – though this is not completely logical as thereby the sequence of left and right letters are changed for these answers.

Figure 8
Item 10 of the first draft of the test Numerical Topologies in both versions (parallel form A on the left side, parallel form B on the right side). The originally intended solution is “C” and “E”, respectively.

In a third study (Kresnik, 2012)[2], only using 16 items out of the original parallel forms A and B for 10 year olds (items 11 and 27 included), 292 trainees of either a police security academy or two nursing schools were tested (their age being 18 to 54 years). In this study only item 7 (parallel form B) had to be deleted in order to get a non-

---

[1] This study was carried out for a Master Thesis and supervised by the first author as the responsible university advisor.

[2] This study was carried out for a Master Thesis and supervised by the first author as the responsible university advisor.

significant Andersen's Likelihood-ratio test with respect to all the three applied partition criteria. Figure 9 shows this item, which points at another possible construction flaw. The neighboring digits "1" and "6", as given in the left and the center tableau and also occuring with the intended solution "6" due to radical 1) "Continuation: the missing figure (number) occurs in both the first tableaus and has to be continued in the third tableau" (be aware that then also radical 9) a. applies), may unforeseeably compel a testee to think that rather the number "16" than both the digits' topology relation is relevant. In this case, the testee will discover that within the left tableau 5 plus 16 sums up to 21 and within the center tableau 16 plus 7 sums up to 23, therefore he/she looks for the number which, with 8, will sum up to 25 in order to continue the sequence 21, 23, … Consequently such a testee will find the result 17, leading him/her to the answer "7". Aside from discovering this construction flaw, this item teaches: a test which aims for measuring *fluid* but not *crystallized intelligence* must avoid topologies which gives the impression that arithmetic calculations are necessary – of course, using the natural sequence of numbers should not require a high degree of *crystallized intelligence*.

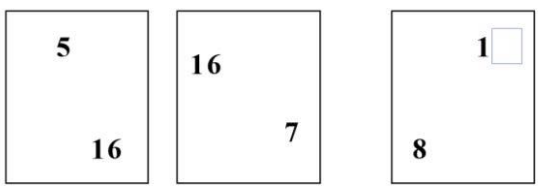


Figure 9
Item 7 of the first draft of the test Numerical Topologies in parallel form B. The originally intended solution is "6".

Finally, in a fourth study (Schnait, 2015) [3], only 15 items out of the original parallel forms A and B were used, items 7, 11, and 27 included. 265 high school students aged between 16 and 21 years were tested. Applying three partition criteria, no significant Andersen's Likelihood-ratio test resulted, at once.

So much for the calibration the test according to the Rasch model.

For testing the hypothesis that the elementary operation parameters sufficiently explain the obtained (Rasch model) item parameters, only the data of the first presented study were taken into account. The result of the above mentioned Likelihood-ratio test is χ^2 = 52.50 (df = 39 – due to 58 item parameters and 18 elementary operation parameters; the 19th parameter in Table 1 has to be canceled when items 20 and 27 have been deleted) which leads to p = .073. That proves that

[3] This study was carried out for a Master Thesis and supervised by the first author as the responsible university advisor.

the applied item generating rules and the accordingly hypothesized elementary (cognitive) operations do actually explain the difficulty of items of *Numerical Topologies*.

Discussion

The conceptualization of the reasoning test *Numerical Topologies* can be, basically, considered usable. Deleting a few items of a first draft did not only result in a Rasch model conforming test but also proved the chosen item generating rules as proper.

Nevertheless a lot of research work is still necessary until such a test can be handed to practitioners for consulting work. Apart from deleting items the (relevant) figures of which are not numbers (cf. the items 9-1 to 9-4, 4, 13, 15, and the instruction item c) in the appendix) and items which obviously demand arithmetical calculations (cf. the items 16 and 25 in the appendix) in particular new generated items have to be constructed more carefully, mindfully and audited with respect to alternative solution processes. That is, a lot of construction flaws have emerged by applying the Rasch model. In order to avoid such flaws during further development of the test *Numerical Topologies* the application of the method of thinking out loud seems imperative. With this exploratory method a subject is asked to work on the items while he/she verbally expresses all thought processes and action strategies. However, other than using just a few subjects (e.g. about 10 as advised by experience) it is highly recommended to use more subjects (maybe up to 100) for this test construction. After respective item revisions, the next step should be some pilot study for a frequency analysis of all the testees' given answers. This will also help to detect unexpected alternative solution processes.

As concerns future psychometric analyses, a "kind of cross-validation" (cf. Kubinger, 2005; see also Rasch, Kubinger, & Yanagida, 2011) of the Rasch model calibration result is highly recommended: that is, if a few items have to be deleted and the remaining final item pool proves to fit the model, then new data should be used in order to test whether the model's validness for that pool can be confirmed – according to Popper's "degree of corroboration/confirmation" (Popper, 2001). For the time being, the second and the third study mentioned already did that, to some extent. By the way, we now suggest an alternative handling of the parallel forms, when testing the Rasch model. Because from the test constructor's point of view both versions of every item only differ in a solution irrelevant manner, each such item pair can also be analyzed as if they were the same item, instead of as two separate items. By applying "parallel form A *vs.* parallel form B" as an (additional) partition criterion for Andersen's Likelihood-ratio test, then an eventually established differential item functioning impact may indicate an unexpected alternative solution process or it may indicate some interfering effect due to a certain item composition (cf. item 11 in Fig. 7, see above). With respect to future LLTM-analyses it seems necessary to use quite more than 30 items, that is rather 100; this is because 18 elementary operation parameters are hypothesized to explain 58 (real 29) item parameters but that small number of the latter supplies for too little information to estimate the former.

Of course, using a free response format but no multiple choice format has the great advantage that lucky guessing hardly takes place. But strictly speaking, as the asked-for solution for the items of the test *Numerical Topologies* is regularly a digit, there is, even with a free response format, still a so-called *a-priori* guessing probability of a tenth. For this it is worthwhile to further consider letters and geometric objects as possible item figures in order to reduce this *a-priori* guessing probability.

The future will show, to what extent the test *Numerical Topologies* finally measures *fluid-numerical* reasoning as intended; the question is, if the test is actually measuring a specific intelligence factor, or if it is instead covered by other factors or even covered by a general reasoning factor. Although the first draft of this test still employs a few times arithmetic computations and thus including the *crystallized* facet, Kubinger (2023b) tentatively answers this question.

References

Gamsjäger, C. (2012). *Entwicklung des Tests „Gleichungen" zur Erfassung von Reasoning* [Construction of the reasoning test „(Mathematical) Equations"]. Unpublished master thesis, University of Vienna, Austria.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37,* 359-374.

Fischer, G. H. (2005). Linear logistic test models. *Encyclopedia of Social Measurement, 2,* 505-514.

Kresnik, Z. J. B. (2015). *Tripelweiser Dimensionalitätsvergleich der Reasoning-Tests Numerische Topologien, Family-Relation-Reasoning-Test und Wiener Matrizen-Test 2* [On the dimensionality of three reasoning tests: Numerical Topologies, Family Relation Reasoning-Test, and Viennese Matrices 2]. Unpublished master thesis, University of Vienna, Austria.

Kubinger, K. D. (2005). Psychological Test Calibration using the Rasch Model - Some Critical Suggestions on Traditional Approaches. *International Journal of Testing, 5,* 377-394.

Kubinger, K. D. (2008). On the revival of the Rasch model-based LLTM: From constructing tests using item generating rules to measuring item administration effects. *Psychology Science Quarterly, 50,* 311-327.

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

Kubinger, K. D. (2016). Adaptive testing. In K. Schweizer & C. DiStefano (eds.), *Principles and methods of test construction* (pp. 104-119). Göttingen: Hogrefe.

Kubinger, K. D. (2019). *Psychologische Diagnostik – Theorie und Praxis psychologischen Diagnostizierens* (3rd ed.) [*Psychological Assessment – Theory and Practice of Psychological Consulting*]. Göttingen: Hogrefe.

Kubinger, K. D. (2023a). Guest Editorial: Promising reasoning test ideas not yet published. Special issue: *Promising reasoning test ideas not yet published. Psychological Test and Assessment Modeling, 65*, 315-320.

Kubinger, K. D. (2023b). On the dimensionality of Reasoning. *Psychological Test and Assessment Modeling, 65*, 437-447.

Kubinger, K. D. & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem Rasch-Modell [Problematic issues when calibrating a psychological test according to the Rasch model]. *Diagnostica, 53*, 131-143.

Mair, P., Hatzinger, R., & Maier M. J. (2015). eRm: Extended Rasch Modeling. 0.15-5. http://erm.r-forge.r-project.org/

Popper, K.R. (2001, reprint). *The logic of scientific discovery*. London: Routledge.

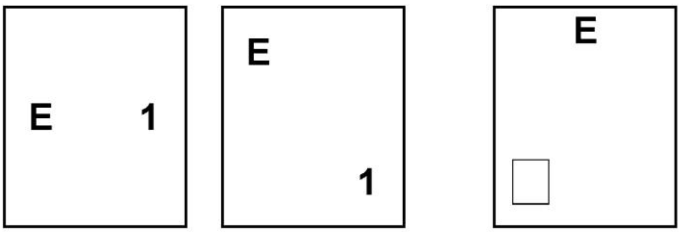
Rasch, D., Kubinger, K.D. & Yanagida, T. (2011). *Statistics in Psychology – Using R and SPSS.* Chichester: Wiley.

Schnait, B. (2015). *Zur Dimensionalität von Reasoning-Tests: Ein Beitrag zur Validierung einer neuen Testbatterie zur Erfassung des logisch-schlussfolgernden Denkens* [On the dimensionality of reasoning tests: first results of a new test-battery’s validity]. Unpublished master thesis, University of Vienna, Austria.

Appendix: The Rasch model conform items of the first draft of the test *Numerical Topologies* (the introduction and those introduction items not given in the text included). The items of parallel form A are given first, followed by that of parallel form B – the version up to 9 years and the version from 10 years are pooled. (It is taken into account, that in the first draft also letters and geometric objects were used).

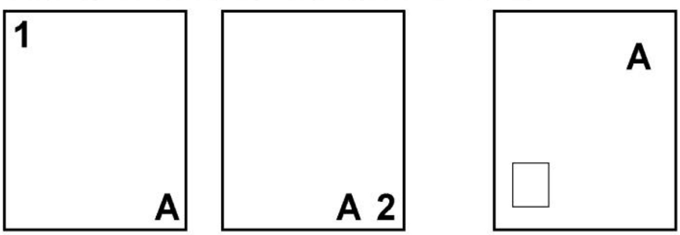
In this test regularities and logically connections should be recognized and applied. The symbol searched for is to be entered in the given box. Five instruction items do demonstrate the task.

Example a)



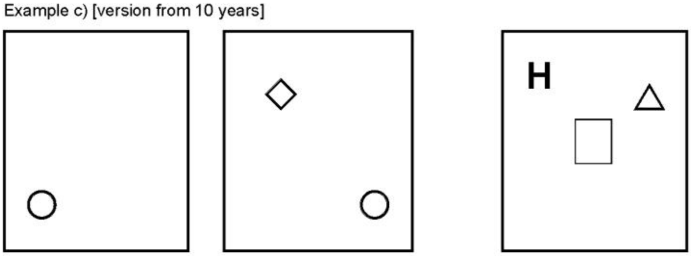
The solution is „1". In every tableau there occurs the symbol „E", in both the left one also the symbol „1", consequently there must also in the third tableau occur: „1".

Example c) [version up to 9 years]; Example d) [version from 10 years]



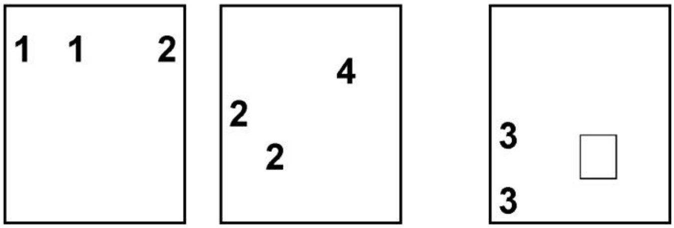
The solution is „3". In every tableau the symbol „A" occurs, in den beiden ersten auch das Symbol „1" bzw. das Symbol „2", also muss folgerichtig im dritten Rahmen das Symbol „3" enthalten sein.

Example c) [version from 10 years]



The solution is „O“. Nothing else is common in both the first, the left tableaus. Consequently, there has to occur „O“ as well in the third tableau. The symbols „◇“, „H“ and „△“ are of no relevance because they do not have any relation; there might be also quite different symbols instead.

Example e) [version from 10 years]

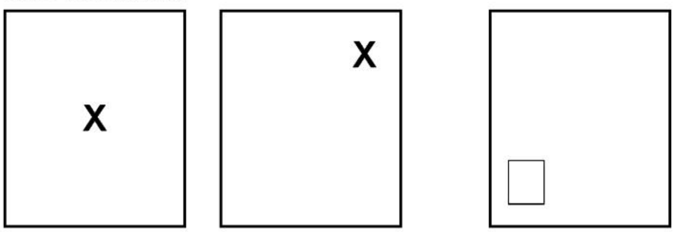


The solution is „6“. In the first tableau there occur rather left-aligned two times the number 1, right of it the sum of them, i.e. 2; in the second tableau there occur rather left-aligned two times the number 2 and right of it the sum of it, i.e. 4. As in the third tableau there occur rather left-aligned two times the number 3, the sum of them, 6, is the missing number.

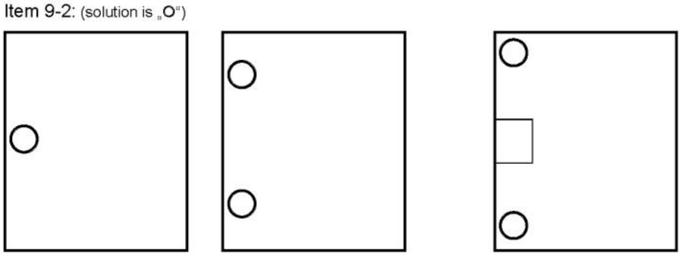
Also in the following items there are certain regularities which determine unequivocally the missing symbols.

Parallel form A

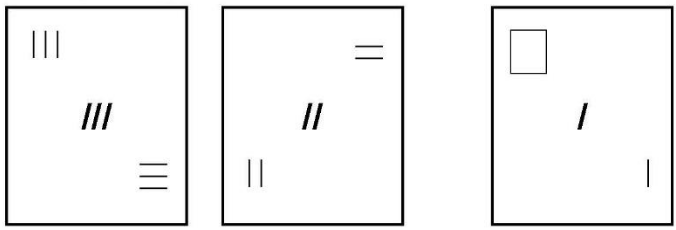
Item 9-1: (solution is „X“)



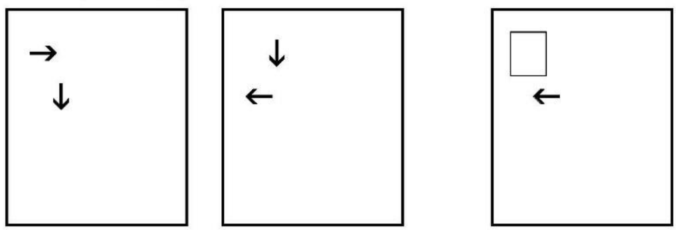
Item 9-2: (solution is „O“)



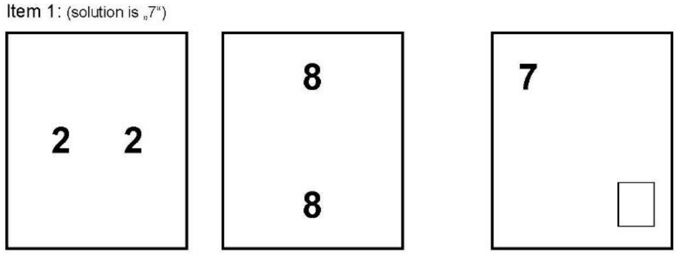
Item 9-3: (solution is „–“)



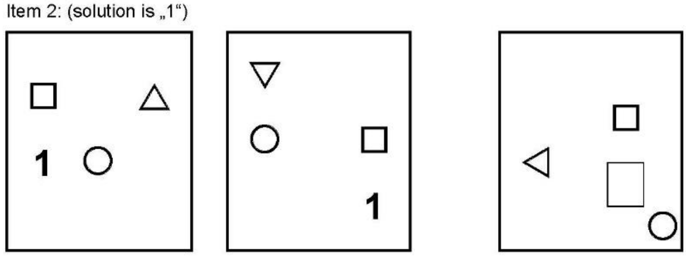
Item 9-4: (solution is „↑“)



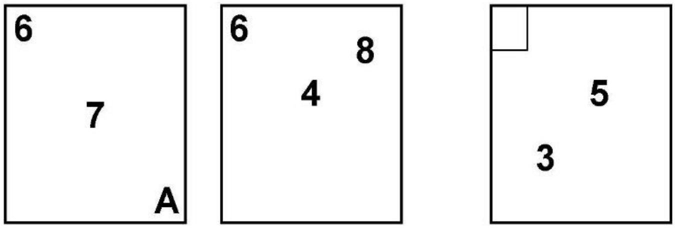
Item 1: (solution is „7“)



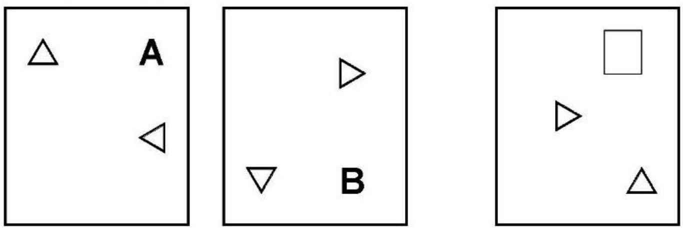
Item 2: (solution is „1“)



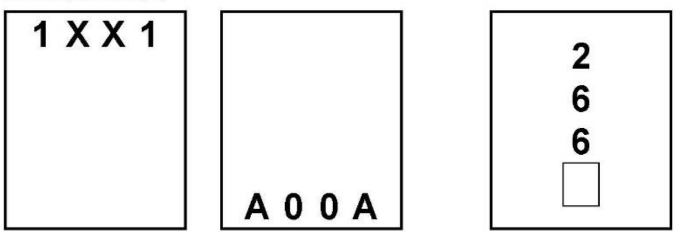
Item 3: (solution is „6“)



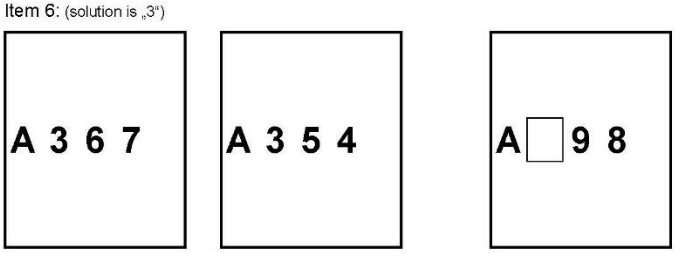
Item 4: (solution is „**C**“)



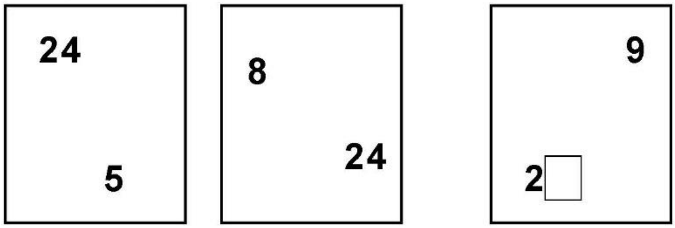
Item 5: (solution is „2“)



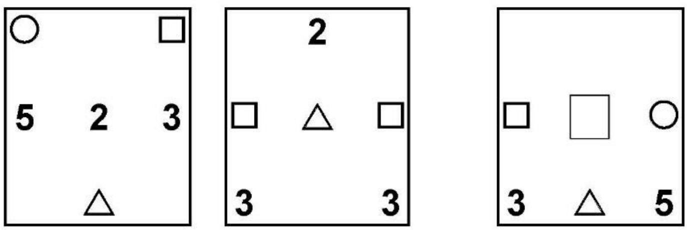
Item 6: (solution is „3“)



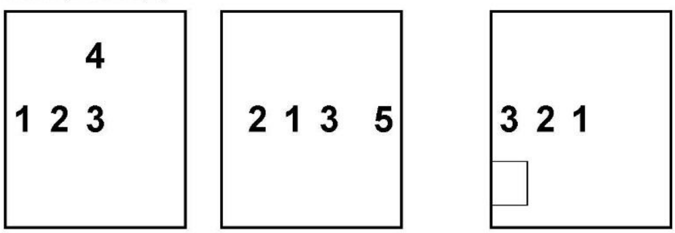
Item 7: (solution is „4")



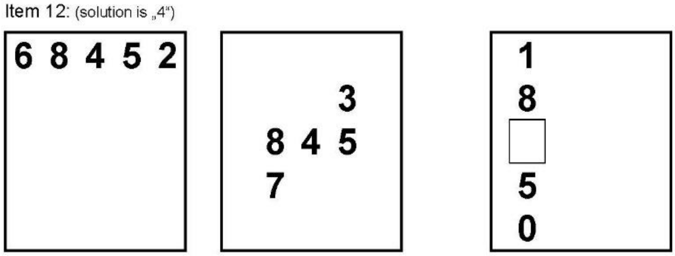
Item 8: (solution is „2")



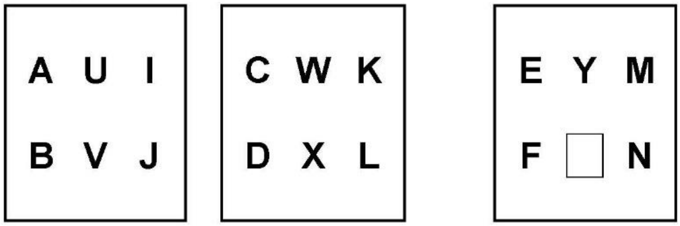
Item 9: (solution is „6")



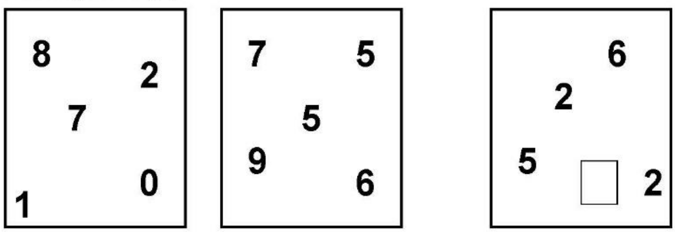
Item 12: (solution is „4")



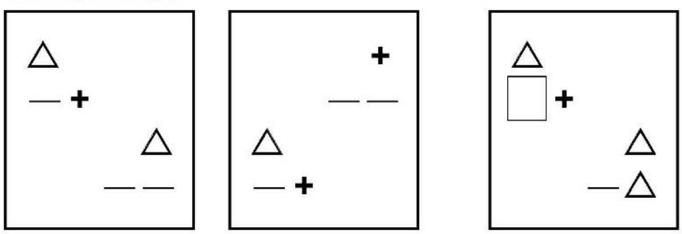
Item 13: (solution is „Z“)



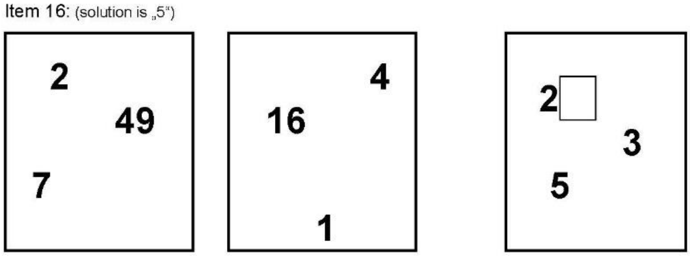
Item 14: (solution is „7“)



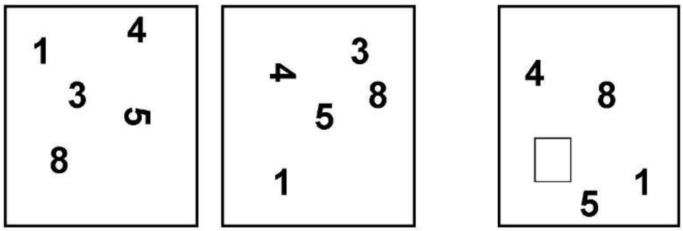
Item 15: (solution is „–“)



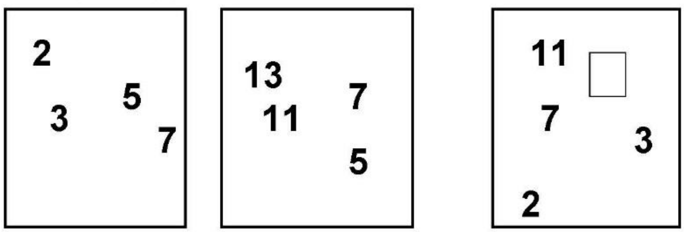
Item 16: (solution is „5“)



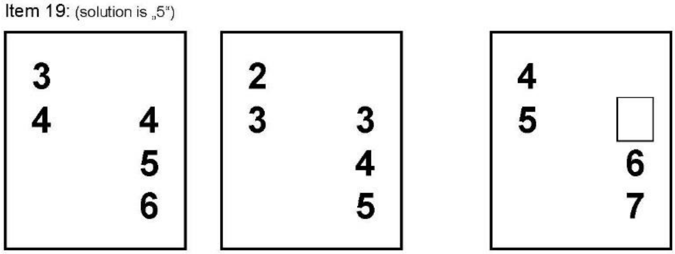
Item 17: (solution is " ω ")



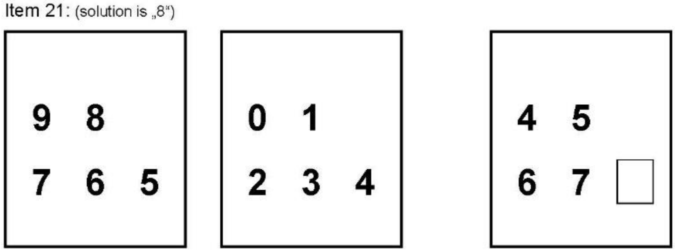
Item 18: (solution is „5")



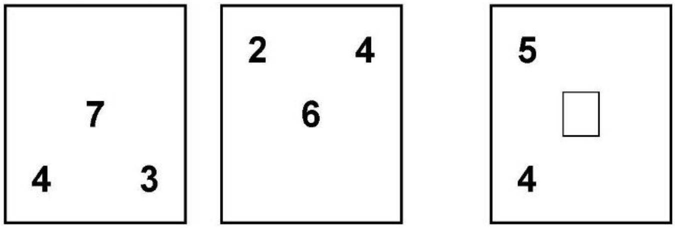
Item 19: (solution is „5")



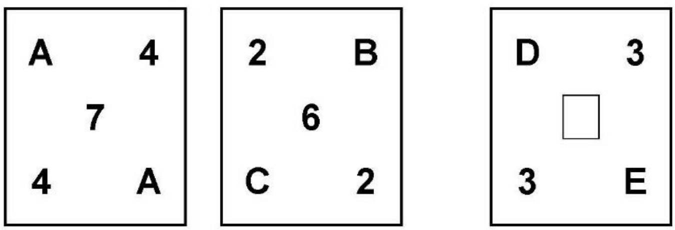
Item 21: (solution is „8")



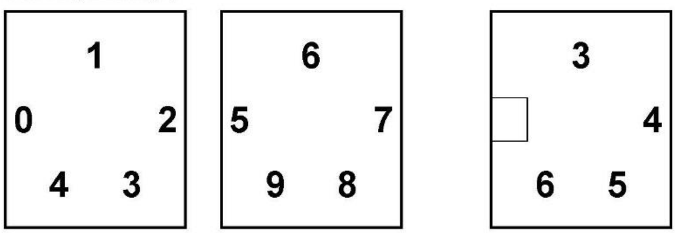
Item 22: (solution is „9")



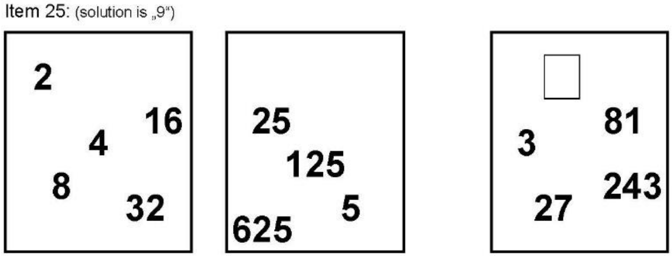
Item 23: (solution is „5")



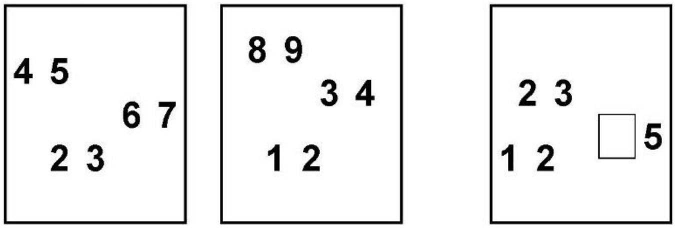
Item 24: (solution is „2")



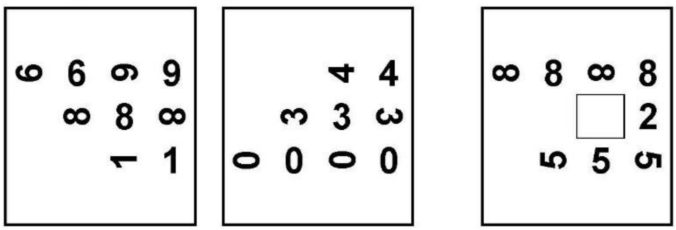
Item 25: (solution is „9")



Item 26: (solution is „4“)

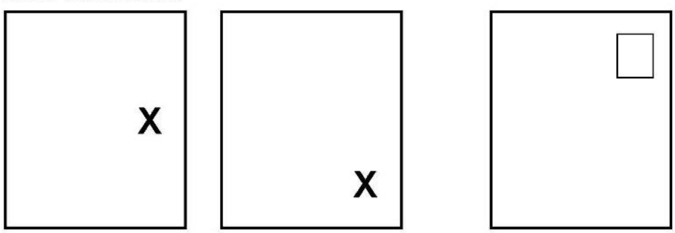


Item 28: (solution is “ 2 “)

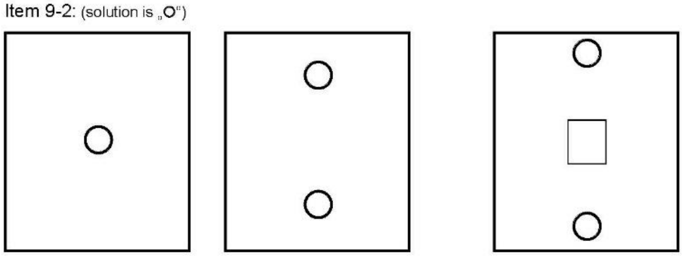


Parallel form B

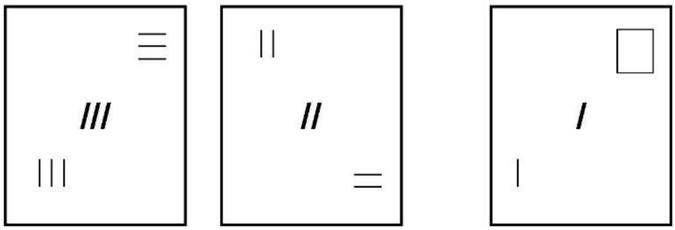
Item 9-1: (solution is „X“)



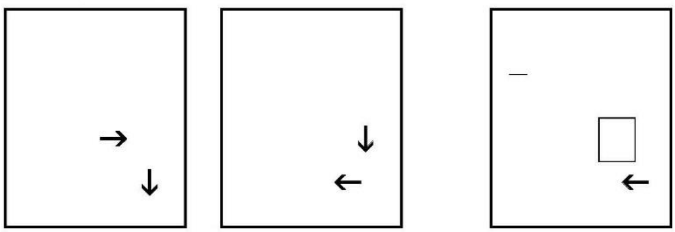
Item 9-2: (solution is „O“)



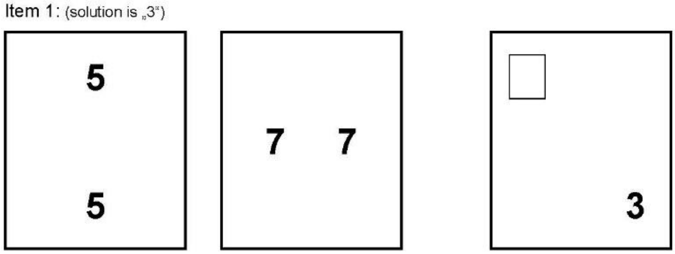
Item 9-3: (solution is „–")



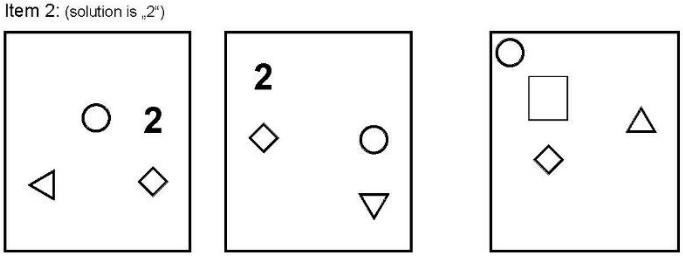
Item 9-4: (solution is „↑")



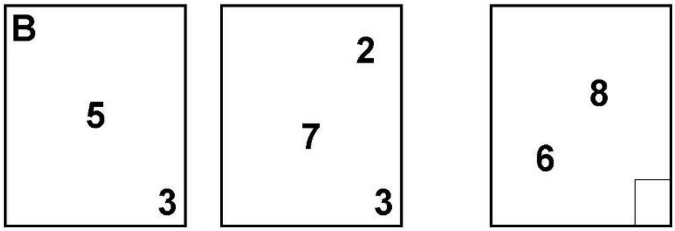
Item 1: (solution is „3")



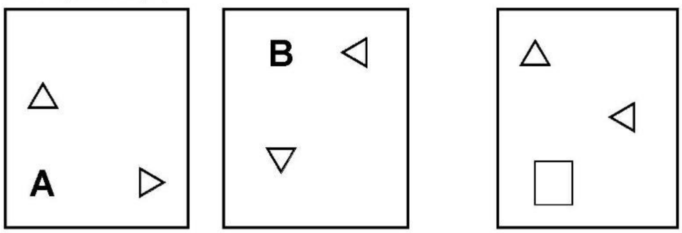
Item 2: (solution is „2")



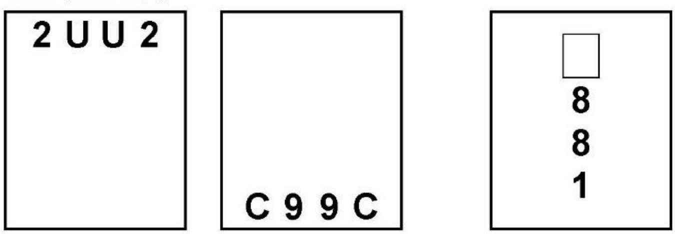
Item 3: (solution is „3“)



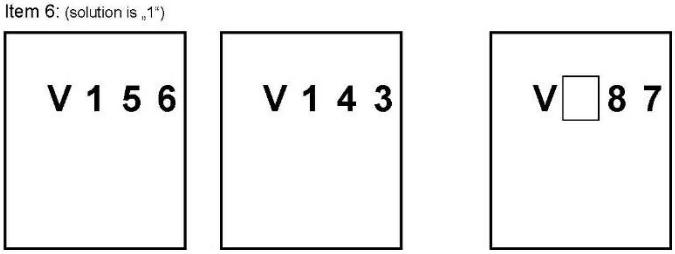
Item 4: (solution is „C“)



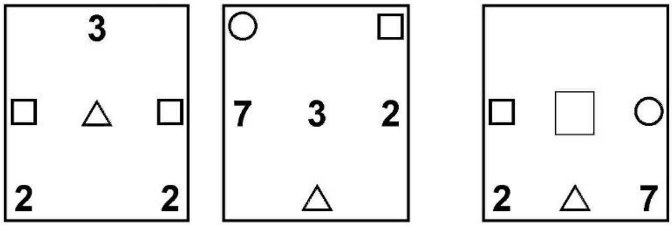
Item 5: (solution is „1“)



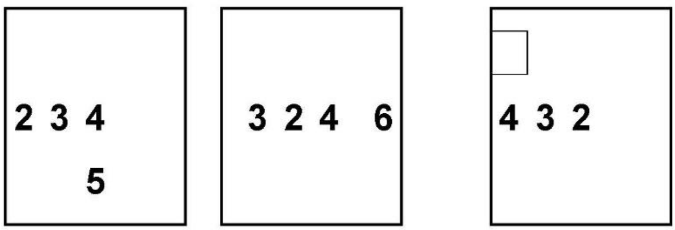
Item 6: (solution is „1“)



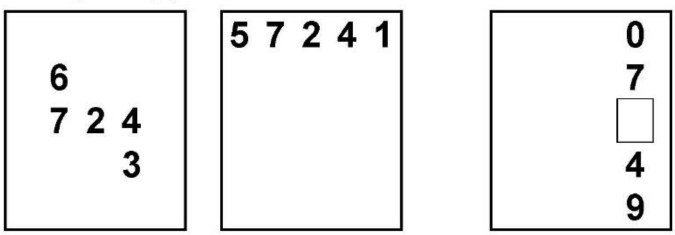
Item 8: (solution is „3“)



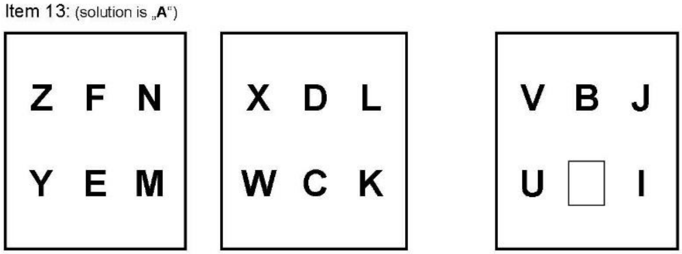
Item 9: (solution is „7“)



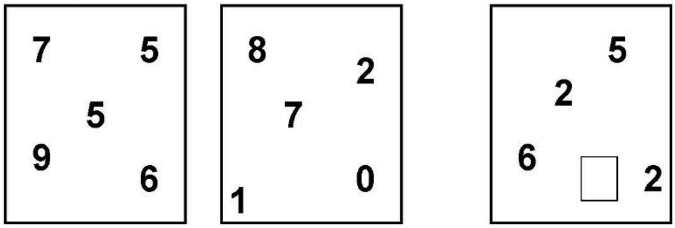
Item 12: (solution is „2“)



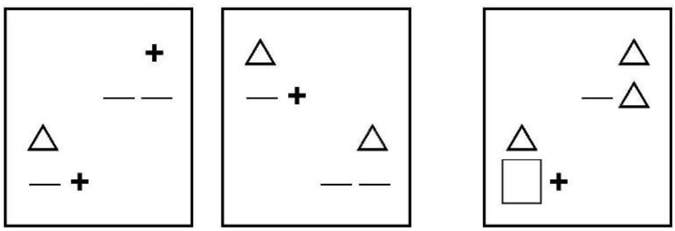
Item 13: (solution is „**A**“)



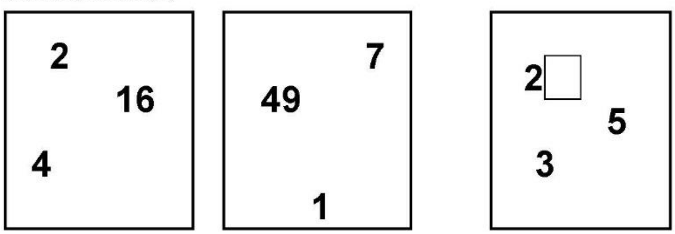
Item 14: (solution is „7")



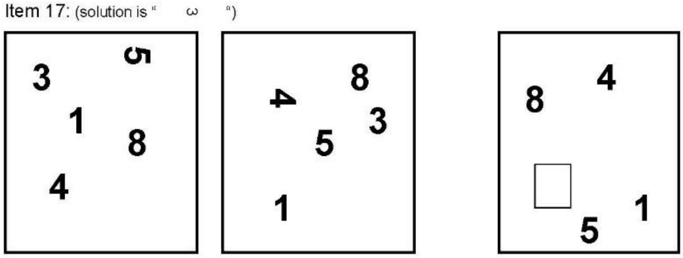
Item 15: (solution is „–")



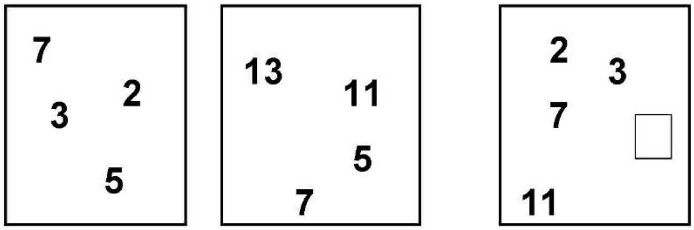
Item 16: (solution is „5")



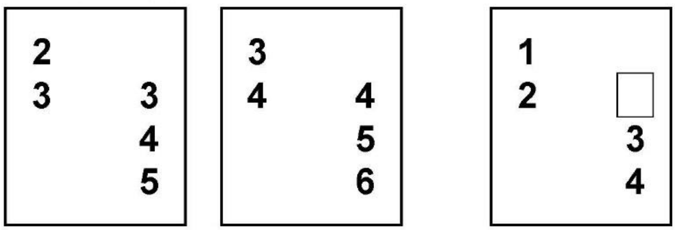
Item 17: (solution is " ω ")



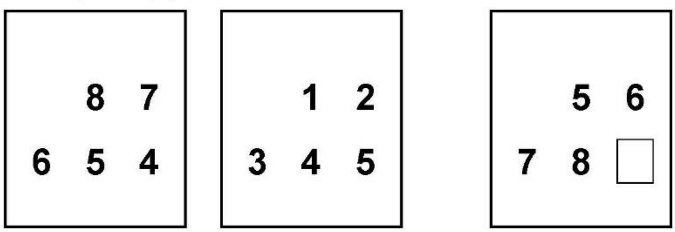
Item 18: (solution is „5“)



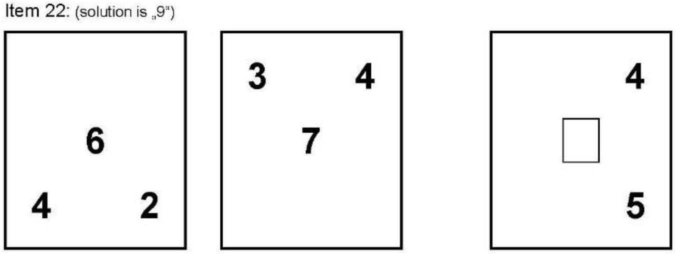
Item 19: (solution is „2“)



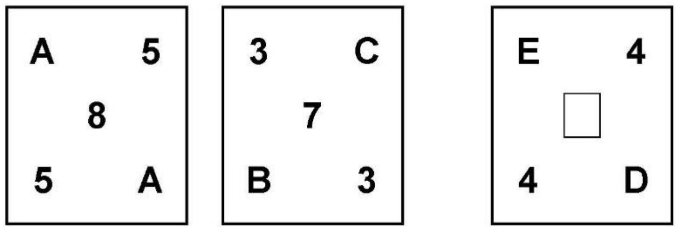
Item 21: (solution is „9“)



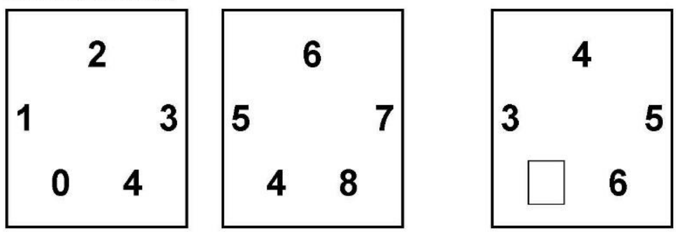
Item 22: (solution is „9“)



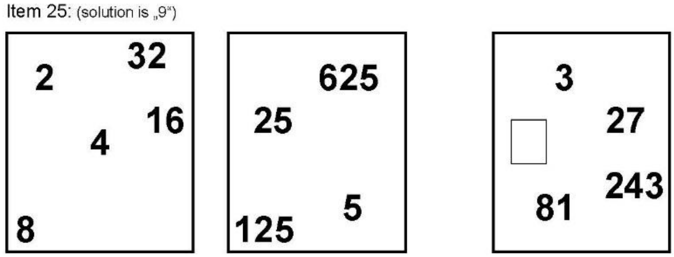
Item 23: (solution is „6")



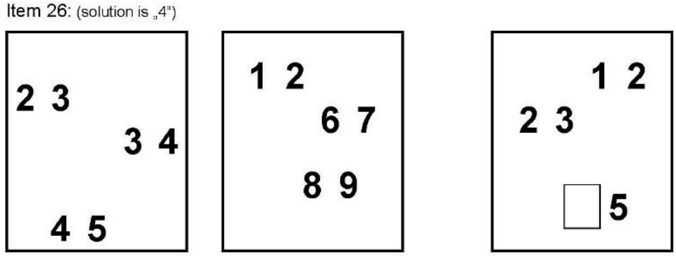
Item 24: (solution is „2")



Item 25: (solution is „9")



Item 26: (solution is „4")



Item 28: (solution is „ 2 “)