Tilburg University

An ordinal scale for transitive reasoning by means of a deductive strategy

Verweij, A.C.; Sijtsma, K.; Koops, W.

Published in:
International Journal of Behavioral Development

Publication date:
1999

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Verweij, A. C., Sijtsma, K., & Koops, W. (1999). An ordinal scale for transitive reasoning by means of a deductive strategy. *International Journal of Behavioral Development*, 23(1), 241-264.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

International Journal of Behavioral Development

<http://jbd.sagepub.com>

An Ordinal Scale for Transitive Reasoning by Means of a Deductive Strategy

Anton C. Verweij, Klaas Sijtsma and Willem Koops
International Journal of Behavioral Development 1999; 23; 241
DOI: 10.1080/016502599384099

The online version of this article can be found at:
<http://jbd.sagepub.com/cgi/content/abstract/23/1/241>

Published by:

<http://www.sagepublications.com>

On behalf of:

[International Society for the Study of Behavioral Development](#)

Additional services and information for *International Journal of Behavioral Development* can be found at:

Email Alerts: <http://jbd.sagepub.com/cgi/alerts>

Subscriptions: <http://jbd.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations (this article cites 31 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
<http://jbd.sagepub.com/cgi/content/refs/23/1/241>

An Ordinal Scale for Transitive Reasoning by Means of a Deductive Strategy

Anton C. Verweij

Vrije Universiteit Amsterdam, The Netherlands

Klaas Sijtsma

Tilburg University, The Netherlands

Willem Koops

Vrije Universiteit Amsterdam, The Netherlands

Transitivity tasks with either equality or inequality relations between three, four, or five objects were presented by means of a successive and a simultaneous presentation procedure to 124 third-grade primary schoolchildren. Equality tasks were easier than inequality tasks. Inequality tasks were more difficult when presented successively than simultaneously. Difficulty of equality tasks was not affected by presentation procedure. These findings were interpreted in terms of the Fuzzy Trace Theory. Successively presented inequality tasks induced more deductive reasoning strategies, and simultaneously presented tasks induced both positional and reductional strategies. As perceptual differences between objects increased, more visual strategies were used.

These conclusions were used to construct a series of nine tasks geared at inducing transitive reasoning by means of a deductive strategy. Mokken scale analysis was used to construct an ordinal scale for transitive reasoning. Data were collected from 417 second- through fourth-grade primary school children. The scale score correlated highly with arithmetic skills and insight into the quantity concept. Possibilities for future transitivity research, supported by advanced psychometric methods, are discussed.

Transitivity tasks typically consist of a series of three, four, or five objects which may differ in either length, weight, size, or area. Pairs of adjacent objects may be presented successively or simultaneously, that is, a pair is

Requests for reprints should be sent to K. Sijtsma, Department of Research Methodology, FSW, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; e-mail: k.sijtsma@kub.nl.

presented without or with the presence of the other objects, respectively. This way, children are familiarised with the relation (larger than, heavier than, and so forth) between adjacent objects from a series. This is the premise information. In the test phase, children are asked to infer the transitive relation between nonadjacent objects on the basis of the premise information. We will denote objects as A, B, C, ...; and the physical property on which they are compared as X. If object A is shorter than B, this is indicated as $X_A < X_B$, and so on.

Originally, transitive inference was understood from a Piagetian framework, later criticised and replaced by an information-processing approach (notably by Trabasso and others). The more recent Fuzzy Trace Theory (FTT), developed by Brainerd and others, offers a broader view by separating memory and reasoning performance. FTT will be introduced here as a general, theoretical framework to study the preference for reasoning strategies as a function of task format and presentation procedure.

Transitivity tasks vary in format (equality or inequality relations, and mixtures) and in number of objects. Several presentation procedures for the premise information have been used. Different types of tasks and presentation procedures make different demands on memory capacity and/or may induce different reasoning strategies. It can thus be hypothesised that difficulty level of transitive reasoning tasks depends on use of reasoning strategies, which in turn depend on task format and presentation procedure.

The aim of this study is: (1) to introduce FTT and to explain the core concepts of "task format", "presentation procedure", and "reasoning strategy", and their relations (Experiment 1); (2) to study the preference for reasoning strategies as a function of presentation procedure and task format (Experiment 1); and (3) to construct a unidimensional scale for transitive reasoning on the basis of deduction from the premises, and to explore the validity of this scale (Experiment 2).

Approaches to Transitive Reasoning: Fuzzy Trace Theory

Piaget (1942; Piaget & Inhelder, 1941; Piaget & Szeminska, 1941) assumed that premise information had to be encoded literally and had to be integrated to obtain the transitive relationship between nonadjacent objects. However, Piaget did not control experimentally for memory effects (Thayer & Collyer, 1978). To control for such effects, Bryant and Trabasso (1971) included an extensive training procedure in their experiments. The results showed that during training the extreme pairs (e.g. AB and DE from a 5-terms series) were learned first, followed by the middle pairs (BC and CD). Riley and Trabasso (1974) hypothesised that the premises were encoded as a series, following an "end-inward" strategy. Sternberg (1980) developed a

linguistic-spatial model to understand such an internal representation. These studies contradicted the Piagetian deductive interpretation of transitive reasoning.

The information-processing approach was seriously questioned on the basis of characteristics of the training procedure. Among others, DeBoysson-Bardies and O'Regan (1973), Kallio (1982), and Perner and Aebi (1985) found that training induced nondeductive strategies. Trabasso (1975, 1977) argued that in the training phase the internal representation of the series could be constructed only on the basis of transitive inferences (also see Halford & Kelly, 1984). Breslow (1981), however, demonstrated that an ordered series could also be constructed without transitive inferences. Furthermore, Perner and Mansbridge (1983), and Perner, Steiner, and Steahelin (1981) reported that the construction of an internal representation depended on visual feedback provided during training. Bryant and Trabasso's claim that poor reasoning performance was due to memory failure was contradicted by Halford and Galloway (1977) and Russell (1981) who found that one-third of their subjects could not infer a transitive relation between nonadjacent objects despite recall of the premises. In addition, Brainerd and Reyna (1990) found evidence that memory of the premise information and reasoning are independent abilities. Brainerd and Kingma (1985) and Reyna and Brainerd (1995) found this memory independence not only for transitivity, but also for class inclusion and conservation. They therefore proposed FTT to explain their findings.

According to FTT the information from the premises is encoded along a continuum of varying exactness (Brainerd & Reyna, 1990). One end is defined by vivid traces that literally contain recently encoded information, and the other end by fuzzy traces that contain schematic or degraded information. Reyna and Brainerd (1995) supposed that the premise information is encoded at various levels of exactness, to be compared with levels of measurement scales: For a transitivity task of length, at the lowest level (cf. the nominal level) only the presence or absence of length is represented, for example, by the trace "All objects are equal"; at the highest level (cf. the ratio level) the length relation between adjacent objects in a series is represented, for example, "Object B is longer than A, and object C is longer than B"; and at an intermediate level (cf. the ordinal level) global qualitative patterns of information are represented, for example, "Things get smaller to the left". According to FTT, children prefer to process premise information at the least precise level of exactness (Brainerd & Reyna, 1990; Reyna & Brainerd, 1995). A deductive reasoning strategy can be conceived of as processing the premise information at the ratio level of exactness. Reyna and Brainerd (1990) pointed out that encoding of premise information at this higher level simultaneously parallels encoding at less precise levels. For example, if children compare the length of objects A and

B in the presentation phase, at the same time they encode at a nominal level: "Objects are different". Contrary to the Piagetian view that a transitive judgement is based on a deduction at the ratio level, and to the information-processing view that transitive judgements are read out directly from internal representations, Reyna and Brainerd (1995) argued that children will use the lowest level of exactness necessary for a correct answer. FTT thus permits to predict relations between reasoning strategies on the one hand, and task formats and presentation procedures on the other.

Presentation Procedures and Reasoning Strategies

Piaget and Inhelder (1941) presented the premises *successively*. For example, in a 3-term inequality task of length ($X_A < X_B < X_C$) object A was first compared with B, and after removal of these objects, object B was compared with C. To test the nonadjacent AC pair, the three objects were arranged in a random order with 30 centimetres between adjacent objects. The relationship between the objects A and C could not be visually perceived because the differences in length were small relative to the distance. Assuming that for a correct solution a deduction from the premise information is needed, we refer to this type of reasoning as *deductive* reasoning.

Brainerd (1974; Brainerd & Kingma, 1984, 1985) presented the objects *simultaneously* in an ordered ascending ($X_A < X_B < X_C$) or descending ($X_A > X_B > X_C$) series. Children were familiarised with the premises by first placing two adjacent objects (e.g. objects A and B) close together so that the length difference was visible. The child was asked to indicate the longest object. Next, the objects were placed in their original position. This procedure was repeated for the objects B and C. When the adjacent pairs (AB and BC) and the nonadjacent pair (AC) were tested, the order of the series was not changed and the complete series was visible to the children. Chapman and Lindenberger (1992a,b) argued that a simultaneous presentation of the premises allows children to rely on the correlation between the objects and their position in the ordered series. They did not consider reasoning based on a serial ordering of the objects (functional reasoning) as transitive reasoning in the usual sense (operational reasoning), which involves the integration of premise information. Brainerd and Reyna (1992a,b; Reyna & Brainerd, 1990) replied that there is no evidence for a distinction between functional and operational reasoning because reasoning is independent of memory for premise information.

The simultaneous presentation procedure thus seems to facilitate encoding of premise information at an ordinal level of exactness. As the length of the objects and their position in the series are perfectly related, we

assume that for processing the ordinally encoded information a *positional* reasoning strategy is needed.

Task Format and Reasoning Strategies

In transitivity research, inequality tasks, equality tasks, and mixed tasks are distinguished. *Inequality* tasks used by Piaget (e.g. Piaget & Inhelder, 1941) consisted of three objects with different weights but equal volumes ($X_A < X_B < X_C$). Bryant and Trabasso (1971; also see Brainerd & Kingma, 1984, 1985; Chapman & Lindemberger, 1988; DeBoysson-Bardies & O'Regan, 1973; Trabasso, 1977) used inequality tasks consisting of four, five, or six objects. Piaget and Inhelder (1941, Piaget & Szeminska, 1941) also used *equality* tasks with the format $X_A = X_B = X_C = X_D$. Murray and Youniss (1968) and Youniss and Murray (1970) used *mixed* tasks of the format $X_A = X_B < X_C$ and $X_A < X_B = X_C$. Harris and Bassett (1975) used a mixed task with the format $X_A = X_B < X_C = X_D$.

Thayer and Collyer (1978) pointed out that the use of different task formats led to contradictory findings concerning the age at which transitive reasoning emerged. Different task formats thus corresponded to different difficulty levels. Because 3-term inequality tasks and 3-term mixed tasks appeared to require the same reasoning strategy (Sijtsma & Verweij, 1992; Verweij, Sijtsma, & Koops, 1996) the present study only used inequality and equality tasks.

The presentation of a 3-term equality task ($X_A = X_B = X_C$) of length may result in encoding the premise information at a nominal level of exactness (e.g. "All objects are equal"). Obviously, such a *reductional* reasoning strategy is only applicable for *equality* tasks (Verweij et al., 1996). The presentation procedure (successive or simultaneous) is irrelevant here because reduced premise information can be encoded in both cases. Because children prefer to process the least precise level of exactness, it can be expected that they will use this nominal information for the solution of the AC item.

If the length differences between the objects are large enough to be visually perceptible, *inequality* tasks also can be solved using a *visual* strategy. Piaget and Inhelder (1941) did not consider this strategy to reflect transitive reasoning because it does not integrate the premise information. To prevent children from using a visual strategy, the length difference of adjacent objects in several experiments was 0.5cm, whereas their distance was 30cm (e.g. Brainerd, 1974; Brainerd & Kingma, 1984; Kingma, 1983; Youniss & Dennison, 1971). It can be argued, however, that visual cues only have an impact on processing the relevant information at the ratio level of exactness. For example, if premise information is encoded at a nominal level

(e.g. “All things are the same”) or at an ordinal level (e.g. “Things get longer to the right”), a correct answer can be found without the use of visual cues (also see Reyna & Brainerd, 1990). It thus seems that a visual strategy will be used if for a correct solution on a transitivity item children are forced to rely on verbatim premise information, for example, if an explanation is required. Because processing information at the ratio level was included in this study and an explanation was required for responses, the visual strategy is relevant here.

The Present Study

To summarise, a deductive reasoning strategy may be used to solve equality tasks and inequality tasks, both under the successive and the simultaneous presentation mode. A positional strategy may be used for inequality tasks under the simultaneous presentation mode. A reductional strategy may be used to solve equality tasks under both presentation modes. A visual strategy may be used to solve equality and inequality tasks under both presentation modes. Children may prefer the strategy that uses the least precise level of exactness of encoded premise information. They may use the highest level of precision only if they are forced to do this. Thus, in practice a visual strategy may frequently be used for the solution of successively presented inequality tasks.

The first study analysed the effects of task format (inequality and equality) and presentation procedure (successive and simultaneous) on reasoning performance. The identification of the four strategies was done using verbal explanations of the judgements. Although explanations may reflect a confounding of reasoning ability and verbal ability (Brainerd, 1973, 1977), in many situations useful information about strategies may be obtained (Chapman & Lindenberger, 1988). To assess the possible switching between strategies, and the possible preference for the use of the visual strategy, 3-, 4-, and 5-term transitivity tasks were used. Dominance of strategies under the successive and simultaneous presentation mode, and for inequality and equality tasks, was also studied.

The second study used results from the first study to construct by means of Mokken scale analysis (Meijer, Sijtsma, & Smid, 1990; Mokken, 1971, 1997; Mokken & Lewis, 1982) an ordinal scale for transitive reasoning based on a deductive strategy. Also, the relation was investigated between the scale score and several measures of arithmetic skill. Mokken scale analysis has been used for the construction of ordinal scales for the Piagetian concepts of seriation (Kingma & Reuvekamp, 1984) and conservation (Kingma & TenVergert, 1985). Verweij et al. (1996) used Mokken scale analysis to construct a preliminary scale for transitive reasoning.

EXPERIMENT 1

Method

Participants

To avoid a confounding of effects of age or education with task performance only children ($n = 124$) from the third grade (Dutch primary school) participated in this study. The mean age expressed in months was 104.6 with a standard deviation of 4.5. Both sexes were approximately equally represented.

Tasks

Tasks of length were used because length has most frequently been manipulated in transitivity research (e.g. Braine, 1959; Brainerd & Kingma, 1984; Trabasso, 1975, 1977), and because length was expected to induce a visual strategy sooner than, for example, weight and area. Two identical series of six tasks were constructed. The premises of Tasks 1, 3, 5, 7, 9, and 11 were presented successively, and the premises of Tasks 2, 4, 6, 8, 10, and 12 simultaneously. Within each series task format (inequality: Tasks 1 through 6; and equality: Tasks 7 through 12) and number of objects (three objects: Tasks 1, 2, 7, and 8; four objects: Tasks 3, 4, 9, and 10; and five objects: Tasks 5, 6, 11, and 12) were varied systematically. Thus, a task was constructed for each combination of presentation mode, task format, and number of objects.

The objects of the equality tasks had a length of 10cm. The mean length within the inequality tasks was 10cm, with a 0.5cm difference between adjacent objects. The distance between adjacent objects in a series was 30cm. Objects of the same task had different colours to identify them in conversation. Colours were randomly assigned to objects such that the same objects of corresponding tasks under the two presentation modes had different colours, thus excluding order effects due to colour. Different materials were used to increase the attractiveness of the tasks and to keep children motivated. Materials were: plastic tubes (diameter 19mm), copper tubes (diameter 12mm), and round wooden sticks (diameter 8mm). Corresponding tasks under different presentation modes consisted of different materials.

Procedure

Administration was individual. For each child, the presentation orders of the tasks and, within tasks, of the premises were randomly determined.

Successive presentation went as follows. Suppose that for a child the premises of Task 5 were presented in the order BC, AB, DE, and CD. The experimenter handed over the objects B and C and asked: "Which object is

longer, B or C, or are they equally long?" (in fact, the objects were identified by colour). After an answer was given, B and C were removed. This procedure was successively repeated for the other three premises. In the test phase, the child was asked to judge and explain the relation between the six pairs of objects that were nonadjacent in the ordered series. The objects of the task were arranged in a random order. The order in which the pairs of nonadjacent objects had to be assessed also was randomly determined. Children were not permitted to touch the objects. For each pair, the relation between the objects had to be judged and, next, a verbal explanation was requested. Judgement and explanation were recorded in writing.

With simultaneous presentation, the objects of a series were laid down in front of the child. For inequality tasks, length increased from left to right. The experimenter placed the objects from a premise close together so that the length relation was visible and could be memorised. Children were allowed to pick up both objects to verify their relation. For each premise, the experimenter asked: "Which object is longer, or are they equally long?" After an answer was given, the experimenter placed the objects in their original position and repeated this procedure for the other premises. Next, for each task the pairs of nonadjacent objects were tested in a randomly determined order. The experimenter first asked for a judgement, and then for an explanation.

Scoring

Tasks and Items. The pairs of nonadjacent objects of a task were called transitivity items or T-items, for short. A T-item that required the knowledge of m premises for a deductive solution was a T_m -item. For example, an AD item is a T_3 -item. Children were required to provide a judgement and an explanation for 40 T-items in total (one, three, and six items for each of the 3-, 4-, and 5-term tasks, respectively). A score of 1 was obtained if the judgement of a T-item was correct, and a score of 0 if the judgement was incorrect. For the 3-term tasks the dependent judgement variable had 0–1 scores (one T-item); for the 4-term tasks this variable had 0–3 scores (sum score on three T-items); and for 5-term tasks this variable had 0–6 scores (sum score on six T-items).

Reasoning Strategies. A deductive strategy was inferred if the explanation suggested retention of all premises which were necessary for a solution by deduction. For example, the explanation for a T_3 -item (e.g. AD) had to be: "B is the same as A, C is the same as B, and D is the same as C; therefore, A and D are equally long". Deductive solutions were possible for each item type under each presentation mode.

A positional strategy was inferred if the explanation suggested use of the spatial position of the objects. For example, for a T_2 -item (e.g. AC) the

explanation could be: "C is longer than A, because C is to the right of A". Because for an equality task the length of the objects is irrelevant for their position in a series, a solution based on position was not plausible under simultaneous presentation.

A reductional strategy was inferred if the common property of the objects (i.e. length) was used to solve the item. An example of an explanation is: "Everything is the same" or "Everything is equal". Inequality tasks also can be solved using a reductional explanation (e.g. "Everything is different" or "Nothing is equal").

A visual strategy was inferred if the judgement was based on a visual comparison of the objects. For example, for a T_3 -item (e.g. AD) the explanation could be: "D is longer than A, because I can see that" or "because it looks that way". Equality tasks also can be solved using a visual explanation (e.g. "Everything is equal because I can see that").

Table 1 summarises the types of explanations per type of task. Explanations which could not be classified as deductive, positional, reductional, or visual were assigned to a miscellaneous category which contained explanations such as: "I just guessed"; "I'll take that one"; and "I don't know".

Results

Influence of Presentation and Task Format on Judgement

Because for each subset of four tasks with the same number of objects the judgement variable had different numbers of score categories (see Scoring), separate analyses were done for subsets to determine the influence of presentation procedure (P; two levels) and task format (F; two levels) on judgement. Table 2 displays the task means of the judgement variable. The pattern of means per subset of four tasks with the same number of objects suggests that there may be a weak interaction between factors P and F.

TABLE 1
Explanations Expected for T-items as a Function of Presentation Mode and Task Format

<i>Presentation:</i>	<i>Successive</i>		<i>Simultaneous</i>	
	<i>Inequality</i>	<i>Equality</i>	<i>Inequality</i>	<i>Equality</i>
<i>Explanation</i>				
Deductive	+	*	*	*
Positional	—	—	+	—
Reductional	*	+	*	+
Visual	*	*	*	*

+, theoretically most likely explanation; —, explanation seems impossible; *, explanation is possible as an alternative.

TABLE 2
Mean Judgement Score for Each Combination (Task) of Number of Objects, Task Format, and Presentation Mode

	<i>Number of Objects</i>					
	<i>3</i>		<i>4</i>		<i>5</i>	
	<i>Task</i>	<i>Mean</i>	<i>Task</i>	<i>Mean</i>	<i>Task</i>	<i>Mean</i>
<i>Inequality Tasks</i>						
Successive	1	0.88	3	2.61	5	4.94
Simultaneous	2	0.95	4	2.78	6	5.37
<i>Equality Tasks</i>						
Successive	7	0.97	9	2.90	11	5.62
Simultaneous	8	0.98	10	2.87	12	5.61

For the 3-term inequality Tasks 1 (successive presentation) and 2 (simultaneous presentation), the null-hypothesis of no relation between presentation modes of inequality tasks could not be rejected (Fisher's Exact Test, two-tailed $P = .25$). The performance on inequality tasks thus is different under different presentation modes. For the 3-term equality Tasks 7 (successive presentation) and 8 (simultaneous presentation), however, the null-hypothesis of independence was rejected (Fisher's Exact Test, two-tailed $P = .00001$). For equality tasks, different presentation modes thus do not lead to differences in performance. Combination of both results led to the conclusion that there is an interaction between factors P and F.

For the 4-term tasks a significant but weak interaction [$F(1,123) = 4.63$, $P = .03$] was found. The effect of factor P was thus different for inequality tasks and equality tasks (Table 2). The mean score of the simultaneously presented inequality Task 4 was higher than the mean score of the successively presented inequality Task 3. The difference between the means of the simultaneously presented equality Task 10 and the successively presented equality Task 9 had the opposite sign but was only 0.03.

For the 5-term tasks a significant interaction [$F(1,123) = 6.84$, $P = .01$] was found. Performance on the simultaneously presented inequality Task 6 was superior to that on the successively presented inequality Task 5, whereas performance on the equality Tasks 11 and 12 had the opposite sign but differed by only 0.01 (Table 2).

Influence of Length Difference on Type of Reasoning Strategy

Four subsets of three tasks, each subset corresponding to a particular combination of factors P and F, were analysed separately to study the influence on the explanation of variation in length difference (inequality tasks) and distance (equality tasks; 0 length difference) between the two

extreme objects of a task. Per subset of inequality tasks, length differences were 1cm for the AC-item of a 3-term task; 1.5cm for the AD-item of a 4-term task; and 2cm for the AE-item of a 5-term task. For equality tasks, distances were 60cm, 90cm, and 120cm, respectively.

For each subset of tasks, children used three strategy categories, including the miscellaneous category (Table 3). For each task, the percentages in the three categories were based on 124 observations: Thus, per task the distribution across the three categories is multinomial. Because statistical testing to compare such dependent distributions is awkward and, moreover, an alternative comparison per subset of tasks of the percentages in a column is complicated by the same dependence between these percentages, the results in Table 3 are discussed without formal statistical testing.

The deductive strategy was used to solve successively presented inequality tasks, and the positional strategy was used to solve simultaneously presented inequality tasks. Use of the deductive and the positional strategies decreased monotonically with increasing length difference. Use of the visual strategy increased monotonically with length difference. For equality tasks, irrespective of presentation mode the reductional strategy was dominant. Use of the reductional strategy decreased modestly with distance between objects. The visual strategy was rarely used.

TABLE 3
Percentages of Explanations for the T-items with the Largest Difference/
Distance between Objects, for Each Combination of Presentation Mode
and Task Format

<i>Task No.</i>	<i>Diff. cm/T-item</i>		<i>Strategy^a</i>	<i>Visual</i>	<i>Other</i>
<i>Successive, Inequality</i>					
1	1	AC	50.8	41.1	8.1
3	1.5	AD	37.9	47.6	14.5
5	2	AE	6.4	82.3	11.3
<i>Simultaneous, Inequality</i>					
2	1	AC	81.5	12.9	5.6
4	1.5	AD	67.7	25.8	6.5
6	2	AE	48.4	46.0	5.6
<i>Successive, Equality</i>					
7	0	AC	93.5	2.4	4.1
9	0	AD	91.1	4.0	4.9
11	0	AE	83.9	6.5	9.6
<i>Simultaneous, Equality</i>					
8	0	AC	95.2	1.0	3.8
10	0	AD	89.5	2.4	8.1
12	0	AE	84.7	4.0	11.3

^aFor "Successive, Inequality" this is "Deductive"; for "Simultaneous, Inequality" this is "Positional"; and for "Successive, Equality" and "Simultaneous, Equality" this is "Reductional".

Influence of Strategy Use on Judgement

For each subset of three tasks defined by a combination of factors P and F, the null-hypothesis that different reasoning strategies (including the miscellaneous category) had no effect on the judgement score (0, 1) distribution of a T-item was tested by means of a chi-square statistic in a 3×2 cross-table. A summary of the detailed results is given.

The null-hypothesis was rejected ($P \leq .001$) for all 10 T-items of the successively presented inequality Tasks 1, 3, and 5. For the T_2 -items, the deductive strategy was dominant. The π -value (proportion of correct judgements) of these items ranged from 0.79 to 0.90. For the T_3 -items and the T_4 -item the visual strategy was dominant. For these items, the π -values ranged from 0.82 to 0.92. Children thus used strategies that led to high probabilities (π -values) of solving the items. For all 10 T-items of the simultaneously presented inequality Tasks 2, 4, and 6 the null-hypothesis was rejected ($P \leq .001$). The positional strategy was dominant for all T-items. This strategy led to π -values varying from 0.87 to 0.95.

The null-hypothesis was rejected ($P \leq .001$) for all 10 T-items of the successively presented equality Tasks 7, 9, and 11, and for all T-items of the simultaneously presented equality Tasks 8, 10, and 12. Here the reductional strategy was dominant, leading to π -values ranging from 0.90 to 0.97, and from 0.90 to 0.98, respectively.

Discussion

Presentation affected judgement on the 3-, 4-, and 5-term inequality tasks (see Table 2). Successive presentation requires the retention of the literal premise information, whereas simultaneous presentation allows reliance on the relation between the length of the objects and their position in the ordered series. Successive presentation thus renders inequality tasks more difficult than simultaneous presentation. Presentation did not affect performance on the equality tasks. These tasks could be solved under both presentation modes without encoding and retrieving the literal information from the premises.

Strategy use depended on presentation procedure, task format, and distance between objects (equalities) or length difference between objects (inequalities) (see Table 3). For inequality tasks, successive presentation induced deductive reasoning for small length differences; otherwise, children tended to use a visual strategy. Simultaneous presentation was strongly related to a positional strategy for all length differences considered. For larger length difference, however, simultaneous presentation also induced substantial use of a visual strategy. For equality tasks, presentation procedure and distance did not affect use of a reductional strategy. Deductive strategies were not used, and visual strategies were rarely used.

Children appeared to use the simplest strategy for solving a task. Choice of strategy depended on presentation mode, task format, and length difference or distance between objects. The theoretical expectations (Table 1) were confirmed by the empirical results (Table 3). Deductive reasoning by approximately half of the children was induced only by successively presented inequality tasks with a small (1cm) length difference between the objects to be compared. For larger differences this strategy was less frequently used.

If transitive reasoning is considered to be an aspect of logical reasoning, then it should be based on deductive reasoning. A test was constructed for the measurement of transitive reasoning, conceived of as a logical reasoning ability. From the results of Experiment 1 we concluded that: (1) only inequality tasks should be used; (2) the premises should be presented successively; (3) the length difference between the objects of the T-items should be small; and (4) an explanation should be required to check whether deductive reasoning was used.

EXPERIMENT 2

Method

Participants

The participants were 417 second- through fourth-grade primary schoolchildren. The mean ages expressed in months were 100.6 ($n = 139$), 114.4 ($n = 140$), and 122.2 ($n = 138$), respectively. The standard deviations were 5.8, 5.4, and 5.2, respectively. In each grade, both sexes were approximately equally represented.

Tasks

Length, weight, and size were used in this study, because each of these properties was useful and representative for the measurement of transitive reasoning (Verweij et al., 1996). The nine inequality tasks of length, size, and weight consisted of three, four, and five objects, thus representing all nine combinations of these characteristics. The objects of the tasks had different colours (white, red, yellow, green, and blue), which were randomly assigned.

For the tasks of length (Tasks 1, 2, and 3) round wooden sticks with a diameter of 0.6cm were used. The difference between adjacent sticks was 0.2cm. The difference in a T_2 -item thus was 0.4cm. The mean length of the objects per task was 9.5cm.

The tasks of size (Tasks 4, 5, and 6) consisted of round wooden discs with a thickness of 0.4cm. The diameters of adjacent objects differed by 0.2cm. The mean diameter per task was 5.1cm. Compared with diameter area is a squared measure. As it was not known whether children would respond to

differences in diameter or area, use of a visual strategy could not be excluded.

For the tasks of weight (Tasks 7, 8, and 9) clay balls with a diameter of 5cm were used. Adjacent objects differed by 30 grams. The mean weight in a task was 105g. As all the objects in a task had the same size, use of visual strategies was not plausible.

Procedure

All tasks were successively presented, and administration was individual. For each child the presentation orders of the tasks, the premises, and the T-items were randomly determined. To control for use of visual strategies, judgements and explanations were required only for T₂-items. T₃- and T₄-items were not tested.

Scoring

Tasks and Items. A score of 1 was obtained if a correct judgement was followed by a deductive explanation, or if an initially incorrect judgement was revoked on the basis of a deductive explanation. The maximum sum score based on the T₂-items per task was 1 (three objects), 2 (four objects), and 3 (five objects).

Scaling Method. Mokken scale analysis (Mokken & Lewis, 1982) for polytomous items (Hemker, Sijtsma, & Molenaar, 1995; Molenaar, 1997; Verweij et al., 1996) assumes that a higher task score indicates a higher ability level (e.g. transitive reasoning). For example, for a 5-term task the failure to solve any of the three T₂-items by means of a deductive strategy represents the lowest ability level for this task (score 0). In this case, the first *item step* was failed. The correct solution of at least one T₂-item, two T₂-items, or three T₂-items by means of a deductive strategy represent increasingly higher ability levels, scored 1, 2, and 3, respectively. Correspondingly, only the first, only the first two, and all three item steps have been taken.

Important questions in Mokken scale analysis are whether this scoring procedure is justified, and whether the sum score on all nine tasks can be used to order persons on a scale for transitive reasoning (Hemker, Sijtsma, Molenaar, & Junker, 1997). We used the *Mokken model of double monotonicity*(DMM; Molenaar, 1997; Sijtsma & Junker, 1996) to answer these and other questions. If the DMM fits the data it allows an ordering of persons on the ability scale by means of their total score on the nine tasks, and an ordering of the item steps by difficulty level.

The DMM assumes that the set of items measures one ability (unidimensionality). It also is assumed that the probability of taking an item

step is nondecreasing in the ability (monotonicity). For example, the higher the transitive reasoning ability, the more likely the child is to solve at least, say, two T_2 -items of the same task by means of a deductive strategy. Finally, it is assumed that the ordering of the *item steps* by difficulty level is the same in each subgroup of the population of interest (invariant item step ordering), for example, boys and girls, different grades, and subgroups representing different ability levels. The ordering of the *tasks* in different subgroups may be more important than the item step ordering. The task ordering has to be investigated separately. A task ordering which is the same in all possible subgroups is an *invariant task ordering*.

Unidimensionality implies that all covariances between the tasks must be non-negative (Mokken & Lewis, 1982). The investigation of monotonicity amounts to investigating whether a *person ordering* is justified. The scalability coefficient H is used for this purpose (Molenaar, 1997). Hemker et al. (1995) consider a set of items to be unscalable for practical purposes if $H < 0.30$; weak scalability is obtained if $0.30 \leq H < 0.40$, medium scalability if $0.40 \leq H < 0.50$, and strong scalability if $0.50 \leq H \leq 1.00$ (maximum H). To support the H analysis, monotonicity was also investigated using the item step-restscore regression: This is the regression of the item step score on the total score on the other eight tasks. Invariant item step ordering and invariant task ordering were investigated across relevant subdivisions of the sample. The program MSP 3.0 (Molenaar, Debets, Sijtsma, & Hemker, 1994) was used for data analysis.

Results

Analysis of the Verbal Explanations

The mean proportion of children who gave correct judgements on the six T_2 -items of the tasks of length, size, and weight equalled 0.70, 0.92, and 0.62, respectively. Length and weight tasks were predominantly correctly solved using deductive reasoning (Table 4). Size tasks were most often correctly solved using visual reasoning. The incorrect visual explanations for length tasks appeared to be due to guessing caused by the large distance between the objects of some of the items (for a 5-term task the maximum distance was 120cm). In 44% of the cases explanations were lacking for the T_2 -items of weight, which may be due to the absence of visual cues.

Mokken Scale Analysis

Person Ordering. Unidimensionality was supported by positive covariances between all nine tasks. The MSP program selected the nine tasks into the same scale with a common H -coefficient of 0.75. The task H_i -coefficients varied from 0.62 to 0.84 (Table 5). Monotonicity was

TABLE 4
Total Proportion of Explanations for Three Task Types of Length, Size, and Weight

	<i>Correct Explanation</i>		<i>Incorrect Explanation</i>		<i>No Explanation</i>
	<i>Deductive</i>	<i>Visual</i>	<i>Deductive</i>	<i>Visual</i>	
<i>Type of Task</i>					
Length	0.51	0.12	0.18	0.09	0.10
Size	0.21	0.68	0.04	0.01	0.05
Weight	0.40	0.01	0.13	0.01	0.44

investigated by inspection of the regression of the observed probability of taking a particular item step on the total score on the other eight tasks (the restscore). For each of the 18 item steps, the regression was inspected. An observed probability followed by a *smaller* probability in a *higher* restscore group contradicts monotonicity. MSP detected only one significant result [test proposed by Molenaar (1970, Ch. 4); second item step of Task 2, length]. With 18 regressions in total, this single result can be ignored. The results from the *H*-analysis and the regression analysis both support the usefulness of the total score for ordering persons on the latent ability scale.

Item Step Ordering and Task Ordering. The invariant item step ordering was investigated: (1) for boys and girls; (2) in Grades 4, 5, and 6; and (3) in the groups with relatively low total score (total score ≤ 6) and relatively high total score (total score ≥ 7). Table 6 shows the item steps ordered according to increasing easiness (higher proportion taking the item step) in the total group. This is the *expected item step ordering*. The item step orderings in subgroups were compared with the expected ordering. No reversals in the expected orderings were found for boys and girls. Grade 4 showed two violations of the expected ordering, which were not significant [test proposed by Molenaar (1970, Ch. 3)]. In the group with the lowest total scores two *nonsignificant* deviations were found. The item steps thus have an invariant ordering.

Table 7 shows the increasing task easiness for the nine tasks in the total group. This is the *expected task ordering*. The task orderings in the subgroups

TABLE 5
H-values and Mean Scores of Nine Tasks Selected in One Scale

	<i>Task Number</i>								
	6	5	1	9	4	8	7	2	3
H_i	0.63	0.78	0.82	0.78	0.81	0.78	0.84	0.67	0.62
Task Mean	0.15	0.36	0.56	0.69	0.76	0.83	0.91	1.20	1.29

TABLE 6

Item Steps Ordered According to Increasing Easiness (Proportion Taking Item Step g) in the Total Group, and Several Subgroups

<i>Item Step</i> $P[X_i \geq g]^a$	<i>Item</i> ^b <i>N:</i>	<i>Total</i> (417)	<i>Gender</i>		<i>Grade</i>			<i>Total Score</i>	
			<i>Boys</i> (211)	<i>Girls</i> (206)	<i>4</i> (139)	<i>5</i> (140)	<i>6</i> (138)	<i>Low</i> (186)	<i>High</i> (231)
$X_5 \geq 2$	Si-2T	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$X_6 \geq 3$	Si-3T	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$X_6 \geq 2$	Si-3T	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$X_9 \geq 3$	We-3T	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
$X_3 \geq 3$	Le-3T	0.01	0.01	0.01	0.00	0.01	0.02	0.00	0.02
$X_9 \geq 2$	We-3T	0.06	0.06	0.05	0.05	0.06	0.06	0.00	0.10
$X_8 \geq 2$	We-2T	0.06	0.06	0.07	0.06	0.06	0.07	0.00	0.11
$X_6 \geq 1$	Si-3T	0.15	0.17	0.14	0.14	0.16	0.17	0.01	0.27
$X_2 \geq 2$	Le-2T	0.23	0.21	0.25	0.22	0.21	0.26	0.05	0.38
$X_3 \geq 2$	Le-3T	0.28	0.28	0.28	0.19	0.31	0.33	0.05	0.47
$X_5 \geq 1$	Si-2T	0.36	0.38	0.35	0.33	0.34	0.42	0.02	0.64
$X_1 \geq 1$	Le-1T	0.56	0.56	0.56	0.45	0.58	0.65	0.08	0.95
$X_9 \geq 1$	We-3T	0.63	0.64	0.62	0.52	0.66	0.72	0.20	0.98
$X_4 \geq 1$	Si-1T	0.76	0.74	0.78	0.73	0.75	0.79	0.48	0.98
$X_8 \geq 1$	We-2T	0.77	0.74	0.80	0.68	0.79	0.83	0.48	1.00
$X_7 \geq 1$	We-1T	0.91	0.91	0.91	0.88	0.90	0.94	0.80	1.00
$X_2 \geq 1$	Le-2T	0.97	0.97	0.97	0.95	0.96	0.99	0.93	1.00
$X_3 \geq 1$	Le-3T	0.99	1.00	0.99	0.99	0.99	1.00	0.98	1.00

^a $P[X_i \geq g]$ is the probability of having at least a score g on task i .

^b Si, size; We, weight; Le, length.

were compared with the expected task ordering. The ordering was identical for boys and girls. Only Grade 6 showed one reversal of adjacent tasks of 0.03 (Tasks 1 and 2, length) and one tie (Tasks 3, length; and 8, weight) which, however, did not contradict the expected ordering. The subdivision on the basis of total score showed one small (0.01) and one large (0.26) reversal of adjacent tasks in the low group, and one small (0.05) and one large (0.26) reversal of adjacent tasks in the high group. The large violations pertained to the length tasks with three and one T_2 -items in the low group, and the length tasks with one and two T_2 -items in the high group. Possible explanations are highly speculative and are, therefore, avoided here.

Reliability and Distribution of the Scale Score. MSP calculated a reliability estimate (Sijtsma & Molenaar, 1987) equal to 0.89. In total, the percentage of 0 and 1 scale scores was 2.4, and the percentage of 16–18 scale scores was 0. Floor and ceiling effects thus were almost absent. The mean scale score varied across age (see bottom of Table 8). A one-way ANOVA revealed a significant difference [$F(4,412) = 3.46$, $P < .01$] between age

TABLE 7

Tasks Ordered According to Increasing Easiness (Mean Score) in the Total Group, and in Several Subgroups

Item No.	Item ^a	Total N: (417)	Gender		Grade			Total Score	
			Boys (211)	Girls (206)	4 (139)	5 (140)	6 (138)	Low (186)	High (231)
6	Si-3T	0.05	0.06	0.05	0.05	0.05	0.05	0.00	0.09
5	Si-2T	0.18	0.19	0.17	0.16	0.17	0.21	0.01	0.32
9	We-3T	0.23	0.24	0.22	0.19	0.24	0.26	0.07	0.36
8	We-2T	0.41	0.40	0.43	0.36	0.43	0.45	0.24	0.55
3	Le-3T	0.43	0.43	0.43	0.39	0.44	0.45	0.34	0.50
1	Le-1T	0.56	0.56	0.56	0.45	0.58	0.65	0.08	0.95
2	Le-2T	0.60	0.59	0.60	0.58	0.59	0.62	0.49	0.69
4	Si-1T	0.76	0.74	0.78	0.73	0.75	0.79	0.48	0.98
7	We-1T	0.91	0.91	0.91	0.88	0.90	0.94	0.80	1.00

^aSi, size; We, weight; Le, length.

Comparisons between adjacent age groups yielded a significant difference (Scheffé test, 5% level) for age group 100–107 and age group 108–115. A significant difference [$F(2,414) = 5.03$, $P < .01$] was also found for grades. Differences between adjacent groups were found for grades 4 and 5, and grades 5 and 6. Because the mean scale score increased across grades but not consistently across age groups this may be an indication that “development” in transitive reasoning is influenced by education.

Correlation of Scale Score with Other Variables. The transitive reasoning scale scores were correlated with: (1) a test score for the quantity concept [measured by a subtest from a Dutch intelligence battery (R-AKIT: Bleichrodt, Drenth, Zaal, & Resing, 1985) which requires comparisons of objects with respect to number, length, volume, weight, and size]; (2) a score on a standardised arithmetic test (requiring operations such as addition, multiplication, and other tasks involving quantities; Cito, 1979); (3) a score representing insight into the number line (Cito, 1979); (4) the teacher’s rating of arithmetic performance; and (5) the teacher’s rating of general performance level. The correlations were: 0.78 (Rakit quantity subtest), 0.70 (Cito arithmetic test), 0.81 (number line comprehension and insight into number structure), 0.53 (rating of arithmetic skills), and 0.43 (rating of general performance level).

Discussion

Approximately 60% of the group used deductive reasoning to solve the T₂-items of length and weight, and 68% used a visual strategy to solve the T₂-items of size. It appears that for solving size items children responded to

TABLE 8
Frequency Distribution, Mean, and Standard Deviation (SD) of Scale Score on Nine Transitivity Tasks, for Age Group, Grade, and Total Group

	<i>Age Group (months)</i>					<i>Grade</i>			<i>Total Group</i>
	<i><100</i>	<i>100-107</i>	<i>108-115</i>	<i>116-123</i>	<i>>124</i>	<i>4</i>	<i>5</i>	<i>6</i>	
<i>Scale</i>									
<i>Score N:</i>	64	70	98	120	65	139	140	138	417
0	1	1	1	0	0	2	1	0	3
1	1	4	0	2	0	5	2	0	7
2	3	5	7	3	5	7	10	6	23
3	3	7	6	10	8	11	10	13	34
4	7	9	6	9	5	18	9	9	36
5	9	10	6	12	2	19	9	11	39
6	9	8	13	9	5	16	20	8	44
7	11	3	15	15	11	15	16	24	55
8	4	9	12	14	14	12	20	21	53
9	7	5	12	14	3	13	17	11	41
10	5	3	14	16	6	9	17	18	44
11	3	3	3	10	5	8	5	11	24
12	1	3	1	5	1	4	3	4	11
13	0	0	0	1	0	0	0	1	1
14	0	0	1	0	0	0	1	0	1
15	0	0	1	0	0	0	0	1	1
16	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0
Mean	6.41	5.80	7.01	7.12	6.72	6.18	6.81	7.26	6.75
SD	2.63	3.02	2.83	2.85	2.75	2.81	2.81	2.80	2.87

area rather than diameter and thus chose to use a simple visual strategy (also see Flavell, 1972, 1977). The scale found measured transitive reasoning and allowed a reliable person ordering based on total score without floor and ceiling effects, and orderings of item steps and tasks that are approximately invariant across gender, grade, and total score groups. Transitive reasoning is closely related to arithmetic skills (Ginsburg, 1977). The high correlations between transitive reasoning and the number line and number structure tasks is an indication that the ordering principle is relevant for both sorts of tasks (also see Kingma & Koops, 1983). Ginsburg (1977) and Levinova (1977) argued that mastering of the ordering principle is a necessary condition to participate successfully in the education of arithmetic. For example, to solve addition or subtraction tasks correctly children must have the insight that numbers are ordered on the number line. The transitive reasoning test developed here might be used to diagnose mastering of the ordering principle.

GENERAL DISCUSSION

The first experiment studied the relations between presentation mode, task format, and reasoning strategy. Inequality tasks were more difficult when presented successively than simultaneously. For equality tasks such a difference was absent. A simple hypothesis for these differences is that a successive presentation of inequality tasks requires the retention of the literal premise information, whereas a simultaneous presentation allows for reliance on the relation between the length of the objects and their position in the ordered series. Therefore, simultaneous presentation of inequality tasks induced positional reasoning.

For successively presented inequality tasks with hardly perceptible differences between the two objects (T_2 -items) deductive reasoning was used more often than visual reasoning. Visual strategies were preferred when the differences between the objects were larger (T_3 - and T_4 -items). For simultaneously presented inequality tasks, however, the children switched to a positional strategy; but proceeding from T_2 - to T_3 - and T_4 -items the greater perceptibility of the differences between the two objects induced visual reasoning. A reductional strategy was consistently used to solve equality tasks.

Deductive reasoning is used only if there is no alternative. The simpler visual strategy is used if the perceptibility of the differences is sufficiently large (T_3 - and T_4 -items). For simultaneously presented tasks the strategy requiring the smallest memory capacity is preferred; for inequality tasks this is the positional strategy and for equality tasks this is the reductional strategy.

These findings support Fuzzy Trace Theory. The bottom line regarding FTT is that it seems to predict closely the results obtained in Experiment 1 (Brainerd & Reyna, 1992a,b; Reyna, 1992; Reyna & Brainerd, 1990, 1995). The changes in strategies across tasks, which in FTT is called "task calibration", actually capture what FTT would predict for transitive reasoning. The fuzzy processing preference indicates that children tend to reason at the least precise level of gist that permits correct solutions. For the equality tasks, this is the reductional strategy (a categorical or nominal level of gist). For the simultaneously presented inequality tasks, this is the positional strategy (relational or ordinal gist; more refined than categorical or nominal gist). For the tasks that differ visually in length, the visual strategy is used (discussed in detail by Reyna & Brainerd, 1990).

The traditional information-processing approach did not offer a full picture of the aspects of task calibration from FTT. Trabasso and his colleagues did not use simultaneous presentation, but at the same time they took a preventive measure against forgetting the premises by building up a representation in the memory of their subjects through excessive training of

the premises. Visual strategies were prevented by not showing the length differences, and reductional strategies by not presenting equality tasks. In general, none of the studies according to the traditional information-processing approach systematically analysed the relations between presentation procedure, task format, and reasoning strategy. For a theoretical understanding of these relationships FTT appears most helpful. Our first study offered empirical support for the central idea of task calibration; that is, children reason at the least precise level of gist that permits correct solutions. As this study was restricted to tasks of length, future studies might also include tasks of size and weight.

Piaget (Piaget & Inhelder, 1967, 1968) considered transitive reasoning to be a symptom of the ability to reason logically. We decided to stay as close as possible to this original concept of transitivity. We thus concluded from our first study that a test for transitive reasoning should minimise switching to other than deductive strategies. In our second study we therefore only used tasks: (1) with inequality relations; (2) with minimal length differences; (3) that were presented successively; and (4) for which the use of the deductive strategy was checked by requesting an explanation of a response.

Mokken (1971, 1997; Mokken & Lewis, 1982) scale analysis was used to construct a strong ordinal scale based on nine transitivity tasks. The nine tasks all required the same ability characterised as verbal reasoning about transitive relations on the basis of a deductive strategy. The scale permits the reliable ordering of individuals and is, therefore, an efficient instrument with which to investigate individual differences.

Our second study had several limitations. First, use of other than the deductive strategy could not be completely avoided. For length and weight deductive reasoning was induced by 69% and 53% of the T-items, respectively (Table 4). However, 69% of the T-items for size induced a visual strategy (Table 4). This may be the result of the visually more impressive differences between the areas of the objects compared to their diameters. Area thus seemed easier to distinguish visually, hence the increasing use of visual strategies. In future research the “just noticeable” difference between areas should be determined in advance.

Individual differences in transitive reasoning had a weaker relation with age than with grade. The largest increase was between Grades 4 and 5. In the Dutch educational system it is customary to introduce fractions and transitive reasoning in the fifth grade, and this may be the main cause of this difference. The fact that so little age-related development was found may be due to the sample. According to the Piagetian tradition transitive inference should first emerge in the concrete-operational period. It is recommended, therefore, that younger children (from five years on) be included in future research samples. Some interesting relations were found between performance on the transitive reasoning scale and aspects of arithmetic and

early mathematics, such as insight into the number line and number structure. The results are limited, however, and call for additional validity research.

It was demonstrated how important disagreements in the literature can be reduced to differences in task formats and presentation. We constructed a series of tasks triggering children's transitive reasoning skills by means of a deductive strategy. These tasks constituted a unidimensional scale, offering an important starting point for the construction of an optimal device to investigate the development of transitive reasoning in a more advanced, psychometric way than is customary in the developmental research tradition.

Manuscript received August 1996

Revised manuscript received February 1998

REFERENCES

- Bleichrodt, N., Drenth, P.J.D., Zaal, J.N., & Resing, W.C.M. (1985). *Revisie Amsterdamse kinder intelligente test* [Revised Amsterdam intelligence scale for children]. Lisse, The Netherlands: Swets & Zeitlinger.
- Braine, M.D.S. (1959). The onthogeny of certain logical operations: Piaget's formulation examined by nonverbal methods. *Psychological Monographs*, 73(5), whole No. 475.
- Brainerd, C.J. (1973). Judgments and explanations as criteria for the presence of cognitive structures. *Psychological Bulletin*, 79, 172-179.
- Brainerd, C.J. (1974). Training and transfer of transitivity, conservation, and class-inclusion of length. *Child Development*, 45, 324-334.
- Brainerd, C.J. (1977). Response criteria in concept development. *Child Development*, 48, 360-366.
- Brainerd, C.J., & Kingma, J. (1984). Do children have to remember to reason? A Fuzzy-Trace theory of transitivity development. *Developmental Review*, 4, 311-377.
- Brainerd, C.J., & Kingma, J. (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology*, 17, 210-247.
- Brainerd, C.J., & Reyna, V.F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, 10, 3-47.
- Brainerd, C.J., & Reyna, V.F. (1992a). Explaining "memory-free" reasoning. *Psychological Science*, 3, 332-339.
- Brainerd, C.J., & Reyna, V.F. (1992b). The memory independence effect: What do the data show? What do the theories claim? *Developmental Review*, 12, 164-186.
- Breslow, L. (1981). Reevaluation of the literature on the development of transitive inferences. *Psychological Bulletin*, 89, 325-351.
- Bryant, P.E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature*, 232, 456-458.
- Chapman, M., & Lindenberg, U. (1988). Functions, operations, and decalage in the development of transitivity. *Developmental Psychology*, 24, 542-551.
- Chapman, M., & Lindenberg, U. (1992a). Transitivity judgments, memory of premises, and models of children's reasoning. *Developmental Review*, 12, 124-163.
- Chapman, M., & Lindenberg, U. (1992b). How to detect reasoning-remembering dependence (and how not to). *Developmental Review*, 12, 187-198.

- Cito (Centraal Instituut voor Toetsontwikkeling). (1979). *Rekentoetsen leerjaar 2, 3 en 4 basisonderwijs* [Arithmetic tests for grades 2, 3 and 4 of primary education]. Arnhem, The Netherlands: Cito.
- DeBoysson-Bardies, B., & O'Regan, K. (1973). What children do in spite of adults hypothesis. *Nature*, 246, 531–534.
- Flavell, J.H. (1972). An analysis of cognitive developmental sequences. *Genetic Psychology Monographs*, 86, 279–350.
- Flavell, J.H. (1977). *Cognitive Development*. Englewood Cliffs, NJ: Prentice Hall.
- Ginsburg, H. (1977). *Children's arithmetic: The learning process*. New York: Van Nostrand.
- Halford, G.S., & Galloway, W. (1977). Children who fail to make transitive inferences can remember comparisons. *Australian Journal of Psychology*, 29, 1–5.
- Halford, G.S., & Kelly, M.E. (1984). On the basis of early transitivity judgments. *Journal of Experimental Child Psychology*, 38, 42–63.
- Harris, P.L., & Bassett, E. (1975). Transitive inference by four-year-old children. *Developmental Psychology*, 11, 875–876.
- Hemker, B.T., Sijsma, K., & Molenaar, I.W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Applied Psychological Measurement*, 19, 337–352.
- Hemker, B.T., Sijsma, K., Molenaar, I.W., & Junker, B.W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika*, 62, 331–347.
- Kallio, K.D. (1982). Developmental change on a five-term transitive inference. *Journal of Experimental Child Psychology*, 33, 142–164.
- Kingma, J. (1983). Seriation, correspondence and transitivity. *Journal of Educational Psychology*, 75, 763–771.
- Kingma, J., & Koops, W. (1983). Piagetian tasks, traditional intelligence and achievement tests. *British Journal of Educational Psychology*, 53, 278–290.
- Kingma, J., & Reuvekamp, J. (1984). The construction of a developmental scale for seriation. *Educational and Psychological Measurement*, 44, 1–23.
- Kingma, J., & TenVergert, E.M. (1985). A nonparametric scale analysis of the development of conservation. *Applied Psychological Measurement*, 9, 375–387.
- Levinova, L.A. (1977). Oriëntaties van kleuters in relaties tussen grootheden [An introduction for kindergarteners to the relations between quantities.] In C.F. van Parreren & J.M.C. Nelissen (Eds.), *Rekenen*. Groningen, The Netherlands: Wolters-Noordhoff.
- Meijer, R.R., Sijsma, K., & Smid, N.G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement*, 14, 283–298.
- Mokken, R.J. (1971). *A theory and procedure of scale analysis*. Berlin: De Gruyter.
- Mokken, R.J. (1997). Nonparametric models for dichotomous responses. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 351–367). New York: Springer.
- Mokken, R.J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, 6, 417–430.
- Molenaar, I.W. (1970). *Approximations to the Poisson, binomial and hypergeometric distribution functions*. Amsterdam: Mathematisch Centrum.
- Molenaar, I.W. (1997). Nonparametric models for polytomous responses. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 369–380). New York: Springer.
- Molenaar, I.W., Debets, P., Sijsma, K., & Hemker, B.T. (1994). *User's manual MSP*. Groningen, The Netherlands: iecProGAMMA (Rijksuniversiteit Groningen).
- Murray, J.P., & Youniss, J. (1968). Achievement of inferential transitivity and its relation to serial ordering. *Child Development*, 39, 1259–1268.
- Perner, J., & Aebi, J. (1985). Feedback-dependent encoding in length series. *British Journal of Developmental Psychology*, 3, 133–141.

- Perner, J., & Mansbridge, D.G. (1983). Developmental differences in encoding length series. *Child Development*, 54, 710–719.
- Perner, J., Steiner, G., & Steahelin, C. (1981). Mental representation of length and weight series and transitive inferences in young children. *Journal of Experimental Child Psychology*, 31, 177–192.
- Piaget, J. (1942). *Classes, relations et nombres: Essai sur les groupements de la logistique et sur la réversibilité de la pensée*. Paris: Collin.
- Piaget, J., & Inhelder, B. (1941). *Le développement des quantités chez l'enfant*. Neuchâtel, Switzerland: Delachaux et Niestlé.
- Piaget, J., & Inhelder, B. (1967). *La genèse des structures logiques élémentaires*. Neuchâtel, Switzerland: Delachaux et Niestlé.
- Piaget, J., & Inhelder, B. (1968). *Le développement des quantités physiques chez l'enfant*. Neuchâtel, Switzerland: Delachaux et Niestlé.
- Piaget, J., & Szeminska, A. (1941). *La genèse du nombre chez l'enfant*. Neuchâtel, Switzerland: Delachaux et Niestlé.
- Reyna, V.F. (1992). Reasoning, remembering, and their relationship: Social, cognitive and developmental issues. In M.L. Howe, C.J. Brainerd, & V.F. Reyna (Eds.), *Development of long-term retention* (pp. 103–127). New York: Springer.
- Reyna, V.F., & Brainerd, C.J. (1990). Fuzzy processing in transitivity development. *Annals of Operations Research*, 23, 37–63.
- Reyna, V.F., & Brainerd, C.J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, 7, 1–75.
- Riley, C.A., & Trabasso, T. (1974). Comparative, logical structures and encoding in a transitive inference task. *Journal of Experimental Child Psychology*, 22, 1–22.
- Russell, J. (1981). Children's memory for the premises in a transitive measurement task assessed by elicited and spontaneous justifications. *Journal of Experimental Child Psychology*, 31, 300–309.
- Sijtsma, K., & Junker, B.W. (1996). A survey of theory and methods of invariant item ordering. *British Journal of Mathematical and Statistical Psychology*, 49, 79–105.
- Sijtsma, K., & Molenaar, I.W. (1987). Reliability of test scores in nonparametric item response theory. *Psychometrika*, 52, 79–97.
- Sijtsma, K., & Verweij, A.C. (1992). Mokken scale analysis: Theoretical considerations and an application to transitivity tasks. *Applied Measurement in Education*, 5, 355–373.
- Sternberg, R.J. (1980). Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General*, 109, 119–159.
- Thayer, E., & Collyer, C. (1978). The development of transitive inference: A review of recent approaches. *Psychological Bulletin*, 85, 1327–1343.
- Trabasso, T. (1975). Representation, memory and reasoning: How do we make transitive inferences? In A.D. Pick (Ed.), *Minnesota symposia on child psychology* (Vol. 9). Minneapolis, MN: University of Minnesota Press.
- Trabasso, T. (1977). The role of memory as a system in making transitive inferences. In R.V. Kail & J.W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 333–336). Hillsdale, NJ: Erlbaum.
- Verweij, A.C., Sijtsma, K., & Koops, W. (1996). A Mokken scale for transitive reasoning suited for longitudinal research. *International Journal of Behavioral Development*, 19, 219–238.
- Youniss, J., & Dennison, A. (1971). Figurative and operative aspects of children's inference. *Child Development*, 42, 1837–1847.
- Youniss, J., & Murray, J.P. (1970). Transitive inference with nontransitive solutions controlled. *Developmental Psychology*, 2, 169–175.