Training in Measurement

WELKO TOMIC
The Open University, The Netherlands

JOHANNES KINGMA
ELISABETH M. TENVERGERT
University Hospital Groningen, The Netherlands

ABSTRACT
The study investigated the effects of Obuchova’s (1966, 1972) method of teaching children how to measure. The subjects were 30 kindergarten children who showed no pretest knowledge of either conservation or seriation. Children from the training condition (n = 15) received 3½ weeks of training. Training appeared to be highly effective. A broad near-far transfer was observed; that is, skills were transferred to conservation tasks not taught in training. Far-far transfer (i.e., transfer to concepts not included in training) was also observed, because the children were able to solve a broad range of seriation tasks for which they had received no training. This is a noteworthy result, because far-far transfer has rarely been reported in training research. These effects persisted for 4 months. The educational importance of this result is that by means of a broadly designed course of training, strong and long-lasting near-far and far-far transfer effects may be induced. Training did not, however, evoke sleeper effects, because trained and untrained children performed at the same level 2 years after training.

For example, to have learned conservation skills during training, a subject must (a) give correct answers on tasks set during training (near-near transfer tasks, which concerns identical or almost identical tasks used in the training and in the (post) test); (b) give correct answers to questions testing a concept that uses stimuli different from the stimuli used in training (near-far transfer tasks); (c) give logically acceptable explanations for the answers given on the tests in a and b; and (d) pass a test for at least one other concept related theoretically to conservation, for example, correctly anticipating the water level in horizontality tasks, and so forth (far-far transfer tasks); and (e) continue to perform Steps a to d successfully several days or weeks after training. Brainerd (1978a) argued that this Genevan standard is insensitive, because it detects only massive changes in behavior.

In his review, Brainerd (1978a) found that investigators outside Geneva usually view Steps a, b, and c as a more adequate indication that learning has occurred. (In addition, in many training studies these training effects appeared to persist weeks or even months after training.) In other words, most researchers outside Geneva are using a range of either near-near or near-far transfer tasks as a standard to evaluate the effectiveness of training, whereas the Geneva standard consists of near-near, near-far and far-far transfer tasks. Progress on far-far transfer tasks is considered particularly credible evidence that a change in the child’s cognitive structure has occurred. Far-far transfer effects have seldom been reported or investigated outside Geneva (Field 1998). The purpose of the present study was to show that far-far transfer effects may be induced when training nonconservers to use Gal’perin’s (1957, 1966, 1967, 1972) principles of measurement.

According to the Genevan perspective on learning, a child’s stage of development determines the amount that he or she can learn (Inhelder & Sinclair, 1969; Inhelder,
Sinclair, & Bovet, 1974; Piaget, 1957, 1964, 1975; Sinclair, 1973). Applying their standard to training experiments involving nontransition children (i.e., children who do not show any pretest knowledge of the concept in which they will be trained), Genevan researchers have become skeptical about training effects. Specifically, non-transition children make little or no progress as a result of training. In contrast, children who have a partial knowledge of the concept in which they will be trained (i.e., children who are "in transition" between preoperational and operational thought) may benefit from training and make moderate-to-good progress on both near-far and far-far transfer tasks. However, training methods based on Galperin's 1967 theory have produced far-far transfer effects in nonconservers (Burmenskaja, 1976; Lider, 1978; Obuchova, 1966, 1972). According to Galperin, when a child fails in a Piagetian task, it is because he or she does not possess the "rational object scheme." This scheme is a rational structure that can be superimposed on the appearance of objects and allows the child to compare objects quantitatively. A child who does not possess the rational object scheme is misled by the dominant feature of the object (e.g., after transformation in a conservation task), because the characteristics of the object are not perceived as independent features.

Obuchova (1966, 1972) taught nonconservers the concept of measurement to induce the rational object scheme. Children were trained not to rely on their direct perception of differences and similarities between objects, but rather on a well-based relational judgment achieved by using measuring tools or units (for example, a scale or cardboard strips). Obuchova's training method consisted of five phases. In the first phase, the child was trained to compare two sets of objects. Each element of the first set was assigned a plastic red circle, and each element of the other set a blue circle. Subsequently, the circles were placed in two parallel rows, and the child compared the two rows to determine which of the two sets had more elements. In the second phase, the child was taught to compare two objects on length by means of another object (e.g., a cardboard strip). In the third phase, a measuring unit was introduced for comparing length, volume, area, and so forth. For example, the child was asked to compare the weight of a clothespin and the weight of a pair of scissors. This comparison was not made directly, but by placing one object in the right hand and the other in the left, but rather indirectly, by placing one of the objects in one tray of a scale and coins in the other tray until the two trays balanced. The coins were then arranged in a row. In Phase 4, the child was taught to analyze different features of the same objects (e.g., objects were first compared according to volume and subsequently an analysis was made of their weights with the help of measuring units).

According to Obuchova (1966, 1972), by using measuring units in these four phases, a child can construct a new model for the relationships between objects, which becomes apparent by comparing the two parallel rows of these units. This schematic representation of the essential relationship between objects forms the outward embodiment of what will subsequently become the child's internal plan of relational judgment (the rational object scheme). In the fifth phase, the child was taught to apply the rational object scheme while performing Piagetian conservation tasks.

In the beginning, some children were misled by the transformation of the object. However, by using the measuring units, they were able to perform these conservation tasks correctly. Subsequently, the children were able to solve all types of conservation problems (both first- and second-order invariants) correctly without taking measurements. Furthermore, the training effect persisted one month after training. Burmenskaja (1976) obtained similar results by applying Obuchova's method. In addition, she also observed a far-far transfer effect, that is, trained children were able to solve horizontality in anticipation of seriation tasks on the posttest. Lider (1978) found that this method was also successful in inducing class-inclusion.

The present study investigates the effect of Obuchova's (1966, 1972) training method on conservation and seriation performance. The main concern was to analyze both near-far and far-far transfer. The study also corrects certain methodological weaknesses of Obuchova's and Burmenskaja's (1976) experiments; that is, matched control groups were used, and a distinction was made between Phase 5 of the training program and the first posttest, which in previous training studies had been interwoven. A second concern was to study the aftereffect of training. An interval of 2 or 4 months satisfies both Piaget's (1975) and Galperin's (1967) stringent criteria for successful training. Different posttests were therefore administered at different time intervals after training was completed.

Although these two research questions are most important from the educational point of view, we also investigated a possible sleeper effect. In the literature the suggestion has been made that long-term aftereffects or improvements in performance have been observed in some studies. An aftereffect, or sleeper effect, occurs a long time after training has been implemented. In some cases, a sleeper effect is observed years after the course of training has ended. If a sleeper effect is likely to occur, the research design should include at least one delayed posttest in addition to the immediate posttest (Ball, 1985). A large-scale training program, as in the present study, may induce a long-term sleeper effect (Krech, Crutchfield, & Ballachey, 1962). Although we administered a posttest 4 months after training ended, a sleeper effect remained possible. This sleeper effect was therefore also investigated to analyze the long-term—that is, 2 years—impact of training.
Method

Design

During the pretest, children from five kindergarten classes were administered conservation, seriation, and relational-term tasks as well as the Culture Fair Intelligence Test, Form 1, abridged version (Cattell, 1950). One day after training, the first posttest was administered (consisting of conservation and seriation tasks) to the trained children and their counterparts in the control group. The same tasks were also given in the second and third posttest, respectively, 1 week and 4 months after the first posttest. The sleeper effect—the fourth posttest—was assessed 2 years after the administration of the first posttest.

Subjects

The subjects were 30 Dutch kindergarten children selected from a sample of 220 children from five different kindergarten classes who were administered conservation, seriation, and relational-term tasks and the Institute for Personality and Ability Testing (IPAT), Form 1, abridged version (Cattell, 1950). In this sample, 147 children were unable to solve any of the conservation or seriation tasks. From these 147 children, 15 pairs of children (i.e., both complete nonconservers and complete nonseriators) were selected who could be matched on IQ, age, and sex. In each matched pair, one child was assigned at random to the training group and the other to the control group. The training group \( (n = 15, \, 9 \text{ boys and } 6 \text{ girls}) \) had a mean age of 65.5 months, \( SD = 8.2 \), and a mean IQ of 104.2, \( SD = 11.2 \). The control group \( (n = 15, \, 9 \text{ boys and } 6 \text{ girls}) \) had a mean age of 64.5 months, \( SD = 8.2 \), and a mean IQ of 104.5, \( SD = 11.3 \).

Materials

Pre- and posttests. The pretest consisted of 13 conservation, 12 seriation, and 24 relational-term tasks and the Culture Form Intelligence Test, Form 1, abridged version Cattell (1950). Only conservation and seriation tasks were administered during the posttests.

Conservation tasks. Before the 13 conservation tasks were administered, an introductory trial was given to familiarize the child with the questions. Subsequently, the 13 conservation tasks were administered. (For a detailed description, see Kingma, 1983a, 1983b, 1984a; Kingma & Koops, 1983; Kingma & TenVergert, 1985.) The procedure for each task was similar. At first, two identical objects were presented and the child was asked whether these objects were the same for the relevant feature (e.g., length, area). Following this, a transformation of one of these two objects was performed, and finally the child was asked whether the transformed object was the same as the nontransformed one for that feature. The following conservation tasks were completed: area, number, distance, substance, perimeter, quantity, discontinuous quantity, weight, covered area, length, discontinuous length, geometric volume, and physical volume.

In all 13 conservation tasks, two identical objects were used. To avoid inducing a response set, we also administered two nonequivalent conservation tasks (between Tasks e and f, and i and j) in which the two objects did not match on the relevant feature before transformation.

Seriation tasks. Of the 12 seriation tasks, 6 were derived from Piaget (1957, 1964, 1975) (the traditional seriation tasks) and the remaining 6 contained irrelevant cues, that is, the stimuli varied on properties other than the one according to which they had to be seriated (see Figure 1). In

Figure 1. A Schematic Representation of the Seriation Tasks With Irrelevant Cues (Tasks T to Y)
an introductory trial, the child had to seriate (make a "staircase" of) five puppets. The following six traditional seriation tasks (Tasks 1 to 6) were subsequently administered (for a detailed description, see Kingma, 1983b, 1983c; Kingma & Reuvenkamp, 1984): (a) Single seriation according to length consisted of 10 tubes to be seriated according to length. (b) Multiple seriation according to length and color contained five series of five tubes. Within each series the tubes differed according to length, whereas the series in succession showed an increase in color intensity from light to dark blue. (p) Single seriation according to weight was performed with five identically shaped cubes of different weights. (q) Multiple seriation according to weight and color contained five series of identically shaped blocks. Within each series the blocks differed on weight, whereas the series in succession showed an increase in color intensity from light to dark blue. (r) Single seriation according to size contained four plywood squares to be seriated according to increasing size. And (s) Multiple seriation according to size and color contained five series of aluminum squares. Within each series the squares differed on size, whereas the series in succession showed an increase in color intensity from light to dark blue.

Following the traditional tasks, six single seriation tasks (1 to 6) with irrelevant cues were administered (for a detailed description of these tasks, see Kingma, 1984b; Kingma & Loth, 1985; Kingma & Reuvenkamp, 1984). (i) The single-seriation tasks were according to length with width as an irrelevant cue, (u) according to width with length as an irrelevant cue, (v) according to height with both length and width as irrelevant cues, (w) according to length with weight, width, and thickness as irrelevant cues, (x) according to weight with height, length, and width as irrelevant cues, (y) according to length with smaller differences in length between the successive tubes (2 cm) than those commonly used in Piagetian studies. These tasks were more difficult than the traditional seriation task (Kingma & Reuvenkamp, 1984), enabling us to investigate the range of far--far transfer in the present study.

Relational-term tasks. Eight concept areas (the same used in the conservation tasks) were selected: number, proportion, weight, volume, distance, substance, area, and length. For each concept area, three questions (requiring a response of more, less, or the same) were posed in random order. Altogether there were 24 items (for a detailed description, see Kingma 1983b; Kingma & Loth, 1984).

Training. The main measuring tool in each phase of training was the workshelf, on which the measuring units were placed in two parallel rows to allow the child to compare the two objects along the relevant feature. Two parallel rows with corresponding marks were drawn on the workshelf to provide a clear one-to-one correspondence between the elements of the two rows: coins, paper clips, and so forth (see Figure 2).

In Phase 1, the materials consisted of 28 sheets of paper and two sets of plastic circles, 35 blue in one set and 35 red in the other. More than 18 identical objects were printed on each sheet (e.g., ducks, dogs, cats). In Phase 2, cardboard strips were used to measure pairs of objects printed on 15 sheets of cardboard. In Phase 3, measuring cups were used to assess volume. Squares and triangles served as measuring units for comparing the areas of different school yards. A cardboard strip was used to measure differences in length between staircases printed on a sheet of paper. For the comparisons of weight, both a scale and a lever were used and coins or paper clips served as measuring units. In Phase 4, we used the materials from the preceding phases. The materials used in the six conservation tasks in Phase 5 were different from the materials used in the pretest and posttests (for a more detailed description, see Appendix).

Scoring

Pretest and posttests. A child was given a score of 1 for a relational-term item when he or she correctly picked the object in the row that was more, less, or the same as the standard. A score of 0 was given for an incorrect response.

The scoring system devised by Goldschmidt and Bentler (1968) was used for the conservation items. A score of 1 was awarded when both a correct judgment and an explanation were given (for a detailed description, see Kingma, 1984a). All other cases resulted in a score of 0. This scoring system was chosen because in previous research (Goldschmidt & Bentler, 1968; Kingma, 1984a), it had provided highly reliable tests for conservation skills. In addition, this system is more stringent than the system relying on the judgment-only criterion. Progress as measured according to the former system may therefore be considered more substantial than when it is measured according to the latter system.

A child received a score of 1 for a seriation task when he or she produced a correct series. An incorrect series resulted in a score of 0. The Culture Fair Intelligence Test, abridged form (Cattell, 1950), was scored according to the instructions given in the test manual.

Recording the course of training. For each exercise, the researcher posed questions such as "Are all the paper clips taken together as heavy as the ear?" The child's responses to these questions were recorded on a sheet. When the child gave an incorrect response, he or she was told to repeat the measurement. The number of repetitions needed to reach a correct answer was recorded, as well as the time needed to complete an exercise successfully.

Procedure

The pretest and posttests were administered individually by 14 experienced researchers who, during the posttests, did not know whether a child belonged to the training
group or to the control group. Approximately the same number of children were assigned at random to each researcher. Testing took place in two sessions when the children's motivation and concentration made this necessary. In the majority of cases, the first session focused on the conservation and relational-term tasks, whereas the second session focused on the seriation tasks and, during the pretest, the Cattell, (1950) Form I. The two sessions taken together varied between 25 and 50 min.

While the tasks were being administered, two researchers observed the child independently. The researcher to whom the child was assigned tested the child and also recorded the child's explanations for his or her judgments (after transformation) in the conservation tasks, as well as the results obtained on the seriation and relational-term items. Another researcher (seated separately from the child and the first researcher) also recorded the same observations to assess interrater reliability. Before the start of the present study, these observers had received 1 week of training in administering the tests and coding the explanations given by the children. Prior to actual testing, all the observers had reached a reliability criterion of 95% agreement.

Training. Following the pretest, the children assigned to the training group received treatment each school day for a period of 3 1/2 weeks. Each training session took 10 to 30 min. Training proceeded in fixed order from Phase 1 to Phase 5, as described in the Appendix. The children belonging to the training condition were assigned at random to seven other researchers. Each child was trained individually by the same researcher, who was not one of those administering the posttests. At the beginning of each training phase, the child was given the opportunity to become familiar with the questions and told that the workshelf and the measuring units were important tools for making comparisons between two objects on relevant features. In each phase of training (Phases 1 to 3), the researchers used only the objects and measuring units necessary to complete the task, so that distraction was reduced to a minimum. During Phases 4 and 5 of the training, all units were displayed.

When the child gave an incorrect response in a training trial, the researcher asked: "How do you know that?" or "Can you demonstrate that?" The child was then asked to perform the measurement again until he or she gave a correct response (to stimulate overlearning). When the child gave a correct response, the researcher pointed out that using the measuring units would help to reach a more reliable comparison between the two objects than relying solely on visual inspection.

Within the same time frame, the control group counterpart of a trained child had to complete tasks from the kindergarten program (e.g., drawing) under the supervision of another researcher, so that a comparison could be made between the training group and the control group on motivational factors.

Results

Inspecting the Data

For each pair of researchers, the interrater reliability of the recorded observations was computed separately for the pretest and the four posttests. Applying the Goldschmidt and Bentler (1968) criterion, the researchers recorded four correct explanations—those based on qualitative identity, addition-subtraction, reversibility, and compensation. The product-moment correlations for the seven pairs of researchers ranged from .96 to .99, respectively, for the qualitative identity arguments on the four posttests. Interrater agreement on addition-subtraction, reversibility, and compensation arguments varied for the four posttests between .95 and .98, .95 and .99, and .98 and .99, respectively. (There was perfect interrater agreement on the explanations given on the pretest.) These results agreed with the results of previous research into interrater reliability concerning the accurate assessment of explanations (Kingma, 1984a). Perfect interrater agreement was found for the recorded judgments given for both the conservation and relational-term tasks, as well as for the recorded results of the seriation tasks.

All of the children, in both the training and the control groups, solved all 24 relational-term items correctly on
the pretest. These tasks were therefore not administered in the subsequent posttests.

**Training Effects**

The mean total scores on the 13 conservation tasks, the 6 traditional seriation tasks, and the 6 seriation tasks remained stable for the trained children on the first three posttests, whereas only slight progress in these tasks was observed for the control group (see Table 1). However, no sleeper effect was observed; the performance of the control group improved greatly on the fourth posttest (2 years after training) for these three types of tasks compared with their performance on the first three posttests. The training group also showed a slight increase in the mean total scores on the fourth posttest.

The total scores on the 13 conservation tasks were analyzed in a $2 \times 4$ (Subject Groups $\times$ Posttests) repeated-measures analysis of variance (ANOVA). This ANOVA showed significant effects for groups, $F(1, 28) = 39.3, p < .01$, and posttests, $F(3, 8) = 14.1, p < .01$, and a significant Groups $\times$ Posttests interaction, $F(3, 84) = 5.3, p < .01$. Overall comparisons among groups using the Newman–Keuls procedure indicated that the total scores of the training group on the conservation tasks were significantly higher than those of the control group on the first three posttests ($p < .01$). On the fourth posttest, the conservation performance of the training group did not differ significantly from that of the control group. Thus, it is clear that training in measurement successfully induced near–far transfer in conservation skills, and that this transfer effect persisted at least 4 months after training (i.e., the third posttest). An effect that persists for 4 months is sufficient for educational purposes.

The total scores of the six traditional seriation tasks showed a similar picture for a $2 \times 4$ (Groups $\times$ Posttests) ANOVA with repeated measurements. Significant main effects were found for groups, $F(1, 28) = 61.09, p < .01$, and for the posttests, $F(3, 84) = 45.10, p < .01$, and a significant Groups $\times$ Posttest interaction, $F(3, 84) = 45.14, p < .01$, was also observed.

Concerning the total scores on the six seriation tasks with irrelevant cues, the $2 \times 4$ (Groups $\times$ Posttests) ANOVA with repeated measurements resulted in findings similar to those obtained in the analysis of the traditional seriation data. Significant main effects were observed for Groups, $F(1, 28) = 32.46, p < .01$, and for the posttests, $F(3, 84) = 32.99, p < .01$, as well as a significant Groups $\times$ Posttest interaction, $F(3, 84) = 5.41, p < .01$. Overall, comparisons among groups using the Newman–Keuls procedure indicated that the training group had significantly higher scores on the first three posttests than did the control group for both the traditional and the seriation tasks with irrelevant cues. However, on the fourth posttest (i.e., 2 years after training), the performance of the training group did not differ significantly from that of the control group in both types of seriation tasks. No sleeper effect was observed on the fourth posttest after training.

These results demonstrated that training in measurement had induced a strong far–far transfer, that is, the children were able to generalize their newly acquired skills to different concept areas within the domain of seriation for which they had not received training. This transfer was effective up to 4 months after training, which is sufficient for educational purposes.

**Type of Transfer**

Two nonequivalent conservation tasks (in which the objects differed on the relevant feature before transformation) were also included on each posttest, to investigate whether the children might show a mechanical response pattern. On the first posttest, only 2 children gave the same judgment and explanation for one nonequivalent conservation task as they gave for an equivalent conservation task. The 2 children gave zero and three correct solutions, respectively, for the 13 equivalent conservation problems on the first posttest. The remaining 13 children from the training group solved the nonequivalent conservation problems correctly. Similar results were found for the second, third, and fourth posttests. Thus, it is reason-

<table>
<thead>
<tr>
<th>Posttest</th>
<th>Condition</th>
<th>Conservation</th>
<th>Traditional seriation</th>
<th>Seriation with irrelevant cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Training</td>
<td>7.8</td>
<td>5.0</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.7</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>Training</td>
<td>8.3</td>
<td>5.1</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>1.1</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>Training</td>
<td>7.6</td>
<td>4.9</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>2.1</td>
<td>1.5</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>Training</td>
<td>9.7</td>
<td>5.7</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>7.3</td>
<td>5.5</td>
<td>4.1</td>
</tr>
</tbody>
</table>

Note. Maximum score was 13. Minimum score was 6.
able to suppose that the training effect did not induce a mechanical response pattern.

Furthermore, each posttest task was analyzed for level of difficulty. The children in the training group apparently found it difficult to perform the two conservation tasks of volume, as well as the tasks involving conservation of area and conservation of distance. Although in Phase 3 of training the children were taught to measure volume and area, the six exercises alone may not have been sufficient, because 60% of the trained children were unable to perform these tasks on the first, second, and third posttests. However, although the children had not been taught to measure perimeter, about 46% were able to perform this task correctly on the first three posttests.

Most of the correct solutions to the conservation tasks (on the first, second, and third posttests) were in the concept areas of number, substance, quantity, weight, and length.

With respect to the traditional seriation tasks, the majority of trained children were able to perform both the single- and multiple-seriation tasks correctly on the first three posttests. Only 1 child in this group was able to perform only one or two single seriation tasks on these posttests. This child also had a score of 0 for the conservation tasks on the first posttest. Because the trained children had no pretest knowledge of seriation, one might consider this type of far-far transfer very strong. Upon closer inspection, the data obtained from the first three posttests on the seriation tasks with irrelevant cues led to the same conclusion. On the first posttest, 2 of the children produced a correct seriation in one task alone, whereas the majority of the children produced three or more correct series in these six tasks. The second and third posttests showed similar results. The 2 children with the lowest scores for these tasks also had the lowest scores for the conservation tasks on the first three posttests.

Training Effects and the Course of Training

During training, the researcher recorded the number of repetitions needed to reach a correct solution for each exercise, the amount of time required to complete all the exercises, and the number of times that a child used the workshelf during Phase 5 of training. Arguably, a large number of repetitions might indicate that the child had difficulty learning the concepts during training. The number of repetitions for all training exercises varied between 3 and 150. In general, the majority of the repetitions occurred at the beginning of each training phase. This finding suggests that the children learned the principle of indirect comparisons during each phase by using the measuring units. However, no clear correlation was found between the number of repetitions and posttest performance on the conservation tasks (the Spearman rank correlation was .14, ns). In addition, the amount of time required to complete all exercises did not relate clearly to posttest performance on the conservation tasks ($r = .07$, ns). A moderate correlation was found ($r = .36$, ns) between IQ and the total scores for the conservation tasks on the first posttest.

In contrast, observations about the use of the workshelf and other measuring units in Phase 5 provided the best information about the effectiveness of training. Three of the children, the same 3 who scored lowest for the conservation tasks on the first posttest, used the measuring units as many as four or five times during this phase to perform these tasks. Four of the children used the workshelf once or twice only. The remaining 8 children did not use the measuring units at all.

Discussion

Training in measurement based on Obuchova's (1966, 1972) method was successful in teaching nonconservers both first- and second-order invariants. First-order invariants refer to the conservation concepts of early operational thought (e.g., conservation of number, substance, quantity, and length), whereas second-order invariants are mastered at the end of concrete operational thought (conservation of perimeter, area, and volume). Training produced a broad near-far transfer, because skills were transferred to concept areas of conservation for which no training had been given. These results confirmed the earlier findings of Burmenskaja (1976) and Obuchova (1966, 1972). This transfer is noteworthy, because only in Phase 5 of training was the child taught to perform the conservation tasks. The amount of conservation training may be considered negligible, however, because the child used only the measuring units when he or she was unable to reach a correct solution.

The use of these units (after an incorrect response) may be viewed as a type of feedback. However, the majority of the children ($n = 12$) made little use of these units. The four preceding phases of training in measurement had therefore already induced a cognitive scheme in these children, enabling them to perform the conservation tasks correctly. In contrast, the children ($n = 3$) who frequently made use of the workshelf and other measuring units in Phase 5 had the lowest scores on the posttests. In other words, the feedback obtained by using these measuring units did not improve performance on conservation tasks to any great degree. However, use of the workshelf in Phase 5 may indicate whether a child learned the rational object scheme in the preceding phases of the training.

The training program, which was based on the Obuchova (1966, 1972) method, also induced far-far transfer (i.e., a child can transfer acquired skills to concepts in areas other than the ones in which he or she was trained). The majority of the trained children who did not show any pretest knowledge of seriation reached correct solutions on the posttests for both the traditional seriation tasks (single and multiple seriation) and the seriation tasks with irrelevant cues. The latter tasks were shown to
be more difficult in previous research (Kingma, 1984b; Kingma & Reuvenkamp, 1984) because the stimuli varied on properties other than the one to be seriated. This is a noteworthy result, as far-far transfer has seldom been investigated and thus rarely observed in training research.

The question remains why this training method induced far-far transfer in seriation. Although learning to compare objects along relevant features is an important aspect of learning to construct a correct series, we cannot assume that training in measurement alone is responsible for this far-far transfer. In addition to the ability to make correct comparisons, a correct seriation also requires the ability to integrate objects into a series of transitive relationships. The training exercises also taught the subjects how to make a transitive comparison. For example, after a child had measured objects A and B and saw that they consisted of x and y units, respectively, he or she was able to deduce from the relationship between x and y the transition relationship between A and B (without comparing A and B directly). An explanation for far-far transfer as it occurred in the present study might be that learning both measurement and transitive inference also made it possible for a child to perform the seriation tasks.

With respect to the range of both the near- and far-far transfer effect, a comparison can be made between the numerical magnitude of this training effect and the results of previous cross-sectional and longitudinal research involving the same conservation and seriation tasks (Kingma, 1982, 1983a, 1983b, 1983c, 1983d, 1983e, 1984a, 1984b, 1984c; Kingma & Koops, 1983; Kingma & Loth, 1985; Kingma & Reuvenkamp, 1984; Kingma & Roeilinga, 1984; Kingma & TenVergert, 1985). A comparison between the results of the first posttest for both seriation and conservation and those obtained in a longitudinal study (Kingma, 1984b) revealed that the trained children (mean age 65 months) performed similarly to children aged 78 to 80 months on the first posttest. In other words, previously obtained empirically based standards concerning the development of both conservation and seriation skills provide information about the effectiveness of training. The effect of training for both near-far and far-far transfer persisted for at least 4 months.

A strong aftereffect such as this may be considered sufficient for educational purposes. However, a sleeper effect was not observed 2 years after training had taken place. The training and control groups showed only a negligible difference. From an educational point of view, however, this result cannot be considered surprising, because it agrees with what has been observed in daily educational practice.

Piaget (1975) and Galperin (1967) strongly agree on how to measure the success of a training session: the effect of training must meet a stringent criterion. Both theorists considered far-far transfer evidence for a change in a child's cognitive structure. However, the seriation tasks chosen for the present study may be less adequate for assessing whether operational thought has been induced. Piaget and Inhelder (1967) used anticipation of seriation tasks to assess whether a child was an operational seriator. This approach was also used by Sinclair (1969) to investigate whether her training method induced operational thought. These tasks were not used to measure far-far transfer in the present study, because Kingma (1982) demonstrated that 20% of kindergartners are already able to perform these anticipation tasks correctly, whereas they are unable to construct a correct series. The seriation tasks with irrelevant cues were therefore used to assess transfer of seriation skills. Because previous research has shown that these tasks are more difficult than the traditional seriation tasks (Kingma, 1984b), in the present study they served as an empirically based standard used to analyze the extent of the far-far transfer effect.

Although a number of researchers consider the requirement set by both Piaget (1957, 1964, 1975) and Galperin (1957, 1966, 1967, 1972) for far-far transfer too stringent, in our study, the effects of the training program based on Obuchova's (1966, 1972) method fulfilled this requirement.

APPENDIX

The following exercises were performed in successive phases of training. Phase 1. The child was shown a piece of cardboard with two different images: ducks and fish scattered criss-cross. The number of each image exceeded the child's counting ability. The child was asked which of the two types of images had more elements. Subsequently, the child was given red and blue plastic circles. The child's task was to put a red circle on each fish and a blue circle on each duck. Following this, the blue circles were placed in a row on the worksheet and an image of a duck was placed at the beginning of the row. This same activity was also performed for the fish. Finally, the child's task was to derive, from the one-to-one correspondence of the two rows of circles, which of the images had more elements. A total of 14 of these exercises (with different amounts and different types of images) were performed. Phase 2. The child was taught to compare the lengths of two objects by means of a third object. The child was given a sheet of paper on which two objects with small differences in length appeared (e.g., 2 keys, 2 brushes, 2 combs). The child was therefore unable to compare the two objects directly. A cardboard strip was used to find which of the two images was longer. The strip was placed on one image. The length of that image was marked down with a pencil and the strip was cut at that point. Following this, the child had to measure the other image with this same strip, and decide whether the two images were of the same length. A total of 15 exercises with different kinds of images were performed. Phase 3. In the preceding phases the images had been compared with the aid of concrete objects, although these did not function as measuring units for the children. In the third phase of training, the child learned to use these concrete objects more specifically as measuring units. For example, the researcher taught the child to compare the length of two staircases with a cardboard strip (which served as a measuring unit). The child's task was to measure the length of the staircase with this cardboard strip. For each step (or part of the staircase whose length corresponded with the length of the strip), a blue circle was placed on the worksheet. The child followed the same procedure for the other staircase and placed a red circle on the worksheet for each step. The child was then asked to compare the lengths of the two staircases. In other tasks, the child learned to use the units measuring volume, surface, and weight in the same way. To compare volume, the child placed a circle on the worksheet for every measuring cup with grain that he or she poured into a container. To measure the area of rectangular school yard, the child covered the rectangles with identical cardboard squares. For each cardboard square, a token was placed on the worksheet. A scale and
coins (or paper clips) were used to measure weight. After balancing the first object with the coins, the child placed these coins in a row on the workshelf. The same procedure was also followed for the other objects. Phase 3 consisted of five types of exercises: (a) six length-comparison tasks; (b) eight weight-comparison tasks using a lever and paper clips; (c) six weight-comparison tasks using a scale and coins; (d) six area-comparison tasks; and (e) six volume-comparison tasks using a measuring cup.

Phase 4. The child was taught to analyze different features of objects. Tasks were assigned in which two objects had to be compared according to different parameters (e.g., both weight and length, volume and weight). The child had to choose the units that could measure these parameters. Before beginning the tasks, the child saw all of the measuring units used in the preceding phase displayed on the table. In each of the eight trials, the child had to compare two objects on different features and demonstrate the comparison with the suitable measuring unit.

Phase 5. The child had to apply the rational object scheme to Piagetian conservation tasks. Six conservation tasks were presented: conservation of weight, quantity, area, length, substance, and number. The child was allowed to use the measuring units employed in the preceding training phases to perform these conservation tasks. There were two differences between Obuchova's (1966;1972) training approach in Phase 5 and that of the present study: (a) When the child reached a correct solution for a conservation task, he or she was not asked to demonstrate that solution in the present study (to minimize response-set learning). Only when the conservation task had been performed incorrectly did the researcher ask the child to demonstrate with the help of a measuring unit whether his or her solution was correct. (b) In the present study, we separated this phase of training and the first posttest to avoid a possible misinterpretation of the training effect.

REFERENCES


