On the effects of training inductive reasoning: How far does it transfer and how long do the effects persist?

Welko Tomic
The Open University, Heerlen, The Netherlands

Karl Josef Klauer
University of Aachen, Germany

Using the same program, two training experiments have been conducted in a Dutch and in a German elementary school. The common expectation was that training in inductive reasoning would transfer both on intelligence tests measuring inductive reasoning and on math performance. Furthermore, it was expected that the training effects would persist for at least some months after training had ended. In experiment 1 (N=34), a rather short training period turned out to be effective with respect to the intelligence test performance but not with respect to math performance. In experiment 2 (N=23), the amount of training in inductive reasoning was systematically varied. It could be shown that transfer on intelligence test as well as on math performance was linearly dependent on the amount of prior training. The training effects were found to persist between four and nine months after training.

Problem-solving is seen to be a crucial human activity that should have a central place in school curricula. It is, therefore, no wonder that a great number of research projects have concentrated on studying children's problem-solving ability and its development. For instance, since the early sixties, cognitive developmental psychology has devoted a great deal of attention to experiments in which children are trained to acquire concepts described in Piaget's theory. The majority of this research has focused on the empirical aspects of the Genevan perspective on learning, i.e., on concepts of mid-childhood (Brainerd, 1978a, 1978b; Tomic, Kingma, & TenVergert, 1993).

Problem solving calls for various types of solution skills (Plomp, 1990; Gephart, 1972). In any given field most problems require the problem-solver to apply a variety of reasoning skills to reach a solution. One important problem-solving strategy is inductive reasoning. Induction is the process whereby one generalizes across a number of instances in order to find a description that applies to them all. This process allows the problem-solver to establish regularities or order. By applying inductive reasoning, we are able to make inferences which some-
times exceed the knowledge we possess at the moment (Holland, Holyoak, Nisbett, & Thagard, 1986). Induction is obviously a process that is very important in allowing us to find regularities in and make predictions about novel environments.

Inductive reasoning plays an important role in the psychometric approach to intelligence and was conceptualized as reasoning ability by Spearman (1904). An early research study on visual inductive reasoning was carried out by Luria in central Asia in the 1930s (Luria, 1976). In order to assess inductive reasoning tasks, one can make use of analogies, series, classifications and matrices (Büchel & Scharnhorst, 1993; Goldman & Pellegrino, 1984; Sternberg & Gardner, 1983).

Klauer’s (1989, see also Klauer & Phye, 1994) assumption is that inductive reasoning tasks can be solved by performing mental comparisons. Comparison processes provide the basis for the discovery of perceivable similarities and differences. By analyzing these similarities and differences, one is not only able to find regularities and order, but is also able to detect only seeming regularities, seeming order. According to Klauer, the comparison process consists of both cognitive and metacognitive components, such as strategy choice, selective encoding, etc. The assumption is that the solution procedures can be learned and that inductive tasks can be solved successfully using these previously acquired procedures.

The aim of Klauer’s program is to help children acquire a strategy of inductive reasoning which can be modified in different ways according to the inductive reasoning tasks distinguished by the author. Transfer is not expected to occur automatically, for instance by a process of stimulus generalization; instead, it is necessary to teach transfer by offering a great number and variety of problem situations. Even the most decided advocates of the (somewhat fashionable) notion of situated learning accept the possibility of transfer, in the event that learning is a kind of guided participation that bridges old and new experiences (Rogoff, 1990), that an action schema is acquired which is related to features of situations and to the activities within the situations (Greeno, Smith, & Moore, 1993), or that students learn to explore common features and relationships between old and new situations (Resnick, 1987). Although these authors are skeptical as far as transfer from traditional school learning to out-of-school activities is concerned, they nevertheless make allowances for the fact that transfer does occur. In a seminal paper, Perkins and Salomon (1989) reminded us of the generality-power tradeoff: general skills are widely applicable but of limited use in a concrete problem situation, whereas very restricted usable strategies are extremely effective where they apply (Anderson, 1987).

The strategy of inductive reasoning, as it is understood in this article, represents neither a general nor a specific strategy but a domain-related one, and it is taught by a great number of, on the surface, very different problems giving the children the opportunity to detect relevant commonalities. Although we are dealing with a domain-specific strategy, it is a rather important domain. Several authors assume that induction processes foster concept formation (Holzman, Pellegrino, & Glaser, 1983), text comprehension (Greeno, 1978), or knowledge acquisition in school settings (Norman, Gentner, & Stevens, 1976; Curtis, 1988; Curtis & Reigeluth, 1984). Hence, it may be concluded that research into the training of inductive reasoning is theoretically as well as practically relevant.

Transfer is considered to be an indication that a more or less long-lasting change has occurred in the child’s ability to solve a certain type of problem induced by a training program. To evaluate the success of training, stringent transfer standards can be used. There is a standard for transfer of training which goes from less stringent to stringent criteria. For example, to have learned particular skills during training, a subject must (a) give correct answers on tasks set during training (near-near transfer tasks, which concern identical or almost identical tasks used in training and in the posttest; (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; (d) pass a test on at least one other concept related theoretically to that particular skill (far-far transfer tasks); and (e) continue to perform steps a to d successfully several days or weeks after training. 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 (Brainerd, 1978a), whereas the
Genevan and Russian standard consists of near-near, near-far and far-far transfer tasks. (Piaget, 1957, 1964, 1975; Inhelder, Sinclair, & Bovet, 1974; Brainerd, 1975; Galperin, 1957, 1966, 1967; Tomic, Kingma, & Ten Vergert, 1993). Progress on far-far transfer tasks is considered particularly credible evidence that a change in the child’s cognitive structure has occurred. Assessing the range of transfer is thus important.

The purpose of the present study is to investigate the effects that Klauer’s training program may have on children’s performance of inductive reasoning tasks. Several questions were addressed within the framework of this study.

1) Can the training program be used effectively by inexperienced trainers who are not under the supervision of the author of the program? This question is important because most of the research on the training program was performed by Klauer’s staff members or students.

2) Does the training program improve performance on intelligence tests assessing inductive reasoning? To answer this question, suitable intelligence tests were administered as pretests and posttests with training and control groups.

3) Does the transfer of the training effects to intelligence test performance last for several months? Delayed posttests were given to investigate the question of durability of the effects.

4) Do the training effects transfer to certain performances in mathematics? Math tests were administered to detect possible effects of far transfer. The durability of this transfer was also tested in a delayed posttest.

The Heerlen experiment

The purpose of the experiment was to replicate Klauer’s training program and measure the effects on the performance of children on tests measuring inductive reasoning. Replications are generally commendable. They give an idea of the stability of the effects and a subsequent meta-analysis leads to an estimation of the effect size to be expected when anybody applies the training program anywhere. This replication differs from many others insofar as the trainers were not experienced in working with the program; neither had they been instructed by the author of the program. Additionally it was intended to assess possible far-far transfer of training on certain mathematics performances.

Method

Subjects. Thirty-four primary school children of approximately seven years from the same class served as subjects in the experiment. The children were randomly assigned to either the treatment or the control condition. The Dutch third grade level was chosen for comparison purposes with similar populations in previous training studies (Klauer, 1989, 1990a, b, 1992; Klauer & Phye, 1994; Phye & Sanders, 1993; Resing & Verbraeken, 1993). Both groups consisted of 17 children and the proportion of girls and boys was held constant between the groups. The descriptive statistics are given in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Training group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age</td>
<td>IQ</td>
</tr>
<tr>
<td></td>
<td>86.1</td>
<td>109.1</td>
</tr>
<tr>
<td>M</td>
<td>3.4</td>
<td>14.6</td>
</tr>
</tbody>
</table>

Table 1

Age in Months and IQ of the Children
Pretest. Cattell's Culture Fair Test CFT 1 was used as a pretest (Cattell, 1950). It is a nonverbal intelligence test consisting of five subtests and a well-known measure of fluid intelligence. Contrary to the training program, the test makes use of abstract, meaningless material. The first two subtests—substitutions and mazes—do not assess inductive reasoning as it is defined by Klauber (1989, see below), while the other three subtests—classifications, similarities, and matrices—do.

Training. Klauber (1989) designed a program for the development of inductive reasoning strategies (see also the American version by Klauber & Phyre, 1994). The rationale of the program is that regularities among objects are induced by strategies of comparison, i.e., by a search for similarities and/or differences, either with attributes of objects, or with relations. Comparing objects in order to find out (a) similarity or (a) difference or (a) both with respect to (b) attributes or (b) relations leads to \(3 \times 2 = 6\) different types of strategies. Accordingly, it is possible to distinguish six kinds of inductive reasoning tasks or problems: generalization (similarity between attributes), discrimination (difference between attributes), cross-classification (similarity and difference between attributes), recognizing relations (similarity between relations), discriminating relations (difference between relations) and system construction (similarity and difference between relations). The relationship between these kinds of tasks are depicted in Figure 1. Sample items are presented in the Appendix.

![Diagram](attachment:image.png)

**Figure 1.** The Generalalogy of Tasks in Inductive Reasoning

The training program consists of 120 problems, i.e., each kind of inductive problem and problem-solving strategy is represented by 20 items. Only 15 percent of the items use abstract, meaningless material. Most of the remaining 85 percent consist either of concrete material, blocks the children can manipulate, or pictorial-figural problems taken from the every-day lives of the children. In this way, the children learn to apply the strategy to a great variety of problem situations but the transfer distance between training and test material remains considerable.

Training procedure. Following the pretest, the children assigned to the training group received the treatment each school day for a period of one week. Each of the five training sessions took about 30 minutes. The children assigned to the training group were trained in three groups of six, six, and five by three researchers, who were not involved in administering the
pretests and posttests. At the beginning of the first training session, the children were given the opportunity to become familiar with the problems. When a child gave an incorrect response in a training session, the researcher asked: “How do you know that?” or “Can you demonstrate that?” The child was then asked to perform the item again until he or she gave a correct response. Within the same time frame, the control group counterpart of a trained child had to complete tasks from the school program.

The training program is designed for ten training sessions. Since only five sessions could be given, the trainers tried to administer more than the usual twelve items per session. In fact, they succeeded in working on 80 to 90 instead of 60 problems. This implies that the children did not complete the whole program and that none of the abstract items could be dealt with because the latter are given at the end of the program. Moreover, it implies that the researchers had to lower their expectations a bit concerning the training effects. The results of this experimental study are, nevertheless, interesting because it may often happen in educational practice that a program cannot be completed.

CFT Posttest. The first or immediate posttest was administered one day after finishing the training procedure. Four months later, the test was given again as a delayed posttest. Since the first two subtests of the CFT 1 do not assess inductive reasoning, the intention was to test the hypothesis concerning the improvement of inductive reasoning using the three remaining subtests. Due to force majeure, it was not possible to carry out this plan. Flooding along the Meuse River destroyed the children’s answer sheets before full use could be made of them. Before this mishap, only one sum score had been recorded: the sum of the three inductive subtests and the substitution subtest. Substitution tests measure the speed with which children learn simple paired-associates. It cannot be expected that training in inductive reasoning fosters this kind of learning ability. Thus, the sum score does not represent an optimal variable to reveal effects of the training.

Mathematics Posttest. During training, the children were encouraged to scan objects in order to find relations between them and to identify the attributes which they had in common. This involved using different kinds of objects with which the children were familiar, but not, for instance, numbers. A test was therefore also constructed requiring the subjects to detect relations between numbers (the first three examples below) or to detect common attributes between numbers (last example). The test consisted of four subtests corresponding to the four examples below and had a total of 21 items.

| Fill in the correct numbers: |
|---|---|---|
| □+□=□ | 9 | 5 | 4 |
| □>[□=□ | 6 | 13 | 12 |
| □−□=□ | 17 | 3 | 20 |

One of the numbers on the right also belongs to the set in the circle:

```
    □□□□□□
    □□□□
    □□□
    □□
```

| 10 | 30 | 40 |
| 50 | 21 | 33 |
| 20 | 17 |

*Figure 2. Sample items of the MATH test*

Children of this age possess the mathematical knowledge required to solve problems of that kind. They are not, however, sophisticated enough to uncover the relationships between such numbers as, for instance, 5, 9, and 4, nor are they used to finding the common attributes
of such numbers (e.g., one-digit numbers). In any case, the transfer distance between training and test items is rather large, because there are no numbers at all included in the training program. The test was administered one day after the children finished their training and again four months later. All of the tests were given as group tests. The person administering the posttests did not know which children belonged to which group.

**Results**

The means and standard deviations of both groups and of the three tests are given in Table 2. In order to adjust immediate and delayed posttest scores in accordance with differential performance on the pretest, a 2 (Group: Trained vs. Nontrained) x 2 (Time: Immediate vs. Delayed Posttest) analysis of covariance was performed with the pretest as covariate. Trained children outperformed the untrained children significantly on the immediate posttest and on the delayed posttest four months after training ($F(1,31)=4.13; p=.026$ and $F(1,31)=3.97$, $p=.028$, respectively, both one-tailed). Corrected effect sizes $E_{SCorr}$ have been calculated. They are corrected for the small pretest differences between the groups:

$E_{SCorr}=E_{posttest}-E_{pretest},$

where $ES=(MTG-MCG/sp)$ and where $sp$ is the pooled standard deviation. The corrected effect size amounts to $E_{SCorr}=0.39$ for the immediate and $E_{SCorr}=0.54$ for the delayed posttest. Four months after training, an average trained child outperformed an average control child by about one half standard deviation.

### Table 2

**Raw Scores of the Training and Control Group (N1=N2=17)**

<table>
<thead>
<tr>
<th>Tests</th>
<th>Training Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CFT*</td>
<td>MATH</td>
</tr>
<tr>
<td></td>
<td>$M$  $SD$</td>
<td>$M$  $SD$</td>
</tr>
<tr>
<td>Pretest</td>
<td>26.24  3.01</td>
<td>---</td>
</tr>
<tr>
<td>Posttest 1</td>
<td>31.76  3.46</td>
<td>16.59  4.98</td>
</tr>
<tr>
<td>Posttest 2</td>
<td>34.59  1.23</td>
<td>17.76  3.85</td>
</tr>
</tbody>
</table>

*Note.* *without subtest “Mazes”*

To assess the far-far transfer of Klauer’s training program, children’s scores on the mathematics test administered following training were analyzed. A 2 (Group: Trained vs. Nontrained) x 2 (Time: Immediate vs. Delayed) analysis of covariance was performed with the pretest as covariate. Results indicated that there was no significant transfer of training on the mathematics achievement.

The effect sizes $ES$ also showed smaller effects on math achievement. Immediately after training, the effect size amounted only to $ES=0.18$. Four months later, the effect size reached a somewhat higher value of $ES=0.31$.

**Discussion**

The experiment was conducted to examine the effects of Klauer’s training program on inductive reasoning in children of average ability. Training in inductive reasoning based on Klauer’s program (1989, Klauer & Phye, 1994) was shown to be successful in teaching the inductive reasoning strategy, at least as it is mirrored by the children’s CFT achievement.
Moreover, the durability of the training effects on CFT achievement has been demonstrated in this study. The effect size yielded certainly did not drop during a follow-up four months later. It even increased, though this rise was not statistically significant. These results confirmed the earlier findings of Klauser (1989), Phye and Sanders (1993), and Resing and Verbraeken (1993).

However, no transfer effect on math performance could be established. The effect sizes found are too small to reach statistical significance, especially with the small number of subjects given. One explanation could be that the training program was not administered completely. Another explanation could be that there was not enough opportunity for the children to apply the newly acquired strategy to math learning because there were only a few math lessons between the end of training and the math test.

The Aachen experiment

This experiment was intended to replicate the main results of the Heerlen experiment under different conditions, and to take an additional aspect into consideration. A control group such as the Heerlen one is a tried-and-tested means for assessing a treatment effect compared with the effect of schooling: Both groups are instructed, the one by means of normal classroom activities, the other by means of the special training program. Any improvement in the control group can be traced back to retest effects and to the effects of schooling, so that any superior effect of the training group can be attributed to the special treatment. This design, however, does not establish any control for warming-up effects (Phye, 1987) or for the effects due to particular teacher-student interactions within a small group session (Klauser, 1989). Although there is no good reason to expect any effect from this interaction, especially on the CFT performance or on a comparable test, one cannot be sure whether it affects general achievement motivation in one direction or another. That is the reason why we introduced non-inductive reasoning training as a control condition. It was set up as a placebo training condition insofar as it did not include any inductive reasoning activities but nevertheless maintained the typical small-group training situation.

Moreover, it was intended that the training and control condition would not be an all-or-nothing affair but that every child received both types of treatment although in different amounts. Under these conditions, even linear relationships can be expected between the amount of inductive-reasoning training as independent variable and the dependent variables. Testing such a strict hypothesis can lead to a strong argument for the effectiveness of inductive-reasoning training.

Method

Subjects. Twenty-eight children enrolled in a German elementary school participated in this training study. They resembled the Dutch children in that the school did not have many children from underprivileged backgrounds and in that the children were, on the average, nearly seven years old. The children were all from one class. They were randomly assigned to four groups which received four particular types of treatments.

Tests. Instead of the CFT, we used Raven’s Coloured Progressive Matrices CPM as a measure of inductive-reasoning ability. All of its items are matrices and its regularities can be detected by looking for common relationships. Hence, it is a valid measure of inductive reasoning according to Klauser’s definition. If students solve the items using different non-inductive processes (Hunt, 1974), the hypothesis would be rejected by the data because inductive-reasoning training would not turn out to be of any use in this case.

The CPM was administered as a pretest, a posttest and as a delayed posttest. The time span between pretest and posttest was two months, and between posttest and delayed posttest nine months.
Not immediately but three months after the training had ended, the same math tests as in
the Heerlen experiment were administered to the children. Six months later, i.e., nine months
after completing training, the children were given another math test, adjusted to their progress
in learning arithmetic. The same item types as in the Heerlen study were used but the numbers
were larger. For instance, one of the addition items was
\[
\square + \square = \square / 35 \quad 80 \quad 22 \quad 37 \quad 43
\]

The subjects had to fill in three of the numbers given so that a correct solution resulted.
Such a task is more easily resolved if one can quickly find the relationships between the num-
bers. Note that not only are the numbers larger, but that two irrelevant options have also been
added, distractors which increase problem difficulty.

Training. One of the training programs used was the same as had been employed in the
Heerlen experiment, namely Klauer's program of inductive reasoning (Klauer, 1989). The
alternative was a perceptual training program similar to the Frostig program (Frostig, Maslow,
& Miller, 1972; Frostig, Horne, & Miller, 1977), for average-ability children (Frostig & Hart,
1967). The training exercises dealt with perception of one's body, with recognizing objects
only by touching them, or by hearing their characteristic noise, and with time and space per-
ception in the form of different social games. Space perception, for example, was practised by
moving from one wall to the other blindfolded, with obstacles such as balls lying in the way
and with the blindfolded child only relying on other children to tell her or him where to turn.
Many of the exercises were taken from Mertens (1991) or from Eggert (1992). None of the
perceptual tasks fulfilled Klauer's definition of an inductive reasoning problem.

The children enjoyed this perceptual training and were eager to participate. Every child
participated in both training programs, but in different degrees. The schedule assigning groups
of children to the training sessions can be seen in Table 3. Unfortunately, most of the children
assigned to group 1 could not be tested because the parents refused permission to do so. For
that reason groups 1 and 2 were combined to form group 2. This change, however, does not
detract from the basic expectation: if inductive-reasoning training has a transfer effect on
dependent variables at all, then the effect should linearly increase with the number of the group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Training Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inductive</td>
</tr>
<tr>
<td>1 (n=2)</td>
<td>1</td>
</tr>
<tr>
<td>2 (n=7)</td>
<td>4</td>
</tr>
<tr>
<td>3 (n=7)</td>
<td>7</td>
</tr>
<tr>
<td>4 (n=8*)</td>
<td>9</td>
</tr>
</tbody>
</table>

Note. * One child could not participate in the delayed posttests.

Both training programs were conducted with small groups of about three to six children.
The instructor was a very experienced psychologist known for achieving fine results in inductive-reasoning training.

Results

The means and standard deviations for each of the groups and for the different dependent
variables are depicted in Table 4. As one can see, there are small differences with CPM pretest
favouring the group which had the largest proportion of perceptual training. The differences are, however, not significant. The means of the immediate posttest of the CPM increase linearly with the increasing number of the groups, whereas for the delayed posttest of the CPM, which was administered nine months later, a curvilinear relation seems to hold (Figure 3).

![Figure 3. Achievement Development of the Aachen Training Groups](image_url)

Three months after training had ended the test MATH 1 was given, and six months later a similar, but more difficult, test MATH 2. The means reported in Table 4 again show a linear relationship between the means of the immediate test MATH 1 and the number of the group, and, possibly, a curvilinear relationship between the delayed test MATH 2 and group number.

### Table 4

**Means and Standard Deviations according to Group and Dependent Variable**

<table>
<thead>
<tr>
<th>Group</th>
<th>CPM</th>
<th></th>
<th></th>
<th>MATH</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest 1</td>
<td>Posttest 2</td>
<td>test 1</td>
<td>test 2</td>
<td></td>
</tr>
<tr>
<td>2 (n=9)</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>3 (n=7)</td>
<td>24.7</td>
<td>4.5</td>
<td>29.2</td>
<td>5.4</td>
<td>30.3</td>
<td>6.3</td>
</tr>
<tr>
<td>4 (n=7)</td>
<td>22.7</td>
<td>3.1</td>
<td>32.3</td>
<td>3.8</td>
<td>37.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Total (n=23*)</td>
<td>23.8</td>
<td>5.7</td>
<td>31.7</td>
<td>4.8</td>
<td>34.5</td>
<td>5.5</td>
</tr>
</tbody>
</table>

Note: * One child could not participate in the test MATH 1 and four children not in the CPM delayed posttest and the test MATH 2

Can these findings be substantiated statistically? To answer this question each dependent variable was dealt with in a two-fold way. First, an analysis of covariance was performed with the criterion as the dependent variable, the group as the independent variable, and the CPM pretest as the covariate. This analysis tells us whether or not there are treatment effects at all. Second, the residual gains were computed by removing the influence of the CPM 1, and subsequently a test of linearity between the groups was performed using the residual gain scores. All of the tests are one-tailed because we expected only improvements due to the inductive-reasoning treatment or, respectively, a linearly increasing instead of a decreasing relationship between group and gain.
As to the Raven tests, the immediate posttest CPM posttest yielded a significant group effect ($F(2,22)=4.01, p=.0175$), and the same was true for the delayed posttest CPM three months later ($F(2,18)=4.15, p=.0185$). The tests of linearity for the expected trends also turned out to be significant: $F(1,20)=.86, p=.0055$ for the immediate posttest and $F(1,16)=4.92, p=.0205$ for the delayed posttest. With the latter test, however, an unexpected tendency to a quadratic component appeared ($p=.068$). The adjusted means of the posttests are given in Figures 4 and 5.
The ANCOVA with the immediate test MATH 1 as the dependent variable failed by only a small margin to show significant group differences ($F(2,21)=1.93$, $p=.0865$), quite in contrast to the ANCOVA with the delayed test MATH 2 six months later ($F(2,18)=3.13$, $p=.0365$). In both cases, however, the linear trend could be established ($F(1,19)=3.63$, $p=.036$ for MATH 1 and $F(1,16)=4.02$, $p=.031$ for MATH 2). The adjusted means for both tests are demonstrated in Figures 6 and 7. Only a faint tendency of a quadratic component can be recognized with test MATH 2 ($p=.13$).

Since no control group without any training was included in this research, regular measures of effect sizes cannot be determined. In order to assess the observed effect sizes, contrasts between group 2 on the one hand and the combined groups 3 and 4 have been made use of. Thus, the effect sizes of Table 5 show the effects of a higher proportion of inductive-reasoning training as compared to a lower proportion. It is possible that the real effect sizes have been somewhat underestimated in this way.

### Table 5

**Effect Sizes ES of the Inductive-Reasoning Training**

<table>
<thead>
<tr>
<th>Variable</th>
<th>CPM posttest</th>
<th>CPM delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MATH 1</td>
<td>MATH 2</td>
</tr>
<tr>
<td>ES</td>
<td>1.13*</td>
<td>1.51*</td>
</tr>
<tr>
<td>CPM posttest</td>
<td>0.80</td>
<td>0.72</td>
</tr>
</tbody>
</table>

*Note: * ES<sub>corr</sub> effect size corrected for pretest differences

### Discussion

Cohen (1988) defines a medium effect size as ES=0.5 and a large one as ES=0.8. In this research, rather large effect sizes have been arrived at, even though all the children received the same amount of training. With respect to the dependent variables used, it can be concluded that inductive-reasoning training turned out to be much more effective in fostering inductive reasoning than perceptual training. In a way, perceptual training can be conceived of as a placebo training condition insofar as one would not expect this type of training to have any effect on the cognitive variables in question. It could only make allowances for effects due to the atmosphere in small groups, to group interaction, to the training experience and so on. Even if the effect sizes are not actually underestimated, they are remarkably large. Presumably,
this is due to the fact that the trainer was very experienced in administering inductive-reasoning training.

The effects on the Raven tests have been replicated in many studies (Klauer, 1993, 1995). It is astonishing, however, that this effect lasted over nine months and that it did not fade at all during this time span. To explain this phenomenon, one can assume that the strategy of inductive reasoning acquired during training was not lost later on but remained in active use. Only in this way can the fact be explained that three months after the end of training, a training effect on math performance still persisted and did not fade even during the following half year. If the inductively trained children had not made use of the acquired reasoning strategy later on, one would have expected effects to disappear at least some months after training.

Concluding Remarks

Looking at the results of the Heerlen and the Aachen studies, the pattern of effects induced by inductive-reasoning training seems quite clear and at least partially consistent. Training was effective in both cases in improving inductive reasoning as measured by Cattell’s CFT 1 or by Raven’s CPM. Note that the transfer distance from training problems to test problems is fairly large insofar as the training program does not offer one single item of the kind found on the tests. In fact, most training items are based upon concrete objects or on pictorial problems taken from the children’s everyday experiences, whereas the test items are meaningless and abstract in nature.

The effect size of the Heerlen study is, however, definitely smaller than the effect size found in Aachen. This difference can, possibly, be traced back to cultural differences between the countries, particularly as the training program was designed for German children. On the other hand, it may be assumed that the cultural differences are not that important since training studies using the same program in the USA and in the Netherlands were rather successful (Klauer & Phye, 1994; Resing & Verbraken, 1993). The small to medium effects found in Heerlen can be accounted for, after all, by the fact that the training program could not be administered in its entirety and additionally by the fact that the dependent variable was not a pure measure of inductive reasoning. Another explanation is possible by stressing that the trainers in the Heerlen experiment were inexperienced in inductive-reasoning training, whereas the German psychologist was known for achieving substantial training effects. Finally, we might question whether there is anything in particular at all to be explained, because there are some German training studies which have led to even smaller and nonsignificant effects (Klauer, 1995).

Noteworthy is the fact that in the Heerlen study a long-lasting effect could also be observed. Four months after training, the training effect was still significant and had even increased in size. Comparable results were found in the Aachen study: the effect remained stable in size and also statistically significant even nine months later. It can be assumed that in both cases the children made use of the acquired reasoning strategy not only during the training period but also in the period thereafter. This seems to be an encouraging finding and has significance for the research literature pertaining to self-regulated learning.

There is a difference between the two experiments as far as the transfer on the mathematical test is concerned. A long-lasting transfer on math performance was yielded in Germany but not in the Netherlands. Two explanations are possible to account for this difference. The first stresses the fact that the training effect in Heerlen was smaller than that in Aachen, and it is plausible that a smaller effect might show less far-reaching transfer value. There are, however, many studies that yielded no larger effects on tests of fluid intelligence but which nevertheless did yield significant effects on learning classroom subject matter (Klauer, 1995).

The second possible explanation refers to the fact that there was a large time difference between training and testing of math performance in the German experiment but not in the Dutch. In the latter experiment, the children had only five days to acquire the strategy of
inductive reasoning, and if little relevant math learning took place during this time, they had little opportunity to apply the recently acquired strategy. For the experiments mentioned earlier that showed a transfer on learning to school subject matter, training was followed by a specifically planned learning phase. After the lesson, tests were given exploring how much of the new declarative knowledge the students had acquired and whether or not the inductively trained subjects had acquired more knowledge. In a learning situation like this, the newly mastered strategy of inductive reasoning could come into play and improve learning of the subjects trained. Within the present Aachen experiment there was no specific lesson but a time span of three months during which the new strategy could enhance math learning and, possibly, other learning, too. It is evident that a newly acquired strategy which is supposed to enhance learning needs opportunities to come into play during learning situations. These opportunities were given in the Aachen experiment simply by administering the math achievement test some months later.

Notes

1 We would like to express our gratitude to the teachers and students of De Nuinhof elementary school for their participation. Special thanks are extended to Ms. Buskens, Rika Verhoeof and Ellie Roolofs for their assistance in training the children and collecting the data. This study was supported by a grant from The Open University.

2 Our gratitude is expressed to the psychologist Mrs. Tatjana Igelmund and to the teachers, pupils, and parents of the Aachen elementary school Gut Kullen.

References


Appendix

No. 5. Which three objects belong together? (Generalization: class formation)

No. 18. Have the animals enter the tent according to size (Recognizing relationships: series)
No. 87. Which figure on the right best fits in the empty square? (System construction: matrix)

Deux entraînements expérimentaux utilisant le même programme ont été conduits dans une école élémentaire hollandaise et allemande. On s’attendait à ce que l’entraînement au raisonnement inductif ait des effets positifs sur la réussite à des tests de raisonnement inductif et sur les performances en mathématiques. On s’attendait par ailleurs à ce que les effets se maintiennent dans les mois suivant l’entraînement. Dans l’expérience 1 (N=34), un entraînement d’une durée relativement limitée eu des effets sur les performances aux tests de raisonnement mais n’en eu pas sur les performances en mathématiques. Dans l’expérience 2 (N=23), on a fait varier systématiquement la durée de l’entraînement. On a constaté que les effets aussi bien au test de raisonnement qu’en mathématiques dépendaient de façon linéaire de la durée des entraînements préalables. Les effets persistent entre quatre et neuf mois après la fin de l’entraînement.

Key words: Inductive reasoning, Math learning, Training research, Transfer of training.

Wenko Tomic. The Open University, Faculty of Social Sciences, P.O. Box 2960, 6401 DL Heerlen, The Netherlands.

Current theme of research:
Research in cognitive development and learning.

Most relevant publications in the field of Psychology of Education:


**Karl Klauer.** Institut für Erziehungswissenschaft, Eilfschornsteinstr, 7, D-52056 Aachen, Germany.

*Current theme of research:*

Training of inductive reasoning in children and youths.

*Most relevant publications in the field of Psychology of Education:*


