

Inducing Inductive Reasoning: Does It Transfer to Fluid Intelligence?

Karl Josef Klauer and Klaus Willmes

University of Aachen, Aachen, Germany

and

Gary D. Phye

Iowa State University

Published online July 13, 2001

Based on a prescriptive theory of inductive reasoning, a training program to foster inductive reasoning has been developed. Children from 12 first-grade classes, mean age about 7 years, $N = 279$, participated in a training experiment. The children of 6 classes were trained to apply a strategy to reason inductively while the children of the remaining classes continued their regular classroom activities. It was expected that trained children would outperform the untrained children with respect to Raven's Coloured Progressive Matrices but not with respect to a vocabulary test, thus indicating convergent and discriminant or domain-specific training effects. Results confirmed this expectation. Moreover, it was expected that training would improve performance on the inductive subtests of Cattell's Culture Fair Test 1, but not influence subtests that did not involve inductive reasoning. Considerable transfer to both kinds of subtests was found on the immediate transfer task. However, with a delayed posttest 6 months later, the expected differential training effect could be observed. Finally, a LISREL model analysis confirmed the hypothesis that training children to reason inductively improved fluid but not crystallized intelligence.

© 2001 Elsevier Science (USA)

INTRODUCTION

Inductive reasoning is usually contrasted to deductive reasoning: "Induction means establishing, deduction means applying rules" (Shye, 1988). Thus, inductive reasoning enables one to detect regularities, rules, or generalizations and, conversely, to detect irregularities. This is one way in which we structure our world.

It seems useful at the outset, to distinguish between inductive reasoning

Address correspondence and reprint requests to Gary Phye, N162B Lagomarcino, Iowa State University, Ames, IA 50011-3190. E-mail: gdphy@iastate.edu.



and inductive inferring. Inductive reasoning is aimed at detecting generalizations or regularities. If, for instance, a number of objects is given and if it is found that all of them are toys made of wood, a generalization or regularity has been discovered. Should we extend this generalization to the totality of toys by stating that all toys are made of wood, then we would have made an inductive inference, although a false inference in this case. An inductive inference extends the generalization beyond the scope of experience by asserting something about a nonobservable universe of objects. Inductive reasoning, however, is confined to the observation at hand. It discovers regularity and order within a given set of objects.

There is agreement among researchers that inductive reasoning constitutes a central aspect of intellectual functioning. Ever since Spearman (1923), there has been no doubt about the close relationship between inductive reasoning and intelligence. Inductive reasoning is usually measured by tests consisting of classifications, analogies, series, and matrices (Goldman & Pellegrino, 1984; Sternberg & Gardner, 1983; van de Vijver, 1991). Many intelligence tests contain one or more subtests of these varieties so that the contribution of inductive reasoning to intelligence test performance is beyond question.

At least four important waves of research contributed to our knowledge about the relationship between inductive reasoning and intelligence. Spearman, the founder of the *factor analytical tradition*, was convinced that his general intelligence factor g was mainly determined by inductive processes ("education of relations"). Thurstone (1938) used a different factor analytic approach, which led him to a concept of multiple intelligence factors. One of these was the factor "Reasoning" that is made up of a combination of inductive and deductive tests. Cattell (1963) found an adequate solution by making the distinction between fluid and crystallized intelligence. Fluid intelligence, g_f , is primarily involved in problem solving, whereas crystallized intelligence, g_c , is involved in acquired declarative knowledge. Fluid intelligence can be understood as at least partially determined by genetic and biological factors, while the crystallized factor is conceived of as a combined product of fluid intelligence and education. Vocabulary tests are typical markers of the crystallized factor, whereas inductive tests typically serve as markers of the fluid factor. Using the method of linear structural equations (LISREL), Gustafsson (1984) as well as Undheim and Gustafsson (1987) confirmed Cattell's theory of fluid and crystallized intelligence. Undheim and Gustafsson also concluded that inductive processes play a major role in fluid intelligence.

Continuing research interest in inductive reasoning and fluid intelligence has prompted *cognitive researchers* to engage in analyzing the processes that occur when subjects solve tasks requiring inductive reasoning (cf. Glaser & Pellegrino, 1982; Goldman & Pellegrino, 1984; Greeno, 1978, 1980; Ja-

cobs & Vandeventer, 1972; Sternberg, 1977, 1985). Further, researchers in the field of *artificial intelligence* have constructed computer programs that attempt to solve certain kinds of inductive-reasoning problems in order to test theories about inductive processes (cf. Ernst & Newell, 1969; Holland, Holyoak, Nisbett, & Thagard, 1986; Kotovsky & Simon, 1973).

In contrast to the research mentioned above, the present contribution is based on a different approach, namely a prescriptive one. A prescriptive theory does not describe how subjects actually proceed when solving problems—there is presumably an infinite number of ways to solve inductive problems, depending on the type of problem as well as on different experiential backgrounds and idiosyncrasies of the problem solver. Unlike descriptive theories, a prescriptive theory delineates what to do when a problem has to be solved by describing those steps that are sufficient to solve problems of the type in question. A prescriptive theory of inductive reasoning specifies the processes considered to be sufficient to discover a generalization or to refute an overgeneralization. Obviously, such a theory can be tested in a straightforward manner by a training experiment for transfer. Participants trained to apply an efficient strategy to solve inductive problems should outperform subjects who did not have this training, given that the subjects are not already highly skilled in solving inductive problems. Thus, children would seem to be likely candidates for the training of inductive reasoning strategies.

Defining a Prescriptive Theory of Inductive Reasoning

A primary purpose of this article is to test a prescriptive theory of inductive reasoning by a training experiment. Inductive reasoning enables one to detect regularities and to uncover irregularities. In Fig. 1 it is suggested that this is accomplished by a comparative process, i.e., by a process of finding out similarities and/or differences with respect to attributes of objects or with respect to relationships between objects. Conceptualizing the definition of inductive reasoning this way implies that inducing adequate comparison processes in learners would improve the learners' abilities of inductive reasoning.

Specifically, Fig. 1 makes use of an incomplete form of a mapping sentence as developed by Guttman (Guttman & Levy, 1991). The three facets A, B, and C consist of 3, 2, and 5 elements, respectively. Accordingly, $3 \times 2 \times 5 = 30$ varieties of inductive reasoning tasks are distinguished. For the present study, the core facets A and B are of particular interest. Facet A refers to the comparison process and facet B specifies the aspects (attributes or relationships) which are compared. Logically, attributes are conceived as one-place predicates, whereas relations are identified as predicates with two or more places. For example, $P(x)$ is a one-place predicate. It means that x has the attribute P . $P'(x,y)$ is a two-place predicate. It means that x is linked

Inductive reasoning consists in finding out regularities and irregularities by detecting

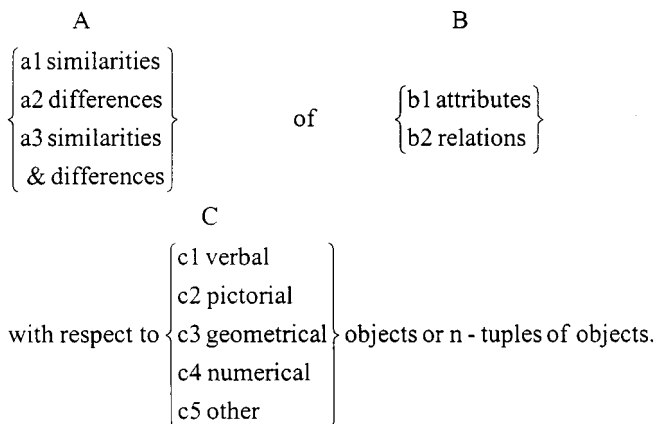


FIG. 1. Definition of inductive reasoning.

with y in the relation P' , for instance, that x is the cause of y . Hence, it is clear that these two kinds of predicates exhaust all possibilities of making statements about objects. This is one reason why inductive processes are fundamentally important and broadly applicable.

Facets A and B constitute six types of inductive reasoning. Table 1 specifies these six in some detail. The table presents the designations given each of the six types of inductive reasoning, moreover the facet identifications, the item formats used in psychological tests, and the cognitive operations required by them.

Figure 2 gives an overview of the genealogy of inductive reasoning tasks for the six types of tasks defined by Facets A and B. The inductive reasoning strategy refers to the comparison process which deals either with comparing attributes of objects (left-hand branch of the genealogy) or with relations between objects (right-hand branch). In any case, one is required to search for similarity, for difference, or both similarity and difference. In this way one detects commonalities and difference. The item classes "cross classification" and "system formation" require one to take notice of both the *same* and a *different* attribute or the *same* and a *different* relationship. That is the reason why these item classes represent the most complex inductive problems—the problem solver must deal with two or more dimensions simultaneously. For more details concerning this theoretical background see Klauer and Phye (1994) or Klauer (1998, 1999). In the Appendix, six sample items of the program are depicted.

The second purpose of this contribution is to determine the convergent/

TABLE 1
Types of Inductive Reasoning Problems

Process	Facet identification	Problem formats	Cognitive operation required
Generalization (GE)	a_1b_1	Class formation Class expansion Finding common attributes	Similarity of attributes
Discrimination (GE)	a_2b_1	Identifying irregularities	Discrimination of attributes (concept differentiation)
Cross-Classification (CC)	a_3b_1	4-fold scheme 6-fold scheme 9-fold scheme	Similarity & difference in attributes
Recognizing Relationships (RR)	a_1b_2	Series completion ordered series analogy	Similarity of relationships
Differentiating Relationships (DR)	a_2b_2	Disturbed series	Differences in relationships
System Construction (SC)	a_3b_2	Matrices	Similarity & difference in relationships

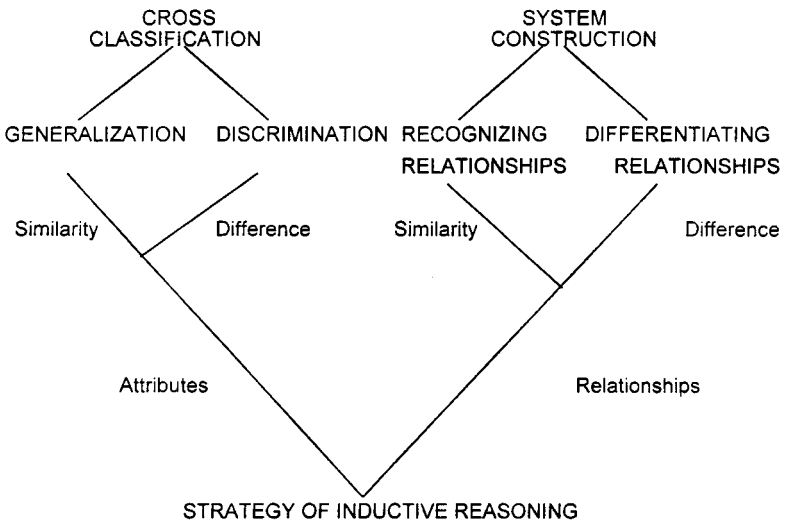


FIG. 2. Genealogy of tasks in inductive reasoning.

divergent validity of the training program based on the prescriptive theory. The question is whether the training program primarily fosters inductive reasoning or whether other cognitive processes are also impacted. The latter should not be the case. If subjects are trained effectively to apply a useful strategy of inductive reasoning, then a domain-specific effect is expected. Theoretically, such training should not transfer to performances that are independent of inductive reasoning. However, if the training improves noninductive variables as well, then one could assume that the training produced—at least additionally—a nonspecific or more general learning-to-learn effect. Such training-induced general effects could be achieved by noncognitive effects of cognitive training such as by (a) the improvement of self-concept and confidence during training, which can lead to improved performance (Hansford & Hattie, 1982; Scheirer & Kraut, 1979); (b) the stimulation of achievement motivation due to training (Hartman & Sternberg, 1993; Sternberg & Gardner, 1983); (c) the special attention students receive while they are trained—a variety of the Hawthorne effect which is sometimes expected to lead to a general improvement in performance (Adair, Sharpe, & Huynh, 1990); or (d) a generalized placebo effect if the control group received no treatment at all instead of a placebo treatment (Lipsey & Wilson, 1993). These effects as well as other nonspecific effects could falsely be attributed to inductive training if not taken into consideration.

The third purpose is an attempt to clarify the question, does inductive training improve participants' inductive competence or only inductive performance? Competencies are not conceived of as behaviors but as unobservable theoretical constructs that are relatively enduring traits that, among other factors, enable one to demonstrate the relevant performance. For instance, one can assume that a real improvement in competence should lead to improved performance that lasts longer than a few days or weeks because warm-up and focusing effects disappear rather soon after training (Dush, Hirt, & Schroeder 1989; Willson & Putnam, 1982).

Hypotheses

Two main hypotheses are tested in this research. The first hypothesis relates to the breadth of the training effect. It is expected that training children to reason inductively does improve inductive reasoning but not performance on tests that do not require this kind of reasoning. Thus, the hypothesis concerns the question, does the training program induce a domain-specific or a more general transfer effect? If only inductive reasoning is fostered by the training, then both convergent and discriminant transfer effect have to be expected. In this case positive transfer should occur with measures of inductive reasoning (convergent transfer) but no transfer should occur with measures that do not require inductive reasoning (discriminant transfer).

Hypothesis 1 concerns the question of convergent and discriminant transfer of the inductive training. Using two different criteria it can be stated in

two different ways. Hypothesis 1.1 states that training to reason inductively should transfer positively to Raven's Coloured Progressive Matrices performance but should not transfer to vocabulary test performance. Hypothesis 1.2 states that there is an analogous difference in transfer effects for subtests of the Culture Fair Test 1. Inductive training should improve performance on subtests requiring inductive reasoning but not subtests that do not require it.

Hypothesis 2 investigates the distinction between performance immediately following training and competence that has a lasting effect. With respect to theoretical and practical considerations, it is of utmost importance that the training and transfer effects should not be confined to only the level of performance. Again, two criteria are used, giving rise to two varieties of the performance–competence hypothesis. The first criterion refers to the stability of a training effect. In training research, it is known that only a few studies were able to show training effects lasting over a longer period of time though this aspect is deemed to be of high importance (see Resnick, 1987a, 1987b). We assume that if an effect is present for some months, this fact can be taken as an indicator of an improvement in competence instead of an improvement only in performance. Whether this criterion is fulfilled can be tested by a delayed posttest. Hence, hypothesis 2.1 states that training to reason inductively improves inductive performance immediately following training and also in a posttest 6 months later.

With hypothesis 2.2, an attempt is made to provide direct evidence that a training of inductive reasoning affects a nonobservable competence instead of open performance only. This objective should be arrived at by employing a structural equations model. Within such a model, a distinction is made between observable *manifest* and not observable *latent* variables (Jöreskog & Sörbom, 1993a). Hence, we would expect that the training affects a latent variable directly and manifest variables only indirectly, i.e., mediated by a latent variable. But which latent variable should be influenced by the training? Since Cattell's CFT and Raven's CPM are classic markers of fluid intelligence and the vocabulary test a classic marker of crystallized intelligence, a LISREL model was constructed such that effects on the posttest latent variables "fluid" and "crystallized intelligence" could be detected. Accordingly, hypothesis 2.2 states that in a structural-equations model the inductive training will improve the latent variable "fluid intelligence" but not the latent variable "crystallized intelligence" as they are defined in this study by their typical markers.

METHOD

Participants

The study was administered in German elementary schools. Two hundred seventy-nine children, mean age 7 years and 1 month, participated. They belonged to 12 first-grade classes in six elementary schools. The headmasters and teachers of the selected classes had to agree to

allow their classes to participate. Within that restriction, two classes were chosen from each school such that there were no significant differences in the mean achievement level of the classes. The two classes were randomly assigned to one of two conditions, the training or the control condition. Thus, the design could be characterized as a quasiexperiment.

Children were pretested about 1 week before and posttested up to 2 weeks after training. Six months after posttest 1 a delayed posttest was administered. At that time only 219 children were available: Meanwhile, a new school year had begun, several children had moved to other places, and others were ill the day when the delayed posttest took place or did not attend school for other reasons on that day. However, attrition did not produce differential groups for the posttest.

Training

The training was based on the program published by Klauer and Phye (1994). Basic cognitive and metacognitive objectives of the program are to teach the children to recognize an inductive problem, to differentiate between the types of problems (not necessarily by labeling them), to apply the adequate solution procedure to the type of problem, and finally to check one's own solution. Particular emphasis was put on teaching for transfer such that the children should become able to apply the cognitive and metacognitive strategies on any inductive problem whenever they met one.

The project started at the beginning of the second half of the school year. In all schools, one class received training and the other class continued with regular classroom activities. The training lasted 5 weeks. In each of the weeks, every child participated in 2 training sessions of maximally 45 min so that 10 training lessons were administered to each child. The training sessions took place in small groups of about three to five children. It was administered in a separate room so that the children had to leave their classroom in order to participate in the training. Trainers were two female psychologists who were very experienced in the application of the training program.

Transfer Assessment

Three tests were administered in order to assess transfer. Two of the tests were used as measures of fluid intelligence, the German versions of Cattell's Culture Fair Test CFT 1 (Weiss & Osterland, 1980) and Raven's Coloured Progressive Matrices (Schmidtke, Schaller, & Becker, 1980). The third test was a vocabulary test (VT) to measure crystallized intelligence. Because the children were not yet able to read fluently enough, a vocabulary test was chosen which required the participants to identify a spoken word by marking the correct choice of four familiar pictures (Kamratowski & Kamratowski, 1969). The children had 20 s/item. The test consisted of 35 items.

The German version of CFT 1 is composed of five subtests: substitutions, labyrinths, classifications, similarities, and matrices. All of them are loaded with factor g_f (Weiss & Osterland, 1980), but only the last three subtests clearly consist of inductive items in terms of Klauer's theory sketched above.

All transfer tests were administered in small groups by psychologists who did not know which children belonged to which group nor which hypotheses should be tested. Raven's CPM and Cattell's CFT were administered two times: the first time some days before training, the second time up to 2 weeks after training. Only the CFT was administered a third time 6 months later as a delayed posttest.

Looking at pretest results, it seemed possible that at least in some cases a ceiling effect could emerge, particularly if the training would be effective. For that reason the first eight items of Raven's Standard Progressive Matrices (SPM) were added to the CPM posttest. Nev-

ertheless, a few children received the maximum possible score with the Raven posttest. No ceiling effect was expected nor found with the CFT subtests.

RESULTS

Hypotheses Concerning Domain-Specific and Discriminant Transfer

According to hypothesis 1 it was expected that inductive training would improve problem-solving transfer performance with inductive tests but not performance with noninductive tests. Hence, with hypothesis 1.1 it is expected that—due to the training—the training group outperforms the control group with respect to the CPM but not with respect to the vocabulary test, VT. Descriptive statistics of the two groups and tests are given in Table 2.

Inspecting pretest differences, one can see that the experimental and control groups differ somewhat from each other. The control group yielded slightly higher scores with both tests. The small differences nevertheless were statistically significant ($p < .05$) because there were rather large samples.

With all analyses, participants were chosen as the unit of analysis. The appropriate unit would have been the groups if the control classes also would have been treated in small groups. However, this was not the case because they continued their regular classroom activities. Another possibility would have been to choose the classes as the unit of analysis but the training was not administered to whole classes. Hence, it seemed to be most adequate to use the subjects as the unit of analyses.

Analyses of covariance were performed using the posttest values as the dependent variable, the pretest values as a covariate, and group as the independent variable. In both cases, the pretest contributed significantly to the posttest variance. As was expected, the group factor contributed significantly to the Raven test CPM, $F(1, 276) = 100.30$, $p < .001$, but not to the vocabulary test VT, $F(1, 276) = 1.18$, $p = .28$.

In order to also consider the size of the effects, effect size measures were

TABLE 2

Means and Standard Deviations of the Two Groups and the Two Test Administrations with Raven's Coloured Progressive Matrices (CPM) and the Vocabulary Test (VT)

Test		Training group		Control group	
		Pretest ($N = 139$)	Posttest ($N = 111$)	Pretest ($N = 140$)	Posttest ($N = 108$)
CPM	<i>M</i>	23.2	33.1	25.1	29.8
	<i>SD</i>	5.46	6.22	5.08	6.10
VT	<i>M</i>	24.7	26.1	25.7	26.4
	<i>SD</i>	4.86	4.45	4.15	4.03

TABLE 3

Means and Standard Deviations of the Raw Scores of the Inductive and the Noninductive Subtests of Cattell's CFT 1 with the Two Groups and the Two Test Administrations

Subtests	Training group		Control group	
	Pretest (<i>N</i> = 139)	Posttest (<i>N</i> = 111)	Pretest (<i>N</i> = 140)	Posttest (<i>N</i> = 108)
Inductive				
<i>M</i>	21.73	35.64	24.16	31.74
<i>SD</i>	6.24	6.91	5.96	7.74
Noninductive				
<i>M</i>	17.60	21.99	20.16	21.36
<i>SD</i>	4.74	2.65	3.61	2.43

calculated as well. Because of the pretest differences between the groups, effect size d corrected for pretest differences and was calculated as $d_{\text{corr}} = d_{\text{posttest}} - d_{\text{pretest}}$ with $d = (M_{\text{TG}} - M_{\text{CG}})/s_p$ where s_p is the pooled standard deviation. For the Raven test CPM, a $d_{\text{corr}} = 0.90$ was obtained and for the vocabulary test VT a $d_{\text{corr}} = 0.15$. One can conclude that hypothesis 1.1 was corroborated.

With hypothesis 1.2 a comparable difference is expected concerning the inductive and the noninductive subtests of Cattell's CFT 1. In Table 3, the means and standard deviations of the sums of the two kinds of subtests are presented for both groups and both test administrations. As one can see, there were considerable gains for the training group and definitely smaller gains for the control group. Analyses of covariances were calculated analogously to those above. For the inductive subtests, the trained group significantly outperformed the control group, $F(1, 276) = 76.40, p < .001$. Unexpectedly, a significant difference was also observed for the noninductive subtests, $F(1, 276) = 24.42, p < .001$. The effect sizes tell a similar story: With the inductive subtests a $d_{\text{corr}} = 0.93$ was yielded and with the noninductive subtests a $d_{\text{corr}} = 0.83$. The difference is negligible.

Hypothesis 1 deals with the question of discriminant transfer. Hypothesis 1.1 was confirmed but hypothesis 1.2 was not. It seems that the transfer effect of the training is limited but not as narrowly as was expected. However, with the next hypothesis this question can be reconsidered because this hypothesis will also be tested using the Cattell test.

Hypotheses 2 concerns the question of whether the training affects competence or only performance. Its effect on performance has just been shown but it is unclear whether the training actually increased the children's competencies. Two hypotheses have been formulated in order to shed light on this question. Because hypothesis 2.2 needs a more detailed presentation, the two hypotheses are dealt with in two different sections.

TABLE 4

Means (Standard Deviations) of the Raw Scores of the Training and Control Group with Inductive and Noninductive Subtests of Cattell's CFT 1 at Three Test Administrations ($N_{TG} = 111, N_{CG} = 108$)

	Pretest	Posttest 1	Posttest 2 (6 months later)
Inductive subtests			
TG	21.58 (6.00)	35.24 (6.62)	36.17 (6.74)
CG	23.80 (6.23)	32.06 (7.53)	31.40 (7.12)
Noninductive subtests			
TG	17.48 (4.89)	22.15 (2.66)	22.72 (2.01)
CG	20.27 (3.77)	21.32 (2.47)	22.49 (2.37)

Hypothesis Concerning Stability of Transfer Effect

With hypothesis 2.1 we expect that a change in competence will lead to a rather stable modification in performance. For that reason, the CFT was administered a third time 6 months after the second administration. Table 4 contains the data for those children who participated in all three test administrations. Again, in Table 4 a distinction is made between inductive and noninductive subtests.

An inspection of Table 4 reveals that there is—possibly—a differential development for the two groups with respect to the two kinds of subtests. Figures 3 and 4 present a better opportunity to compare the development of

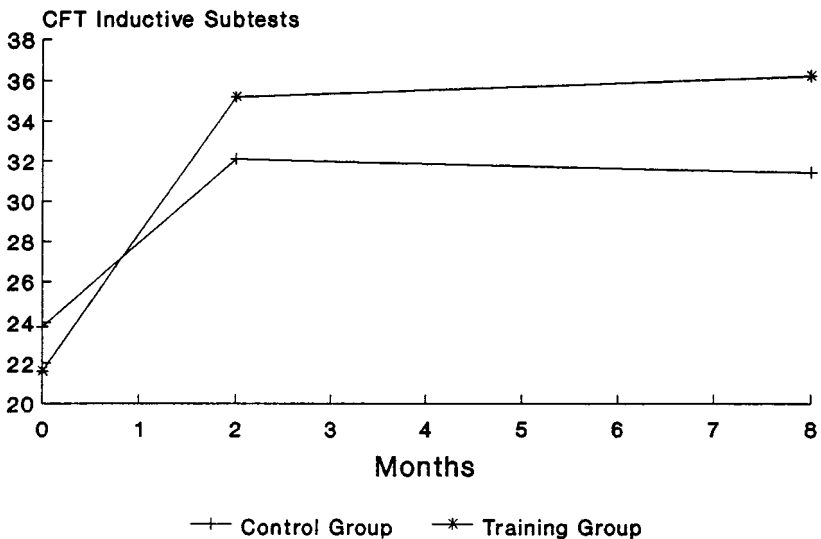


FIG. 3. Development of CFT performance of the two groups: Inductive subtests.

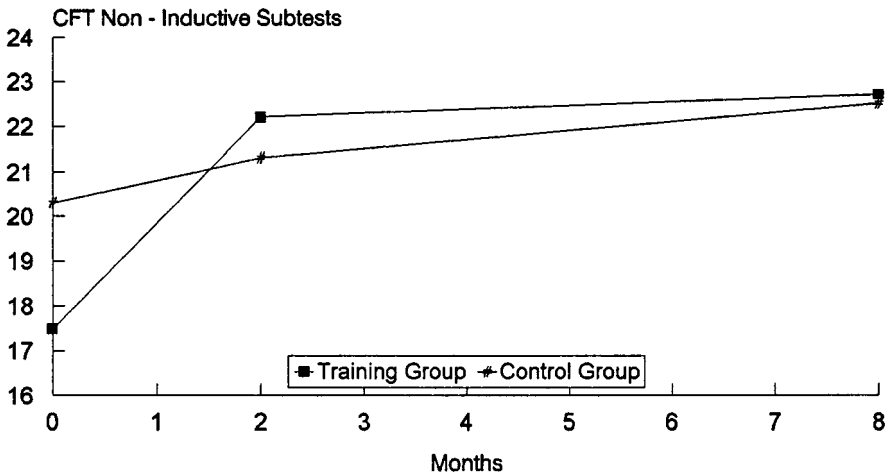


FIG. 4. Development of CFT performance of the two groups: Noninductive subtests.

the groups. With the inductive subtests, both groups improved considerably immediately after training although the training group improved more than the control group, as was to be expected. The control group could just keep its acquired level later on, whereas the training group—possibly—even improved a bit.

For noninductive subtests, however, the picture is different, due to a different developmental pattern for the control group. The control group shows a steady development over the 8-month period, revealing a rather small increase which probably reflects little improvement due to test repetition and regular developmental progress. Although the training group gained more during training, this group maintains its gain over time. Consequently, half a year later there was no longer any difference between groups.

It seems that the training group had a spurious acceleration with noninductive reasoning, whereas the control group had a spurious acceleration with inductive reasoning. Six months after the second administration, there was no difference between both groups with the noninductive subtests but a clear difference with the inductive subtests favoring the training group. The results at the third test administration correspond to the predictions of hypothesis 1.2. The question remains, Is this differential development statistically significant?

This question can be answered by examining the gains between pretest and posttest 2, 6 months later with respect to the two groups of subtests. An analysis of covariance with repeated measures was calculated using the inductive and noninductive subtests of posttest 2 as dependent variables, the

same pretest variables as covariates, group as a between-subjects factor, and test as a within-subject factor. Because of unequal variances, all of the variables were transformed into z scores. The group factor turned out to be significant, $F(1, 216) = 57.57, p < .001$, indicating greater gains of the training group, and—due to the z transformation—there was practically no difference between the subtests ($p = .968$). More interesting, the critical group \times test interaction also was significant, $F(1, 216) = 7.76, p < .01$. This result confirms the impression of a differential transfer between the two groups and the two kinds of tests. Taken together, the inductively trained group gained more than the control group with both kinds of tests during the 8 months. But as far as mean gains are concerned, the training group outperformed the control group on the noninductive subtests by 3.03 raw scores and on the inductive subtests by 6.97 raw scores. In terms of effect sizes corrected for pretest differences, the training effect was estimated to be as large as $d_{\text{corr}} = 0.74$ for the noninductive subtests and $d_{\text{corr}} = 1.05$ for the inductive subtests.

As far as hypothesis 2.1 is concerned, there was no indication of a decline in performance of the training group 6 months after training. However, in light of these findings the conclusion concerning hypothesis 1.2 must be qualified. Although it was confirmed that inductive training also transferred to noninductive subtests of the CFT, the hypothesis of a discriminant transfer effect of the training was supported. Transfer to the noninductive subtests was rather large but remained significantly smaller than transfer to the inductive subtests.

Cross-validation of the stability of transfer effects was conducted in the following manner. Two classes from a school with 44 children originally trained were available for retesting with the CPM a fourth time exactly 2 years after the beginning of the experiment. At this retesting 34 children from the original training participated in the follow-up retesting. Data from this school exhibited the following profile. Immediately after training, i.e., about 2 months after the start of the project, the training group outperformed the control group by a corrected effect size of $d_{\text{corr}} = 1.10$. Both groups improved their mean performances continuously during the following 22 months. However, even at 2 years the difference between training and control group still was statistically significant with an effect size of $d_{\text{corr}} = 0.97$. It seems that the effect sizes found in this school maintained a value of approximately 1 standard deviation without a marked loss over the 2 years.

Since improvement of inductive reasoning was stable over so long a time, it is reasonable to assume that the training fostered not only *performance* of inductive processes but also the *ability* (competency) to induce regularities. However, because of the close relationship between inductive reasoning and fluid intelligence, it seems possible that the training also enhanced fluid intelligence.

Hypothesis Concerning the Performance—Competence Distinction

If the inductive training has an impact on childrens' abilities, then this effect should be detectable immediately after training. It follows that only when the training was operative on the competence level would one expect the effects to last over a longer period of time. Since longer lasting effects of the training have been established, it seems a more rigorous test would be to use the data from the test administration immediately after training, when testing to determine if the training improved ability as well as performance. Consequently, hypothesis 2.2 was tested using the data of the two-wave (prepost) longitudinal study instead of the three-wave (follow-up) study. Specifying hypothesis 2.2, it is expected that the training affects fluid but not crystallized intelligence.

Because vocabulary tests of the kind being used are known as indicator variables of crystallized intelligence and because the inductive variables are known as indicator variables of fluid intelligence, it is possible to set up a structural equation model with the latent variables crystallized and fluid intelligence. A structural equation modeling approach was developed to demonstrate the effect of inductive training at the level of latent variables (competency). Furthermore, since hypothesis 1 concerning a discriminant transfer effect has been demonstrated, it seemed useful to also take this into account by inquiring whether the training exerted a differential influence on different latent variables.

Some technical details need to be reported. The LISREL8 program package (Jöreskog & Sörbom, 1993a) was used for analysis in conjunction with the PRELIS2 program (Jöreskog & Sörbom, 1993b) the latter being employed for initial raw data processing and computation of the covariance matrix. For the actual analyses, LISREL submodel 3 was applied. An error-free observed variable "training" was introduced by assigning "0" to the control children and "1" to the trained children. Then, a corresponding latent variable "training" was introduced by fixing the path between latent and observed variable at 1 and defining the error variance of the indicator variable to be zero.

In order to have a proper measurement model with at least two observed variables for both latent variables "fluid intelligence" and "crystallized intelligence," the items of the vocabulary test were randomly split into two halves and the total scores for both test halves were entered into the analysis. Correspondingly, the items of the Coloured Progressive Matrices CPM also were randomly split into two halves so that two variables were available for the measurement model of "fluid intelligence."

Because the scales of measurement for the latent variables should be fixed, the path to one observed verbal ability was set to 1 for each latent variable. Furthermore, assuming tau-equivalent measurements, both paths to the two

CPM parts were also set to 1. Since the same tests were used in pretest and posttest, the corresponding paths from latent to observed variables were constrained to be equal for pre- and posttest. The respective error covariance parameters were allowed to differ from zero because of possible measurement errors due to repeated measurements. The error covariance between both latent variables at the posttest also was allowed to differ from zero.

Since raw scores were chosen for analysis, path coefficients and error variances of the LISREL8 solution cannot be compared directly. For a numerical comparison of the interesting path coefficients a so-called completely standardized solution was computed as well in which the variances of all observed and latent variables were set to a value of 1.

The major result is depicted in Fig. 5. Latent variables are given in elliptical fields, observed variables in rectangular fields. The nonstandardized path coefficients are included in Fig. 5 as well as the error variances of the observed variables. Significant path coefficients are marked by an asterisk.

Before going into details, it seems necessary to have a look at the model fit. The overall fit of the model is certainly acceptable [$\chi^2(17) = 12.49$, $p = .770$] with a small standard root-mean-square residual (standardized RMR = 0.0213). The goodness-of-fit index as well as the adjusted goodness-of-fit index are close to 1, GFI = 0.990 and AGFI = 0.974. In addition to these overall measures, the squared multiple correlations for the latent variables at posttest are high: $R^2 = 0.816$ for crystallized and $R^2 = 0.711$ for

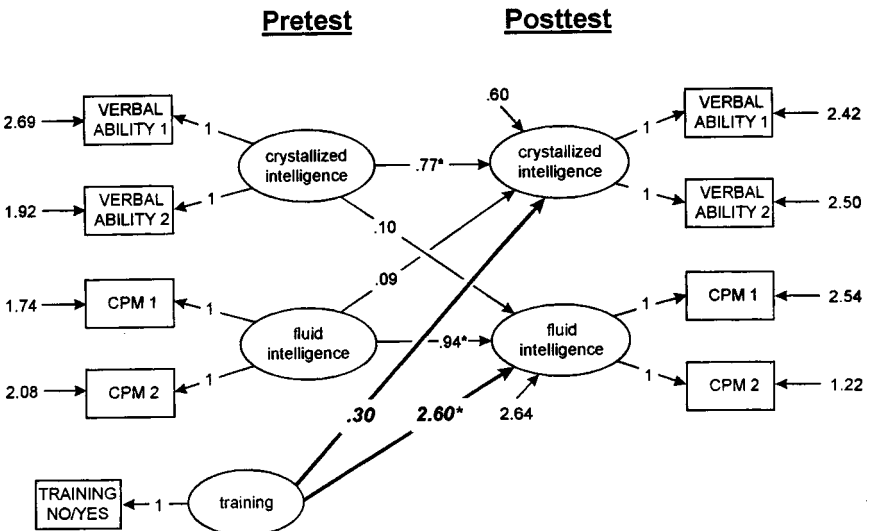


FIG. 5. Structural equations model with nonstandardized path coefficients (significant path coefficients marked by an asterisk).

fluid intelligence. Error covariances between tests and the corresponding retests were somewhat larger than zero for three of the four measures (except for the first random split part of the vocabulary test), which indicates some retest effects. However, there was no significant error covariance between fluid and crystallized intelligence at posttest although the correlation between both variables was 0.48 at posttest and 0.58 at pretest.

More interesting are the path coefficients themselves. If the training affected fluid intelligence and only fluid intelligence, then the path from "training" to "fluid intelligence" in Fig. 5 should be large and the path from "training" to "crystallized intelligence" should be zero. Significance of departure from zero is assessed by dividing the coefficient by its standard error and checking whether the resulting (absolute) t value is larger than 2. For the large path coefficient to "fluid intelligence" one gets $t = 10.70$ and for the path to "crystallized intelligence" one has $t = 1.87$, which is slightly smaller than the cutoff proposed by Jöreskog and Sörbom. Comparing the two path coefficients of the completely standardized solution (0.431 vs 0.090), the coefficient for "fluid intelligence" is 4.79 times larger than the one for "crystallized intelligence." This result points to a strong effect of the inductive training on "fluid intelligence" and a negligible effect on "crystallized intelligence."

As the corresponding path coefficients show, posttest "fluid intelligence" is more or less exclusively influenced by pretest "fluid intelligence" and "training" but not by pretest "crystallized intelligence." The latter almost exclusively influences posttest "crystallized intelligence."

In order to check if the solution can be replicated, two additional, separate analyses were carried out using the data from the first posttest. One analysis was performed for randomly split halves of the CFT and one for both CPM and CFT as indicator variables for fluid intelligence. Both analyses corroborated the findings presented in Fig. 3. They are almost identical in all relevant aspects as is shown in Table 5.

What do these results mean for hypothesis 2.2? This hypothesis stated that the training has an effect on the latent variable "fluid intelligence" but no effect on the latent variable "crystallized intelligence." One can conclude that both parts of the hypothesis were corroborated: The training has had an effect on posttest latent variable "fluid intelligence" but nearly no effect on posttest "crystallized intelligence."

DISCUSSION

Methodological Aspects

Are there reasons to assume that the present findings could have emerged from nonintended factors due to methodological issues that were not taken into consideration? This question could be raised because the control group

TABLE 5

Summary Results of Two Additional Analyses (1) for Two Halves of the CFT and (2) for the CFT and CPM as Indicator Variables of "Fluid Intelligence".

Type of analysis	Latent variables/ parameters	(1) CFT	(2) CFT and CPM
Squared multiple correlation	Crystallized at posttest	$R^2 = 0.811$	$R^2 = 0.819$
	Fluid at posttest	$R^2 = 0.616$	$R^2 = 0.897$
Raw path coefficients	Training \rightarrow cryst. post	0.39 ($t = 1.89$)	0.44 ($t = 2.18$)
	Training \rightarrow fluid post	4.23 ($t = 11.22$)	8.93 ($t = 11.83$)
	Cryst pre \rightarrow cryst post	0.80 ($t = 12.81$)	0.75 ($t = 10.07$)
	Fluid pre \rightarrow fluid post	0.72 ($t = 13.33$)	1.18 ($t = 11.98$)
	Cryst pre \rightarrow fluid post	0.20 ($t = 2.05$)	-0.20 ($t < 1$)
	Fluid pre \rightarrow cryst post	0.03 ($t < 1$)	0.05 ($t < 1$)
Correlation	Fluid pre - cryst pre	0.42	0.58
	Fluid post - cryst post	0.39	0.48
Error covariance	Fluid post - cryst post	0.20 ($t < 1$)	-0.32 ($t < 1$)
Overall fit ($n = 279$)	χ^2 (17 resp. 16)	16.95 ($p = .458$)	18.643 ($p = .288$)
	GFI	0.986	0.986
	AGFI	0.964	0.959
	Standard RMR	0.0255	0.0297
Standardized coefficient	Training \rightarrow cryst post	0.094	0.108
	Training \rightarrow fluid post	0.506	0.578
	Ratio: fluid/cryst	5.38	5.35
	Cryst pre \rightarrow cryst post	0.882	0.83
	Fluid pre \rightarrow fluid post	0.711	0.97
	Cryst pre \rightarrow fluid post	0.112	-0.058
	Fluid pre \rightarrow cryst post	0.060	0.143

yielded relatively high gains with the inductive subtests of the CFT and the training group had relatively high gains on the noninductive subtests (cf. Figs. 3 and 4). Both results were clearly unexpected.

One possible explanation for the unanticipated improvements of the control group could be that this group also took advantage of the training by making informal contacts or by unexpected interventions by their teachers. However, such an explanation can definitely be excluded. The two trainers did not inform anyone about the strategy to be taught nor did teachers have any opportunity to get hold of the training material. Moreover, it is also unlikely that contacts between the children of the two groups could have entailed the unexpected improvement of the control group.

The unanticipated results can possibly be explained by test repetition, particularly since we have to account for unexpected improvements in both cases. It is well known that the effect of retesting is different for different psychological domains as well as for different types of performances. Cognitive tests, particularly if an adequate solving strategy can be acquired with the first test administration, are susceptible to considerable retest gains. However, vocabulary tests are seldom influenced by test repetition (LeGagnoux, Michael, Hocevar, & Maxwell, 1990; Wing, 1980; Willson & Putnam, 1982). The unexpected high gains of the control group with the inductive subtests of the CFT as well as the unanticipated gains of the trained group with the noninductive subtests of the CFT could be explained this way.

Regression effects also have to be taken seriously because the training group scored a bit below the control group with every test administered before the training. Although the pretest differences between groups were small, it could be that regression to the mean spuriously enhanced the training group's gain. However, as Figs. 1 and 2 show, there is no tendency at all for the training group to go down with the third assessment. Quite to the contrary, there is a tendency to slightly improve even after a delay of 6 months. Therefore it can be assumed that regression effects did not play a decisive role.

Convergent and Discriminant Transfer Effects of Training

In the introduction it was mentioned that nonspecific factors also can lead to improvements. To the extent that such influences occur, specific effects of training may be overestimated. That is the reason why it makes sense to test if specific training leads only to specific effects. As the results indicate, a general effect was not observed. According to the LISREL analysis, the inductive training had a strong positive impact on the latent variable "fluid intelligence" but little impact on its counterpart "crystallized intelligence." Moreover, the training did improve performance with Raven's Matrices CPM but not vocabulary performance as was predicted by hypothesis 1.1. However, contrary to hypothesis 1.2 the noninductive subtests of the CFT

also improved as a result of training. Because the effect on the noninductive subtests was significantly smaller than that on the inductive subtests, the hypothesis of a discriminant transfer effect need not be completely rejected, but modified.

Belmont, Butterfield, and Ferretti (1982) examined several cognitive training programs and came to the conclusion that effective programs are those which train metacognitive components in addition to specific content. With our training program, important cognitive and metacognitive components of problem solving are taught concomitantly. For instance, metacognitive components would include analyzing the kind of problem to be solved, looking for the optimal solving strategy, monitoring its application, and checking one's own solution. If, for instance, an analytical and reflexive style of thinking and the metacognitive components just mentioned are trained in combination with the inductive strategy, then one would expect some broader cognitive effects of the training. This consideration of metacognitive transfer might explain why there was an effect on the noninductive subtests of Cattell's CFT. However, this effect was significantly *smaller* than the effect on the inductive subtests.

The Performance–Competence Dichotomy

The present experiment revealed a marked stability for the training effects. With our training program, Hamers and de Koning (1998) using the Raven's CPM found the effects to be stable over 3½ months, whereas Hager and Hasselhorn (1993) found results concerning Cattell's CFT similar to the ones presented here and which were stable over 5 months. Moreover, summarizing 10 training experiments with the training program used in this project, a meta-analysis yielded effect sizes of $d = 0.74$ ($p < .001$) immediately after training and $d = 0.80$ ($p < .001$) 4 to 9 months later (Klauer, 1998). For that reason it can be concluded that the training program leads to rather long-lasting effects.

One potential explanation of the relative stable effects is the assumption that training of a cognitive strategy leads to rather enduring effects similar to that of certain psychomotor skills (e.g., riding a bicycle). However, it seems unrealistic to think of an acquired cognitive skill that would lie dormant for six months and not be activated until the posttest was suddenly administered. Though such an explanation cannot be ruled out, it seems to be more realistic to assume that the children would put the inductive strategy into use during the time between the test administration. Another possible explanation of the relative stable effects is given by the assumption that the training schedule generally leads to over-learning and problem-solving transfer. It also seems realistic to assume that children will put the acquired inductive strategies into use during the time between test administrations when they detected a similar problem in their classwork. Such a deliberate sponta-

neous activation of the strategy between training and delayed posttesting could explain the stability of the improvement.

The second indicator of an improvement in competence which was used in this study was the LISREL analysis. It was assumed that training exerts an influence on a *latent* variable instead of a *manifest* variable and in particular on the latent variable "fluid intelligence" as contrasted to "crystallized intelligence." This assumption was confirmed. The LISREL analysis provides a direct test of the second hypothesis. Because this analysis led to the same conclusion as the analysis of the duration effect, the assumption that inductive training does indeed improve intellectual competence as well as intellectual performance appears to be warranted.

Inductive Reasoning and Intelligence

The strategy of inductive reasoning is called for by many intelligence tests. Moreover, it has often been shown that tests requiring inductive reasoning belong to those tests that define fluid intelligence. On the other hand it is clear that noninductive tests also load on the fluidity factor. One example might be tests of *deductive* reasoning (cf. Sternberg, 1985). For this reason, the extent to which inductive reasoning training really fostered fluid intelligence can be questioned. Regardless it seems clear that training at least fostered processes and competencies that are subsumed under the concept of fluid intelligence. An important implication is that training students to reason inductively should also improve learning in academic settings. In fact, a large-scale study by Csapó (1997) has confirmed that inductive reasoning correlates substantially with school achievement.

Test of the Prescriptive Theory of Inductive Reasoning

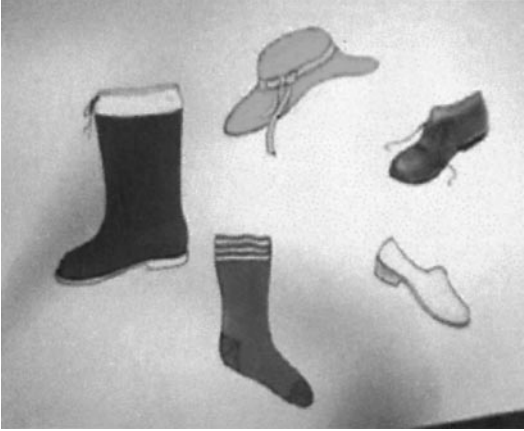
Klauer's prescriptive theory of inductive reasoning sufficiently describes the processes sufficient to solve inductive problems and also defines the range of inductive problems. According to Klauer's definition, there are exactly six different types of inductive problems requiring six varieties of inductive reasoning strategies which can be understood as paradigms of inductive processes. Based on the present findings, it seems possible to conclude that this theory is functionally valid. Hence, teaching to recognize a given problem as belonging to one of these paradigms and teaching how to solve problems of that kind should enable students to solve inductive problems in whichever guise they are encountered.

In summary, empirical evidence has found that a prescriptive theory of inductive reasoning provides us with useful suggestions about how to improve the ability to reason inductively. The available evidence is, however, confined to performance and ability data. Nevertheless, under practical aspects it seems possible to recommend the use of the training program because

the effect on inductive reasoning, fluid intelligence, problem solving, and school-type learning are satisfyingly large, particularly since the training normally requires only 10 lessons.

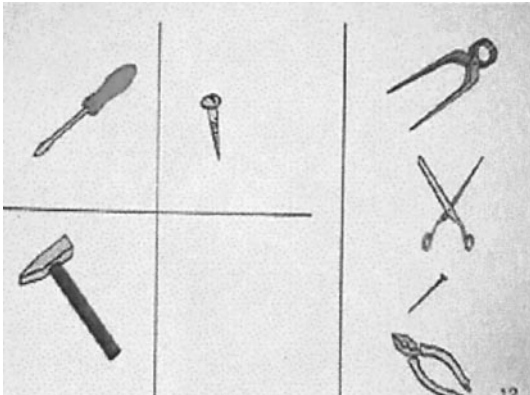
APPENDIX

1. *Generalization of Attributes*



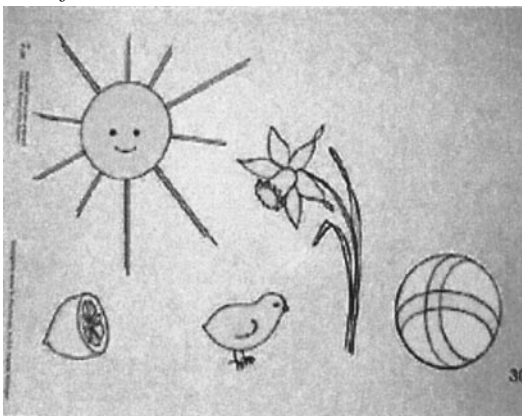
Query: Which three objects belong together?

2. *Recognizing Relationships*



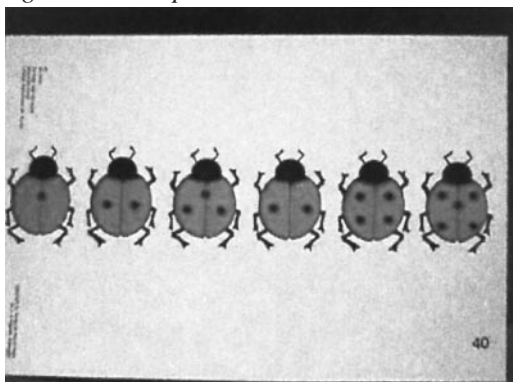
Query: What object belongs in the empty square?

3. *Discrimination of Attributes*



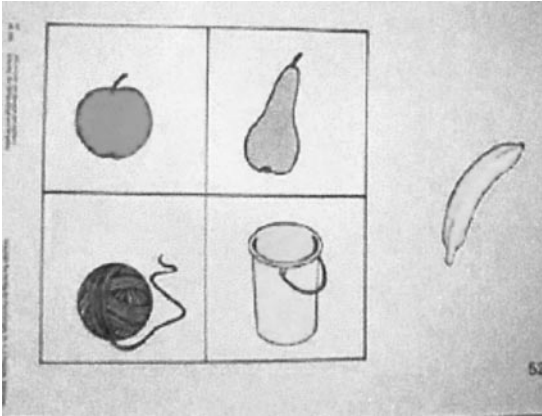
Query: Which picture does not fit in with the other?

4. *Discriminating Relationships*



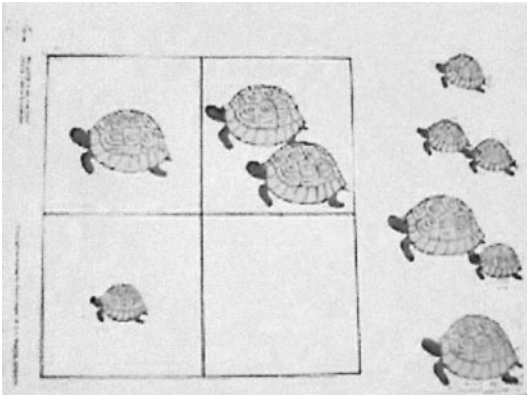
Query: Which bug doesn't fit into the row? Correct the row.

5. Cross-Classification of Attributes



Query: To which of the items on the left does the banana fit?

6. Systems Construction



Query: Which figure on the right best fits in the empty square?

REFERENCES

- Adair, J. G., Sharpe, D., & Huhn, C. L. (1989). Hawthorne control procedures in educational experiments: A reconsideration of their use and effectiveness. *Review of Educational Research*, **59**, 215–228.
- Belmont, J. M., Butterfield, E. C., & Ferretti, R. P. (1982). To secure transfer of training instruct self-management skills. In D. K. Detterman & R. J. Sternberg (Eds.), *How and how much can intelligence be increased* (pp. 147–154). Norwood, NJ: Ablex.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, **54**, 1–22.
- Csapó, B. (1997). The development of inductive reasoning: Cross-sectional assessments in an educational context. *International Journal of Behavioral Development*, **20**, 609–626.

- Dush, D. M., Hirt, M. L., & Schroeder, H. E. (1989). Self-statement modification in the treatment of child behavior disorders: A meta-analysis. *Psychological Bulletin*, **106**, 97–106.
- Ernst, G. W., & Newell, A. (1969). *GPS: A case study in generality and problem-solving*. New York: Academic Press.
- Glaser, R., & Pellegrino, J. (1982). Improving the skills of learning. In D. K. Detterman & R. J. Sternberg (Eds.), *How and how much can intelligence be increased* (pp. 197–212). Norwood, NJ: Ablex.
- Goldman, S. R., & Pellegrino, J. W. (1984). Deductions about induction: Analyses of developmental and individual differences. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 149–197). Hillsdale, NJ: Erlbaum.
- Greeno, J. G. (1978). Natures of problem-solving abilities. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5, pp. 239–270). Hillsdale, NJ: Erlbaum.
- Greeno, J. G. (1980). Some examples of cognitive task analysis with instructional implications. In R. E. Snow, P. Frederico, & E. W. Montagne (Eds.), *Aptitude, learning, and instruction* (Vol. 2, pp. 1–21). Hillsdale, NJ: Erlbaum.
- Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, **8**, 179–203.
- Guttman, L., & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence*, **15**, 79–103.
- Hager, W., & Hasselhorn, M. (1993). Evaluation von Trainingsmaßnahmen am Beispiel von Klauers Denktraining für Kinder [Evaluation of training measures with Klauer's training to reason for children]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, **25**, 307–321.
- Hamers, J. H. M., & de Koning, E. (1998). Inductive reasoning in third grade: Intervention promises and constraints. *Contemporary Educational Psychology*, **23**, 132–148.
- Hansford, R. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, **52**, 123–142.
- Hartman, H., & Sternberg, R. J. (1993). A broad BACEIS for improving thinking. *Instructional Science*, **21**, 401–425.
- Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction. Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.
- Jacobs, P. J., & Vandeventer, M. (1972). Evaluating the teaching of intelligence. *Educational and Psychological Measurement*, **32**, 235–248.
- Jöreskog, K. G., & Sörbom, D. (1993a). *LISREL8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Erlbaum.
- Jöreskog, K. G., & Sörbom, D. (1993b). *PRELIS2 user's guide*. Chicago, IL: Scientific Software International.
- Kamratoski, J., & Kamratoski, J. (1969). *Wortschatztest für Schulanfänger WSS 1*. [Vocabulary test for first graders]. Weinheim: Beltz.
- Klauer, K. J. (1998). Inductive reasoning and fluid intelligence. A training approach. In W. Tomic & J. Kingma (Eds.), *Reflections on the concept of intelligence* (pp. 261–289). Greenwich, CT: JAI Press.
- Klauer, K. J. (1999). Fostering higher order reasoning skills: The case of inductive reasoning. In J. H. M. Hamers, J. E. H. van Luit, & B. Csapó (Eds.), *Teaching and learning thinking skills* (pp. 131–156). Lisse: Swets & Zeitlinger.
- Klauer, K. J., & Pbye, G. (1994). *Cognitive training for children: A developmental program of inductive reasoning and problem solving*. Seattle: Hogrefe & Huber.

- Kotovsky, K., & Simon, H. A. (1973). Empirical test of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology*, **4**, 399–424.
- LeGagnoux, G., Michael, W. B., Hocevar, D., & Maxwell, V. (1990). Retest effects on standardized structure-of-intellect ability measures for a sample of elementary school children. *Educational and Psychological Measurement*, **50**, 475–492.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. *American Psychologist*, **48**, 1181–1209.
- Resnick, L. B. (1987a). *Education and learning to think*. Washington, DC: National Academy Press.
- Resnick, L. B. (1987b). Instruction and the cultivation of thinking. In E. de Corte, H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and instruction* (Vol. 1, pp. 415–442). Oxford; UK: Leuven Univ. Press and Pergamon Press.
- Scheirer, M. A., & Kraut, R. E. (1979). Increasing educational achievement via self concept change. *Review of Educational Research*, **49**, 131–150.
- Schmidtke, A., Schaller, S., & Becker, P. (1980). *CPM Raven-Matrizen-Test*. Weinheim: Beltz.
- Shye, S. (1988). Inductive and deductive reasoning: A structural reanalysis of ability tests. *Journal of Applied Psychology*, **73**, 308–311.
- Spearman, C. (1923). *The nature of intelligence and the principles of cognition*. London: Macmillan.
- Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human ability*. Hillsdale, NJ: Erlbaum.
- Sternberg, R. J. (1985). *Beyond IQ. A triarchic theory of human intelligence*. Cambridge, UK: Cambridge Univ. Press.
- Sternberg, R. J., & Gardner, M. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General*, **112**, 80–116.
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, No. 1.
- Undheim, J.-O., & Gustafsson, J.-E. (1987). The hierarchical organization of cognitive abilities: Restoring general intelligence through use of linear structural relations (LISREL). *Multivariate Behavioral Research*, **22**, 149–171.
- Van de Vijver, F. J. R. (1991). *Inductive thinking across cultures: An empirical investigation*. Tilburg: Catholic Univ. of Brabant.
- Weiss, R., & Osterland, J. (1980). *Grundintelligenztest CFT 1 Skala 1* [Test of basic intelligence CFT 1, Scale 1]. Braunschweig: Westermann.
- Willson, V. L., & Putnam, R. R. (1982). A meta-analysis of pretest sensitization effects in experimental design. *American Educational Research Journal*, **19**, 249–258.
- Wing, H. (1980). Practice effects with traditional mental test items. *Applied Psychological Measurement*, **4**, 141–155.