Synthesis of Research on the Effectiveness of Intellectual Skills Programs: Snake-Oil Remedies or Miracle Cures?

When properly administered, some programs appear to develop thinking skills, but success depends on many implementation-specific variables.

These young girls are participants in the Odyssey project, in which Venezuelan teachers taught thinking skills to 450 7th grade students. The course, originally developed and evaluated by researchers at Bolt, Beranek, and Newman and Harvard University, was shown to result in substantial increases in scores on standardized aptitude and achievement tests.
Suppose you attend a conference on modern developments in health-related research at which a pharmacological salesman discusses a new drug his company manufactures that he claims greatly improves one's general health. The salesman is persuasive but because he seems to be selling a drug at a research conference, and because he works for the company peddling it, you decide to do a little research before buying the drug, which is quite expensive and needs to be used over a fairly long period of time.

The results of your research are disconcerting. There are few studies of the drug's effects, and most of them have been sponsored, supervised, or done "in consultation with" the manufacturer. The reports of the studies are sketchy, and many have inadequate control groups or none at all. Some amount to little more than testimonials about the drug's effects; others use outcome measures that seem to have been selected to maximize the favorability of the reports. Few of the studies have been published in refereed journals; many are reported in a company-sponsored magazine. Those studies that seem better controlled show mixed results. You are skeptical but perplexed. The drug may indeed do everything it is supposed to do, but it's hard to tell from the evidence.

General Observations
Potential consumers of programs to teach thinking skills are in a similar predicament. We conducted a fairly extensive review of the research on five well-known, diverse, and, in most cases, widely used thinking skills training programs: Instrumental Enrichment, Philosophy for Children, SOI, Problem Solving and Comprehension: A Short Course in Analytical Reasoning, and Odyssey. We chose these programs because of their diversity, their reputation and visibility in the thinking skills training arena, and their having been subject to some evaluation. Our goal was to uncover whatever evidence existed to advocate or oppose the use of such programs. Our findings paralleled those for the hypothetical drug.

1. Many of the studies were conducted or sponsored by the program developers, others involved fairly extensive consultation with them.

2. Reporting was usually sketchy and often wholly inadequate. In a majority of instances, detail was insufficient for anything resembling a careful replication of what had been done.

3. Most studies involved inadequate (untrained) control groups, and some entailed none at all. Only rarely was the efficacy of one program compared with that of another program.

4. Some of the evidence offered in support of the various programs amounted to little more than user testimonials. The means of selection of such users was usually unspecified.

5. Outcome measures often overlapped program content, and thus tended to favor the program being tested. Inadequate attention was usually given both to transfer of training and to durability of training over the long term.

6. A very small proportion of the studies was published in refereed journals. Such publication helps ensure independent scrutiny of methods and results.

7. Many studies were unpublished, published in media sponsored (and hence controlled) by the program developers, or available only through the program developers. Availability through such limited channels, of course, raises questions as to just how unbiased the sampling of available studies truly is.

8. Even given these caveats on reporting, the results were generally mixed, some indicating significant gains and others not.

Many other findings emerged from our review of the literature. For example, we were unable to find any evaluations that were useful in determining which aspects of the programs worked and which did not. Such information would be useful not only in further program development, but in considering implementation of only those aspects of a given program that have been shown to work. None of the evaluations adequately assessed motivational effects (i.e., improvements in scores due to greater motivation to do well on intellectual skills tests or to impress examiners). Problems of subject dropout, confounded variables, and sensitivity of the program to teacher quality were rarely discussed. Issues of which populations benefited more and which benefited less from the instruction were generally ignored. Our findings were disconcerting, to say the least.

In our communication with program developers, a few odd experiences reinforced our concerns. One program developer sent us certain materials with the understanding that we would not quote or cite them. In another instance, a program developer's assistant said that results would be presented at a forthcoming conference, but she was unwilling to share any data in advance.

Some of our experiences with program developers were quite positive. One team of developers could not have been more cooperative; they sent us what was one of only a handful of scrupulously designed, carefully conducted, and fully reported studies of an existing program. Another developer readily admitted that the evidence was spotty, but was willing to provide any available information.

We do not wish to imply that the results of our review left us with nothing but skepticism. Several program developers admitted that the existing...
evaluations are inadequate. At least one was frustrated with the difficulty he has experienced in obtaining funding adequate to test his program. Indeed, such funding is difficult to obtain. Some of the program developers had obviously tried to come by adequate evaluations but, confronting limited resources, had to use what was available, which admittedly was not much.

Certainly, anyone who is seriously interested in evaluating the effects of a program should call in a trained program evaluator who has no vested interest in the results of the assessment. Given the constraints of the typical school environments in which evaluations are conducted, even the best of intentions can go awry when school schedules, resources, and administrative support (or the lack thereof) must be reckoned with. Clearly, districts need to conduct their own evaluations, as the way in which an evaluation will proceed depends heavily on the student population, how the course is taught, intended course outcomes, and the extent to which the program is taught as either an infused instructional unit or set of units (i.e., as part of the regular curriculum) or as a separate course.

Although the available evidence is so scant as to preclude any firm decisions, we were able to reach what we believe are informed conclusions about the five programs we surveyed.

Instrumental Enrichment

Feuerstein's (1980) Instrumental Enrichment (IE) is an intervention program based on classroom and individual paper-and-pencil exercises presented for three to five hours per week as a separate curriculum. Completion of the entire set of instruments can easily take two years of instruction. Underlying the program is the belief that the human cognitive system is open and modifiable both through direct exposure to information and, especially, exposure to information that is mediated by an agent who interprets the information for the student. Feuerstein refers to the latter type of exposure as “mediated learning experience.”

Mediated learning experience has associated with it five primary goals: (1) correction of deficient intellectual functions, (2) provision of concepts necessary for progress on the 14 “instruments” (types of exercises) of the IE program, (3) development of intrinsic as well as socially reinforced motivation, (4) production of insight, and (5) achievement of change in the learner's perception from that of being a passive organism to that of being an active one.

The 14 instruments include (1-2) orientation in space I-II (which involves two instruments), (3) comparisons, (4) categorization, (5) analytic perception, (6) family relations, (7) temporal relations, (8) numerical progressions, (9) instructions, (10) illustrations, (11) representational stencil design (spatial relations through recognition of embedded figures), (12) transitive relations, (13) syllogisms, and (14) orientation in space III.

The instruments tend purposely to be very abstract; some might say they are IQ test-like. There is a great deal of paper-and-pencil activity, much of it nonverbal. The problems are challenging, even for teachers. A key element of instruction is “bridging,” which involves relating the intellectual skills taught in the program to real-world problem solving. However, the form and extent of the bridging is pretty much left to the instructor, allowing considerable variability in both its quantity and quality.

We reviewed 38 studies of Instrumental Enrichment, which were conducted by a wide variety of investigators with diverse subject populations with respect to country, social class, age, intellectual level, and educational level. The dosage of the program has been quite variable across studies, with greater amounts of exposure generally leading to greater gains. The number of instruments actually used in any particular implementation has also been variable. Assessments have generally been concerned with academic types of gains. Very little information has been available on the constructs of insight and motivation, both of which the program is intended to

In Making Up Our Minds, a two-year case study of the use of Instrumental Enrichment in five British Local Educational Authorities sponsored by the British Schools Council, evaluators observed that pupils were “interested, motivated, and showed increased self-confidence” as a result of the Instrumental Enrichment intervention. Here students at the Blake School, Somerset, England, work on an instrument intended to develop organizational skills and abilities.
develop. In a large number of studies, we were concerned about teaching to the test, especially because the problems tend to be IQ test-like. We were also concerned about the "justification of effort" effects that can arise from the extensive commitment needed to the program.

Although the results are mixed, we believe that when the full program is administered by carefully trained, intelligent, motivated, and conscientious instructors, gains can be attained on standard kinds of IQ and aptitude measures. Greatest gains are likely to be in the areas of abstract reasoning and spatial visualization, rather than in the areas of verbal skills or "crystallized" abilities. Indeed, the emphasis of the program is nonverbal, and even the verbal items involve a fairly minimal reading load. The gains are not spectacular, but we have not found any existing program that is likely to result in spectacular gains; we believe that expectations for such gains are unrealistic. There appears to be transfer to school work in some cases, although we are less confident of the generality of transfer, in part because the extent of transfer attained will be so much a function of how well teachers are able to conduct the required "bridging." We see no evidence of gains in insightful, creative, or synthetic thinking abilities, and a content analysis of the course materials leads us to believe that such gains are not likely to be attained, if only because they are not built into the instruction contained in the program.

The program seems most appropriate for students of average or below-average ability at the junior high school level. The skills taught are those most likely already to be possessed in generous amounts by above-average and gifted students. The program seems especially appropriate for students who do not grow up with standard white, North American, middle class experiences and, indeed, a positive factor is that it has been so widely used around the world. Thus, the program may be especially suitable for special, including retarded and learning-disabled, as well as normal populations. Potential program users need to take into account its long duration, its requirement of extensive teacher training, the need for separate instruction, and the kinds of abstract-analytical skills that are emphasized in the curriculum.

Philosophy for Children

Lipman's Philosophy for Children has at its core a series of six novels that describe everyday experiences of children in which the characters apply philosophical thinking to their daily lives. The program is taught as a separate course and is aimed toward developing thinking and reasoning skills through classroom discussion of philosophical topics. It is intended to be used for three 40-minute periods per week by teachers extensively trained in its use. The complete program is for children K-12.

Pixie involves reasoning about language and is for students K-5. Kid and Gus is for the same grade levels, but it focuses on reasoning about nature. Harry Stottlemeier's Discovery, the first of the novels to be written and perhaps the core of the program, is intended to develop basic reasoning skills in children in grades 4 through 7. Lisa concentrates on reasoning in ethics for children in grades 7 through 12, although it is probably most appropriate at the junior high school level. Suki deals with reasoning in language arts for grades 8 through 11, and Mark...
training are increased by having a comprehensive K-12 program. Whereas Instrumental Enrichment seems to result in greatest gains on tests of abstract and spatial reasoning, Philosophy for Children seems to produce the greatest gains on verbal tests of critical thinking abilities. This outcome makes sense with respect to the heavy verbal emphasis of the program, both in terms of the reading of the novels and of the philosophically based classroom discussions.

We believe Philosophy for Children is more motivating to children than the large majority of thinking skills training programs, but there seems to be an underlying white, middle class orientation, even though not all the children in the novels are white or middle class. We view this program as appropriate for children with average, above-average, and gifted intellectual abilities, although below-average students might be able to handle novels below their grade level. Potential users should keep in mind that the program is highly teacher-sensitive and requires extensive teacher training. The program is also costly in terms of time and resources and requires a separate course.

SOI

The SOI program was developed by Mary Meeker (1969) at the SOI (Structure-of-Intellec) Institute. It is based on Guilford's (1967) structure-of-intellect theory, according to which intelligence can be understood in terms of the intersections of five mental operations (cognition, memory, evaluation, convergent production, and divergent production), four contents (figural, symbolic, semantic, and behavioral), and six products (units, classes, relations, systems, transformations, and implications). The crossing of the five operations, four contents, and six products yields a total of 120 allegedly separable intellectual skills.

Meeker claims that factor analysis over a 40-year period has demonstrated empirical support for 96 of the 120 factors in Guilford's model, although research by Horn and Knapp (1973), among others, has seriously undermined the credibility of this research. Note that the critical studies have not necessarily undermined the theory, but rather the empirical data base claimed to support it.
SOI deals with 90 of the factors, namely, those that do not involve behavioral content. Extensive materials have been developed for 25 factors regarded as being most relevant to the broad areas of the school curriculum. These abilities are assessed by the Structure of Intellect-Learning Abilities (SOI-LA) test. The test is usually given before the program, and deficient skills are identified through test-profile results. Students complete exercises emphasizing those areas in which they are deficient and then take a post-test. Note that this form of implementation means that different students will receive a somewhat different program, reflecting initial individual differences in SOI-LA test score profiles.

The institute is a clearinghouse for research involving the SOI materials, although the studies are apparently not sponsored by the institute or its personnel. The institute does only norming studies and collection and distribution of validity studies conducted by others. We reviewed 15 studies of the SOI-LA measure and 21 of the SOI training program. Our focus here will be upon the latter set of studies.

The studies we reviewed are poorly reported and appear to have many design problems. A disproportionate number involve gifted children, although we have some difficulty imagining supplementary instruction for gifted children involving little more than training on IQ test items, regardless of the theory by which such items are generated. Although the results are mixed, the training appears to result in some gains from pre- to post-test on the SOI-LA. Of course, training is directed explicitly toward the model on which this test is based. Thus, the training appears to yield some gain, to the extent that one accepts the structure of intellect model as representing intelligence or thinking skills, and to the extent that one accepts the SOI-LA as fully representing the model. There appears to be some transfer, but primarily to other tests that measure skills comparable to those measured on the SOI-LA. The program is even more testlike than Instrumental Enrichment, and we doubt whether transfer would be obtained to non-testlike situations. The SOI training program, unlike the IE one, does not have bridging or its analogue as an integral part of the program. We have reservations about the validity of the SOI model, and hence for the test and training based on it. Because of the mixed nature of the results, consumers should be aware that the main gains to be had may be in performance on SOI-based ability test items. The program might sensibly be used, however, by adherents of Guilford's model who are committed to training the particular skills generated by it.

Potential program users should understand that the exercises are very testlike and perhaps not terribly interesting to many children. At the same time, the program is much closer to being teacher-proof than any of the other programs we investigated because the training is prescribed by the score pattern on the SOI-LA and also because the training is relatively regimented. It does not involve the class discussion elements of the two programs discussed above, which are very extensive in Philosophy for Children and somewhat extensive in Instrumental Enrichment.
Highlights of Research on the Effectiveness of Intellectual Skills Instruction

The following five programs are widely used, diverse thinking skills training programs, which have been subject to some evaluation. Although the evidence is scant and more studies are needed, gains in thinking skills are probably possible from using these programs if districts implement them properly and have strong program commitment.

**Instrumental Enrichment**
- Requires three to four hours per week as a separate curriculum, a great deal of paper-and-pencil activity, and extensive teacher training; emphasizes abstract-analytical skills; completion can take two years.
- Appropriate for students of average or below-average ability at the junior high school level, for special as well as normal populations.
- Gains possible on standard IQ and aptitude measures, particularly in the areas of abstract reasoning and spatial visualization.
- Transfer to school work in some cases.

**Philosophy for Children**
- Should be used for three 40-minute periods per week, is highly teacher-sensitive, requires extensive teacher training and a separate course, and is costly in terms of time and resources.
- Appropriate for children K-12 with average, above-average, and gifted intellectual abilities.
- Greatest gains on verbal tests of critical thinking abilities; highly motivating to children.
- Transfer built into the program.

**SOI**
- More teacher-proof than the other four programs; exercises are very test-like and perhaps not challenging to many children.
- Appropriate for children of varying ages and ability levels.
- Gains from pre- to post-test on the SOI-Learning Abilities Test.
- Some transfer evident, primarily to other tests that measure skills comparable to those measured on the SOI-Learning Abilities Test.

**Problem Solving and Comprehension: A Short Course in Analytical Reasoning**
- Lacks both theoretical rationale for academic problem solving and explicit generalizable instructions; is not as broad in its coverage as other four programs; is closest of the five to standard academic work and best used in conjunction with another program.
- Appropriate for high school and college students.
- Gains primarily in analytical problem-solving strategies applicable to the problems present in the program.
- No transfer studies.

**Odyssey**
- Has been evaluated only in Venezuela; consists of 100 45-minute lessons, each of which is accompanied by a rationale, set of objectives and target behaviors, and class procedure plan; is not based on any one theory of cognitive functioning.
- Appropriate for most 6th to 8th graders.
- An imaginative program that can probably contribute to gains in thinking skills.
- No transfer studies.

Problem Solving and Comprehension: A Short Course in Analytical Reasoning

The authors of Problem Solving and Comprehension (Whimbey and Lochhead, 1979), view thinking as a complex set of skills that can be acquired through practice, and that enables the learner actively to construct concepts and relations. Whimbey and Lochhead draw an analogy between learning physical and thinking skills: both involve demonstration and use of the skills, identification of the components of the skills, practice with feedback, and learning over time.

“Although all but a few of the available evaluations leave a great deal to be desired, there are enough positive results to suggest the potential for gains.”

The program concentrates on what the authors view as four components of problem solving: (1) decoding skills, (2) vocabulary, (3) basic arithmetic operations, and (4) precise thinking. Sources of failure in problem solving that the course attempts to remedy include failure to use all relevant information, making leaps in logical and inferential reasoning that are too large, failure to identify appropriate relationships, and failure to collect sound information. The program seeks to develop at least five attributes of good problem solving, namely, concern for accuracy, positive attitude, problem decomposition skills, distance from guessing, and active problem solving.

While this program stresses academic problem solving, it lacks both theoretical rationale for academic problem solving and explicit generalizable instructions for how to do it. It is not as broad in its coverage as the other programs considered in this article. The program, which is appropriate for high school and college students, emphasizes a teaching method called thinking-aloud pair problem solving (TAPS), in which paired students alternate as problem solver and listener. The problem solver reads and thinks aloud while analyzing worked-out examples and solving new problems, and the listener assists him or her in reflecting upon problem-solving procedures.

Empirical data on this program are...
very scant. Moreover, it is often used in combination with other procedures, rendering problematical isolation of the specific contribution of the program to the results. We located only three evaluations and were unable to draw any conclusions from them. Either the program was used in conjunction with other programs, resulting in confounded variables, or the reporting was too scanty to be useful. A major contribution of the program may be the TAPS procedure, which seems to provide a useful vehicle for learning problem solving. The procedure may be problematical for low-ability students who have difficulty communicating with their partners, or for students who are susceptible to friction or competition with their paired classmates.

This program is the closest of the ones we have surveyed to standard academic work, and shows how fine the line can be between teaching thinking skills and teaching standard academic content. It is probably best used in conjunction with another program, and, indeed, this is the primary way in which it appears to have been used. Although no clear psychological theory is behind the program, the authors seem to base their ideas loosely on information-processing theory. What is learned seems primarily to be a set of problem-solving strategies applicable to the problems that happen to be in the program. These strategies are probably useful in the analytical problem-solving domain, but we question whether transfer studies, which remain to be done, would show much generalization to problems with different surface structures but similar “deep” structures.

Odyssey

Odyssey is based on Harvard’s Project Intelligence, which was implemented in Venezuela in the early 1980s during the days of the Ministry for the Development of Intelligence. The program involves six aspects of cognitive functioning, each comprising a separate student book and teacher’s manual—foundations of reasoning, understanding language, verbal reasoning, problem solving, decision making, and inventive thinking. The program contains 100 45-minute lessons, and each is accompanied by a rationale, set of objectives, set of target behaviors, and class procedure plan. The project was aimed at 7th graders of all ability levels, but Odyssey could probably be used in the 6th to 8th grade levels in most school districts.

The only evaluation (Herrnstein et al. 1986) has been of the program’s success in Venezuela, although it is probably as good as or better than any other single evaluation of any program that we could find. The very thorough and carefully designed evaluation leads us to conclude that Odyssey can probably contribute to some gains in thinking skills. The program is imaginative and is not based on any one theory of cognitive functioning. Like Whimbey and Lochhead’s program, it seems to be based on information-processing principles. We see this as a promising program for future development and evaluation.

Positive Programs: Questionable Evaluations

Some thinking skills training programs are probably not a whole lot better than snake oil, but the good ones, although not miracle cures, may improve thinking skills. A present trend is toward the development of theoretically based programs (e.g., see Bransford and Stein 1984, Sternberg 1986), and we believe that such programs, if founded on sound psychological and instructional theories, yield the greatest opportunities for gains. Although all but a few of the available evaluations leave a great deal to be desired, there are enough positive results to suggest the potential for gains. Although any one study is likely to be problematical, the sources of problems differ across studies, and so the convergence of positive results is heartening about the better programs, if not about the quality of the evaluations of the programs. Ultimately, any school district adopting a program will need to conduct its own evaluation, since our analysis reveals that the success of a given program depends on a large number of implementation-specific factors, such as quality of teaching, administrative support, appropriateness of the program for the student population, and the extent to which the program is implemented in the intended manner.

Although we are not optimistic about the evaluations, we are confident about the possibility for thinking skills instruction. The opportunities exist to increase students' thinking skills, and the time to seize them is now. What is needed to make such instruction succeed is cautious planning, a sound program, effective implementation, strong commitment, and diligent evaluation. When these ingredients are present, instruction in thinking skills is both possible and feasible.

References


Preparation of this article was supported by Contract N00014-85-K-0589 from the Office of Naval Research, and Contract MDA903485K0589 from the Army Research Institute.

Robert J. Sternberg is IBM professor of psychology and education, Department of Psychology, P.O. Box 11A, Yale Station, New Haven, CT 06520-7447. Kastoor Bhana is a visiting scholar to Yale from the University of Durban Westville, P.O. Box X54001, Durban 4000, South Africa.