INTELLIGENCE **10**, 209–230 (1986)

Kvashchev's Experiment: Can We Boost Intelligence?

LAZAR STANKOV
The University of Sydney
Australia

In an effort to improve performance of high school students on intelligence tests, a large-scale study involving 296 students was carried out. Members of the experimental group (N = 149) were given exercises in creative problem solving 3 to 4 times a week over a period of 3 years and performance was assessed on four occasions at the beginning of the experiment, at the end of it, and twice during the fourth year of High School. The test battery contained 28 measures of fluid and crystallized intelligence. The results indicate that it is possible to achieve small improvement in performance and that this improvement remains 1 year after the end of training. The improvement is more pronounced with measures of fluid intelligence than with measures of crystallized intelligence.

INTRODUCTION

By and large, educationists are optimists. Many of them refuse to accept the suggestion that the impact of their practices is restricted to the transmission of knowledge. There is a belief that their activity also leads to an increase in something (maybe capacity or skill or the like) for carrying out cognitive tasks

---

In the early months of 1983, I started making arrangements to spend my 1984 half-sabbatical in Yugoslavia. My intention was to have another look, with Kvashchev, at all the data he had gathered over the years, to reanalyze those sets of data that warranted further analyses, and, maybe, to write a joint monograph with him for the English-speaking audience. By that time I had known him for 27 years. He used to be my high school teacher and I had helped him with statistical analyses many times since my graduation from high school in 1960. I knew and trusted the man and it had become clear to me that his work on practice and training of cognitive abilities had reached almost monumental proportions.

Kvashchev died on May 10, 1983, aged 53. During the last 12 years of his life he was Professor of Educational Psychology at the University of Belgrade. At his death he had published 11 books in Serbocroatian. These books contain detailed descriptions of his work. Although each book features rather long abstracts in both English and Russian, these abstracts suffer from poor quality of translation. On request, I will supply an abstract in English from his last book.

I went to Yugoslavia hoping for a clear, positive statement and, after having another close look at Kvashchev's data, I remain cautiously optimistic but unable to force myself to remove the questioning tone from the title of this paper.

Correspondence and requests for reprints should be addressed to Lazar Stankov, Department of Psychology, The University of Sydney, Sydney, N.S.W. 2006, Australia.

proficiently. Because it is obvious that genetic factors provide a limit to what can be accomplished through education, and new techniques of teaching are being introduced regularly, questions regarding the nature and the extent of their impact have to be addressed repeatedly.

The ideas underlying the study reported here originated in the early 1960s, but the data were gathered in the mid-seventies. These ideas derive from work on creativity and from a belief that it is possible to teach students to develop divergent thinking and other related abilities. This work is important for us now, because even a casual perusal of textbooks used in our schools today shows that similar ideas are present in contemporary educational practices. It is pertinent to ask if their acceptance is justified and, in particular, if exercises in creative problem solving can improve performance on typical IQ tests.

The present study is the last and the most ambitious effort of the late Yugoslavian psychologist, Radivoy Kvashchev.[1] Prior to this, he had carried out several similar projects in an attempt to improve performance on measures of critical thinking, creative thinking, and school learning. The experience gained in these earlier studies was used in the final project and his work deserves our attention because of the scale of its execution. It is based on almost 300 subjects, half of whom were given intensive exercises (three to four school periods per week) over a 3-year period. The effect of this intervention was assessed with a battery of 37 cognitive tests which was given four times in the course of the experiment. In the literature on educational psychology it is hard to find many studies with the same scope.

Kvashchev was convinced that performance on cognitive tasks can be reliably improved through training, provided that this training can be carried out for a prolonged period of time (at least 9 months) and that its theoretical basis is grounded in psychological research findings. We shall see that he was rather eclectic in the choice of theoretical frameworks. Two aspects of Kvashchev's work were derived from his job situation. For a decade he was employed as a school psychologist in a small-town high school. For this reason, the research findings reported in this paper apply to late adolescents and young adults only. Also, he was obliged to stay close to the school curriculum and to use school material in his exercises. Care was taken not to employ exercises that involve either the content or the format of the typical intelligence tests. He wanted to be in a position to make claims about the more general transfer of training.

Current Trends in the Attempts to Increase Intelligence

It appears (see Anastasi, 1981; Detterman & Sternberg, 1982) that more recent efforts to raise intelligence differ in both the choice of subjects who should be

---

[1]The work was published in Serbocroatian (Kvashchev, 1980) and the information regarding the particulars of experimental design to be presented here was obtained from this source. I am grateful to the Institute of Psychology, Belgrade, for permission to reanalyze Professor Kvashchev's data. I am also grateful to Professor A. R. Jensen and to two reviewers of this paper (McCall & Sternberg) for useful comments regarding an earlier version of this manuscript.

trained and in theoretical approaches from the one used by Kvashchev. The choice of subjects commonly reflects the need to improve the conditions of various disadvantaged groups including the mentally retarded. As a consequence, these other studies tend to concentrate on younger kindergarten and elementary school children and children within the lower IQ ranges. Kvashchev's sample involves high school adolescents who, on the average, have a somewhat higher IQ in comparison to the general population. This is because the wide range of trade schools in Yugoslavia tend to attract academically less able students away from the high schools. Nevertheless, because the aim of these studies was to increase global IQ in a similar way to what was done in this study, their outcomes could provide a broad basis for evaluating the results presented here.

A recent review of the literature (Caruso, Taylor, & Detterman, 1982) indicates that between 40% and 60% of the studies that involve various attempts to raise intelligence test performance show no effect. In those studies that show statistically significant improvement, the authors note that the improvement is relatively small and it does not last for a long period of time. According to these authors, research to date indicates that an increase in IQ of 10 to 20 points is the maximum that can be anticipated. Anastasi (1981) is somewhat more optimistic but this optimism seems to be based upon the outcomes of the studies that involved more specific transfer of training. In this context, we may expect that Kvashchev's work will not be particularly successful because he deals with the age and IQ group that is considered less prone to ameliorative influences.

An important conceptual framework for the study of intelligence nowadays is the information-processing approach. Some of the studies in this area have focused on typical intelligence test items and, due to this work, we now have a deeper understanding of the important aspects of performance on many such tasks. Recently, successful attempts have been made to use some of these findings to train people to improve their performance in reading comprehension, vocabulary acquisition, deductive and inductive reasoning, and the learning of—and memory for—lists of words (Belmont, Butterfield, & Ferretti, 1982; Glaser & Pellegrino, 1982; Sternberg, Ketron, & Powell, 1982). These studies differ from Kvashchev's approach in being more specific in their selection of tasks, training procedures and the expected transfer effects. Nevertheless, they have implications for his work since they clearly point to the importance of executive processing, strategy planning, and similar higher order skills rather than the narrower performance aspects. These executive skills presumably have a greater transfer potential across the tasks and, therefore, by training these skills one has a better chance of affecting intelligence.

By building upon the ideas of Piaget, Hebb and others, Kvashchev claimed that his system of training was aimed at developing a set of six "general cognitive schemas," which are akin to the more recent notions covered by the term executive processing (see Sternberg et al., 1982). For example, the general cognitive schema of originality involves, among other things, a realization that concepts can have different meanings in different contexts and can reveal a readiness to depart from the standard meanings of the concepts in order to achieve unexpected solutions, tolerance of ambiguity, and so on. Because the development of cognitive schemas is based on a great variety of tasks and on school material, he expected that training would influence intelligence.

The Nature of the Exercises

All exercises used in this experiment involve what may be called "creative problem solving." The principles upon which they were constructed derive from three main sources: (a) Mednick's (1962) and Maltzman's (1960) analysis of the associative basis of creative process; (b) Gestalt theory of problem solving (Duncker, 1945; Maier, 1945; Maier, Julius, & Thurber, 1967; Wertheimer, 1959); and (c) factor analytic results of the kind summarized in Guilford's (1967) and Guilford and Hoepfner's (1971) writings on the structure of the intellect model. These three sources provided a set of broad (noncontradictory) principles which were used as guidelines in constructing the exercises. A short and selected list of these principles is as follows:

1. Exercises should call for a combination of elements which are remote rather than close in terms of their associational value.
2. Exercises should call for a radical reorganization and reformulation of the problem situation in order to achieve a satisfactory solution.
3. Exercises should call for both convergent and divergent thinking operations, especially the latter.

It can be said that almost every textbook or journal article example that has analyzed thinking and problem-solving processes during the past decades was presented to the experimental group either in its original form or, if possible, was translated to conform to the syllabus of a particular school subject. To illustrate, consider the following problem:

> I was captured by a band of outlaws—said a famous explorer—and their leader had my hands and legs tied up so that I could not move. They did not gag me up though, and I was able to use my mouth freely. The leader of the gang hung a piece of bread exactly five centimeters away from my mouth. He then laughed and said: "If you manage to eat this piece of bread, I'll set you free." He knew that I could get no help. Also, in order to ensure that I cannot roll over or move closer to bread, they tied me to a tree. Nevertheless, I managed to free myself. How?

The subjects had to list as many ways as they could think of to solve the problem. Because of the explicit statement that the mouth was free, the task resembles the well-known "direction of thinking" problem emphasized by some gestaltists (e.g., Maier, 1945) and the divergent thinking or fluency task emphasized by the students of creativity. Also, the problem would encourage answers that have a low associational value as suggested by Mednick (1962). The accept-able solutions include realistic blowing at the bread in order to create a pendulum or solutions that assume some particularly favorable conditions. These include the wind blowing and moving either the rope or the tree to which the explorer is attached (or both)—or some similar quirk of fate.

I believe that the most salient features of the exercises were: (a) They required a very active participation by the subjects; (b) they called the subjects' attention to the principles deemed important for creative thinking (e.g., producing as imaginative and as novel a solution as you can). This was done either implicitly through examples or explicitly by formally defining the principle in the instructions (e.g., transforming the given object in order to find the solution); and (c) subjects were provided with the feedback. This feedback was quite often built into the problem itself, as illustrated by the examples in this method section of this paper.

It should be kept in mind that throughout this experiment all subjects were attending to their normal school activities. The school system in Yugoslavia nowadays is similar to many European systems in that it calls for an acquisition of a large body of factual information (provoking frequent discussions in the press regarding the overburdening of students). Teaching methods for most part involve traditional lecturing and oral assessment of students' ability to understand the material. Typically, students are not encouraged to work on different "projects" and display their own initiative during the indepth enquiries about a particular topic. The exercises in creative problem solving are to be seen against this general background—both experimental and control groups operated within this milieu—and exercises were superimposed on routine schoolwork.

Expectations

The training in creative problem solving was presumed to affect intelligence which, in turn, was defined in a typical psychometric way. During the past 30 years, that is, since breaking political ties with the Soviet Union, Yugoslavian psychologists have been adopting psychological tests developed in the West. At the beginning, these tests were nonverbal but, gradually, original verbal tests became available as well. Kvashchev was keen to represent both in his battery of tests. The assembled tests, in fact, included measures of both fluid and crystallized intelligence. As a consequence, his hypotheses were couched in terms of the theory of fluid and crystallized intelligence.

Typical interpretations of crystallized intelligence (e.g., Cattell, 1971; Horn, 1981) suggest that this broad ability organization should be susceptible to environmental effects and, therefore, it was postulated that the experimental group would outperform the control group on these tasks. However, in some of Kvashchev's previous studies it was found that Raven's Progressive Matrices test was also affected by training similar to that employed here. To some extent this was surprising because a large majority of the exercises involved verbal material and, superficially at least, seemed to involve processes that differ from Raven. This finding prompted the present study.

Although many people believe that fluid intelligence cannot be affected by learning, for a long time John Horn has been explicit in claiming the opposite (see, e.g., Horn, 1985). In his view, individual differences in fluid intelligence are influenced by casual, incidental, and largely idiosyncratic learning processes. Crystallized intelligence, on the other hand, reflects processes of formal education and acculturation—and the available knowledge base (including both content and operations) is an outgrowth of these influences. Because the main aim of the exercises used in this study was to develop general cognitive schemas, the training itself may influence the general acquisition processes tapped by the fluid intelligence markers (more than the knowledge base, per se). Due to the massive nature of the exercises and a great variety of the types of tasks, chances for affecting the casual learning processes in most members of the experimental group may be greater that what is typically available in school environments. If this happens, better performance should be possible to detect with the fluid intelligence tests.[2]

Another expectation was that the induced increases in intelligence would be long lasting. In the context of this experiment, this meant that the experimental group would be superior to the control group one year after the end of training. In Yugoslavia, as in other countries, once high school students leave their homes in order to go to colleges and universities, it is very hard to check on their academic progress. This explains why longer lasting experimental effects could not be checked in this work.

METHOD

Subjects

Participants in this study were all members of a generation that enrolled as freshmen in two high schools. In Yugoslavia, children start their formal education in primary school at the age of 7 and compulsory schooling lasts for 8 years. Thus subjects in this experiment were 15 years of age at the commencement of their high school training.

Since the end of World War II, due to a shortage of classroom space, Yugoslavian schools work in shifts (mornings from 8 a.m. till 1 p.m. and afternoons, from 2 to 7 p.m.) with two schools sharing the same facilities and, typically, alternating shifts every week. However, the main differences between schools is in terms of teaching staff. Every year these two particular schools start five new first-grade classes of approximately 30 students each. Classes themselves are formed on the basis of primary school marks with the objective of having about the same spread of ability in every class. From a total of 10 classes,

---

[2]Horn (1985) presents a review of some recent studies that illuminate the role of genetics and environment in fluid and crystallized intelligence. The above predictions for Kvashchev's experiment are consistent with these findings.

Kvashchev made a random choice of 5 classes for his experimental group. The remaining 5 were declared a control group.

At the beginning, the experimental group had 171 subjects and the control group, 156. In both groups, about half of the samples were males. In Yugoslavia, people are classified into three main socioeconomic status (SES) groups. The breakdown of the total sample in terms of their parents' classification in these groups is as follows: (a) professionals and white-collar workers ($N = 98$), (b) blue-collar workers ($N = 121$), and (c) farmers ($N = 108$). It is apparent that representation of the three groups in this sample does not differ very much and is typical of the total population of high school students in the country. Because the experiment lasted 4 years, some attrition of subject samples took place. As a result, a complete record is available for 149 members of the experimental group and 147 members of the control group. All analyses reported here are based on those subjects who had complete records available.[3]

Procedure

The design of the experiment is relatively simple, involving two parallel groups who were asked to complete a battery of paper-and-pencil tests of intelligence four times. For 3 school years, the experimental group was given exercises in creative problem solving, whereas the control group continued its activities in a typical way. Both groups were tested shortly after their enrollment into the first grade and prior to being exposed to any exercises. This will be referred to as "initial testing." They were given the same battery at the end of the third grade, that is, at the end of the experiment. This will be called "final testing." In order to find out if the effects of training were long lasting, two retests were given. Both retests took place at the beginning, and at the end, of the fourth year, that is, during the very last year of high school.

Some exercises in creative problem solving were administered by Kvashchev himself. School authorities allowed him to work for 1 hour in each experimental class every week. He was also allowed to act as a replacement teacher in these classes whenever there was a need to replace any of the absent teachers.

In addition to his own participation in the exercises, Kvashchev also trained regular school teachers to do the same. Various sections of the syllabus for a particular school subject were chosen and detailed sets of procedures that embody principles outlined in the introduction were applied to the presentation of a given topic. Apart from mathematics, which embodies problem-solving activities in its ordinary mode of presentation, virtually every academic school subject was treated in this way. On the average, every class in the experimental group was provided with 3 to 4 hours of exercises per week.

To illustrate the nature of the exercises he employed, it is interesting to

[3]Attrition was approximately the same for the three SES groups within both samples.

consider the following three examples from a large number of tasks he developed:

Example 1. In one series of tasks he would ask his subjects to try to think of a practical solution to a given problem. For example, he would say: "It will be much easier to move materials from the Moon to the Earth than vice versa because the return trip would require about 97% less energy expenditure. This follows from two facts: First, the Moon's gravity is about 6 times weaker and, second, the Moon's mass is about 18 times smaller than the Earth's mass. Can you develop a method that would allow you to transport the goods from the Moon to the Earth without the use of rockets?" After 5 to 10 min, he would say: "Some of you have expressed the opinion that electric railway lines would be the most convenient means for this purpose. Railway line would be practically levelled but, in fact, it would rise for about 5 km. This would be sufficient to develop a velocity which would allow for the launching. Can you think of yet another system?"

Example 2. In another series of tasks he emphasized the need to organize all the information in a way that would enable the students to see the problem from a variety of viewpoints. For example, he would ask them to list all possible scientific means that can be used to solve the problem of feeding a world population which will double within the next 30 years. Again, after allowing them to work for 5 to 10 min, he would say: "You have mentioned, among the other things, the following: (a) We can use a huge biological factory that would work with the minimum energy expansion and with the minimum effect on the immediate environment. This factory could use agricultural waste to produce sugar, alcohol, proteins, and drugs. (b) We can develop research into ways that would allow for a better utilization of nitrogen by the plants. This would reduce the use of fertilizers and, as a consequence, improve agricultural production and reduce their damaging influence on the environment. (c) We can develop biological control of insect populations which, in some parts of the world, seriously affect the crops. (d) We can develop new plant varieties that can be used for feeding human populations. Can you think of any other solutions?"

Example 3. In one series of tasks he encouraged his subjects to overcome the narrow confines of a given problem and to try to uncover in it new, hidden meanings. For example, he would tell them the following: "Quite often there will be a need to read between the lines, to discover the ideas that have not been explicitly mentioned but are implied by a given text. Try to do this with the following text: 'Infrasound is the oscillation with a frequency range below 20Hz. The word *vibration* means, among other things, quiver, vacillation, a stirring, and so on. If there is a cyclone far away on the ocean, there is an increase in the number of traffic accidents on the shore and the conditions of patients worsen in hospitals. What are the implications of this text?" After allowing them to work for 5 min, he would say: "Some of you have discovered the following hidden ideas in the text: Vibration was the cause of these unpleasant events or, more precisely, infrasound itself may be their cause. This conclusion can be deduced from the given text."

Test Battery

Altogether, 37 different cognitive tests were given to all the subjects. We shall focus our analyses on a subset of 28 variables. All these tests have been used by Yugoslavian psychologists and, for most of them, standardization information is available. The remaining 9 variables represent either achievement or aptitude tests constructed especially for this study, or tests developed abroad and used for the first time in this study.

The 28 variables, in fact, represent subtests from five intelligence scales. Two of these scales are verbal: (a) "B-series" (9 subtests) developed by Z. Bujas and (b) Stevanovic's intelligence scale (7 verbal subtests and 1, nonverbal). The remaining three are nonverbal: (c) Nonverbal Scale of Intelligence (NSI, 4 subtests) developed by Z. Bujas, (d) Dominos test D-48 which produces only one score, and (e) Cattell's Culture Fair test (6 subtests). Brief descriptions of the 28 variables used in this study are as follows:

Verbal Tests

1. *Essential features test.* In each item of this test subjects have to indicate what are the essential features of an object. Example: RIVER(2)______salt, fish, gold, water, ship, bridge, channel. The number in parentheses indicates how many features from the list should be marked. Correct answer: water and channel.
2. *Word classification test.* Given a list of words like KINDERGARTEN,______, HIGH SCHOOL, UNIVERSITY, subjects have to fill in the missing word from the following list: Music School, Trade School, Primary School, College.
3. *Number series test.*
4. *Consequences.* Indicate the most important, common, and regular consequences of a given situation. Example: TOOTHACHE(2)______appetite, unpleasantness, reading, brushing teeth, thinking, visit to a dentist. The number in parentheses indicates the number of alternatives to be chosen from the list.
5. *Relations test.* Example:
   HEN, EGGS = CHICKENS
   EGGS, FRYING PAN = _____
   courtyard, fire, kitchen, scrambled eggs, potato, matches.
   In this test, subjects have to study the first row of words and deduce that one of the relationships between hen and eggs is the production of chickens. In the second row they have to indicate the product of eggs and frying pan—scrambled eggs.
6. *Word meanings test.* Given a word, order a set of 9 words with respect to their similarity to the original word. Synonyms were to be assigned rank one.

7. *Unbalanced structures test.* Given a set of words: COAT, SHIRT, UMBRELLA, CAP, RAINCOAT, subjects have to find out which word does not belong to the list (UMBRELLA) and replace it with an appropriate word from another list: stick, sun umbrella, socks, bag, briefcase. The correct answer is SOCKS since this is another wearable object.
8. *Numeric test.* This test consists of items involving two equations with two unknowns. Example: $a + b = 6$ and $b - 1 = 1$. Find the values for a and b.
9. *Polyprofile test.* Two lists of words are given. Example:
THERMOMETER, BAROMETER, ANEMOMETER, ____ (2)
CLOCK, KILOGRAM, A MILE, PRESSURE, MILLIMETER, WIND.
There are two correct solutions (indicated by the number at the end of the first line) for this problem. One solution is CLOCK because this is a measuring instrument like the other words in the list, and the other one is MILLIMETER because this is the word that ends in "meter."
10. *Meaningful memory test.* In this test, a fable was read by the experimenter and subsequently subjects were presented with the written text in which 65 words were missing. Students had to fill in as many missing words as they could remember.
11. *Word classification "S" test.* Underline the word that does not belong to a series of five words.
12. *Proverbs test.* Each item in this test consists of four proverbs and twice as many possible interpretations of these proverbs. Subjects have to match these.
13. *Verbal analogies test.*
14. *Disarranged sentences test.* Carry out the operation indicated by a sentence with words permuted (e.g., underline a particular word, etc.).
15. *Proverbs interpretation test.* Write down the meaning of a given proverb.
16. *Arithmetic test.* A typical test item of the following kind: "If five people can build a house in 20 days, how many people would be needed to build a house in 15 days?"

Nonverbal Tests

17. *Perceptual reasoning test.* A form of Matrices test.
18. *Pictorial unbalanced structures test.* This is a nonverbal analog of Test No. 7. Each item consists of a matrix of 9 drawings arranged according to a particular rule. Subjects have to find the drawing that does not fit into the matrix and choose from a list of six alternative drawings the one that should replace it.
19. *Combined solutions test.* This test is similar to the Raven's Progressive Matrices test. The main difference derives from the fact that several possi-

ble solutions exist for every item. Subjects are told that there is more than one correct solution and that they should find them. They are not told how many possible solutions exist. Some of the correct answers depend on perceptual, whereas the others depend on more abstract, properties of the matrix.

20. *Multiple solutions test.* Again, this test is similar to Raven's Progressive Matrices test. The main variation consists in the fact that subjects have to rank order the alternatives with respect to their closeness to the correct solution.
21. *Pictorial polyprofile test.* Each item consists of a matrix of 9 drawings and 16 alternative drawings. For every matrix, subjects are told how many possible correct answers exist (2 or 4). The task is to find all possible solutions.
22. *Dominos test, D-48.* In Yugoslavia, this is considered to be the best non-verbal '*g*' measure. The items of this test consist of domino drawings which are arranged in a particular order. Subjects have to find the rule and fill in the last blank domino.
23. *Figure classification test.* A subtest of Cattell's Culture Fair test of intelligence.
24. *Projections in water test.* A subtest of Cattell's Culture Fair test of intelligence.
25. *Figure series test.* A subtest of Cattell's Culture Fair test of intelligence.

26–28. Three different versions of Cattell's *Matrices* test.

Although the above 28 tests have been divided into two groups—verbal and nonverbal tests—it is clear that these two groups correspond only approximately to crystallized and fluid intelligence. For example, the Number Series test is a fluid intelligence marker and the Meaningful Memory and Arithmetic test may, at least in part, represent a measure of short-term acquisition and retrieval function (SAR, see Stankov, Horn and Roy, 1980) rather than crystallized intelligence. There are also some tests (e.g., the Numeric test) that are known measures of both fluid and crystallized intelligence. Preliminary factor analyses of these data indicate that these two broad factors are clearly defined. We are now studying the structure among the variables with confirmatory factor analytic procedures and the results of these analyses will be reported in a separate paper.

Statistical Analyses

The main method of analysis in this paper is the analysis of covariance (ANOCVA). This univariate analysis was accomplished with the multivariate analysis of variance (MANOVA) procedure of the Statistical Package for Social Sciences (SPSS^{X}, Nie, 1985). The results of multivariate analyses will be reported in a separate report.

TABLE 1
Arithmetic Means and Standard Deviations for Both Groups on Four Occasions

	Control Group								Experimental Group							
	Initial		Final		1st Retest		2nd Retest		Initial		Final		1st Retest		2nd Retest	
Variables	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1. Essential features test	30.76	5.45	34.70	4.86	37.37	4.03	37.49	4.21	28.41	6.72	36.24	4.59	37.88	3.79	39.42	2.83
2. Word classification test	10.84	2.52	11.71	2.52	13.07	2.16	12.84	3.51	10.05	2.76	12.48	2.38	13.47	1.99	14.29	1.56
3. Number series test	10.61	2.37	11.53	2.23	12.49	2.10	12.46	2.14	9.87	2.44	12.11	2.20	13.01	1.99	13.77	1.81
4. Consequences test	34.90	4.92	36.77	4.52	38.63	3.84	38.33	3.29	31.91	6.41	36.76	4.26	38.38	3.29	39.21	3.38
5. Relations test	12.56	2.63	13.86	2.78	14.61	1.90	14.87	1.66	10.59	3.04	13.75	1.93	14.89	1.36	15.15	1.21
6. Word meanings test	28.78	6.83	32.39	6.85	36.99	5.90	37.32	5.02	26.63	7.48	34.38	7.05	36.71	5.97	39.20	4.25
7. Unbalanced structures	10.28	2.53	11.48	2.82	13.13	2.32	13.19	2.08	8.85	2.62	11.75	2.29	13.38	1.85	13.66	1.73
8. Numeric test	14.08	2.01	14.61	1.36	14.78	1.56	14.88	1.11	13.60	2.07	14.53	1.64	14.97	1.18	15.12	1.11
9. Polyprofile test	23.00	6.67	27.32	7.00	31.61	6.86	33.01	6.33	20.53	6.00	29.44	5.76	32.41	6.12	35.33	5.23
10. Meaningful memory test	12.96	4.55	16.72	6.38	20.18	7.60	21.72	8.59	13.81	5.53	19.21	7.94	22.36	8.20	25.91	8.63
11. Word classification "S"	14.29	2.79	17.77	2.63	18.94	2.79	18.59	2.97	15.46	2.79	18.05	2.68	20.05	2.65	20.82	2.26
12. Proverbs test	9.50	3.93	13.39	4.61	15.76	3.80	16.22	4.00	9.46	3.95	14.19	4.56	16.85	3.67	17.81	2.92
13. Verbal analogies test	19.51	5.11	23.01	3.86	24.49	3.11	24.73	3.25	20.06	4.76	23.70	3.70	25.52	2.80	26.11	2.66
14. Disarranged sentences	10.42	4.12	13.52	3.57	16.44	3.79	16.84	3.62	10.11	3.88	14.27	4.20	17.02	3.34	17.38	3.52
15. Proverbs interpretation	6.71	2.36	7.88	2.19	9.37	2.00	9.44	2.06	6.93	2.37	8.32	2.08	9.80	2.19	9.89	1.98
16. Arithmetic test	6.85	3.09	8.54	3.38	10.48	4.07	11.76	4.34	7.20	4.43	9.74	3.88	13.60	5.05	13.62	5.42
17. Perceptual reasoning test	23.84	5.28	28.93	4.76	31.90	5.24	33.84	4.43	22.54	6.62	27.50	5.85	33.02	4.91	35.05	4.00
18. Pictorial unbalanced structures	41.36	9.64	45.90	9.97	47.51	10.61	49.94	9.28	38.18	8.16	46.71	8.20	52.67	6.48	53.24	6.08
19. Combined solutions test	39.99	10.09	42.96	10.07	43.44	11.11	45.73	8.95	37.47	9.44	45.40	9.09	49.14	6.69	49.17	7.41
20. Multiple solutions test	44.54	6.55	47.17	6.00	47.98	6.54	50.68	4.81	42.55	7.15	48.24	5.87	50.29	4.95	51.91	3.70
21. Pictorial polyprofile	38.68	6.18	42.44	5.14	44.52	5.17	45.44	5.83	38.03	5.99	42.96	5.82	46.62	5.54	48.28	4.46
22. Dominos test, D-48	26.12	4.98	29.76	5.15	31.67	5.16	32.82	4.96	28.09	4.41	32.21	4.47	32.84	5.72	34.62	4.65
23. Figure classification test	6.62	2.39	8.61	2.37	9.37	2.16	9.89	2.20	6.40	2.41	9.38	2.28	10.11	2.33	10.89	2.14
24. Projections in water test	7.22	1.62	7.69	1.30	8.12	1.14	8.12	1.23	7.00	1.69	8.00	1.24	8.40	0.90	8.43	0.92
25. Figure series test	9.95	2.68	11.40	2.26	11.98	2.42	12.31	2.20	9.42	3.29	12.04	1.93	12.83	1.89	13.38	2.32
26. Matrices I	8.25	1.46	8.95	1.20	9.18	1.37	9.34	0.94	7.85	1.54	9.18	1.03	9.40	0.89	9.56	0.74
27. Matrices II	9.23	1.52	9.85	1.20	10.05	1.20	9.99	1.15	8.24	1.82	10.07	1.23	10.35	0.85	10.50	0.79
28. Matrices III	5.46	2.72	6.80	2.28	7.52	2.29	7.51	2.07	5.46	2.64	7.81	2.06	9.17	1.84	8.90	1.66

RESULTS

Two questions will be addressed with reference to the data gathered in this study. First, is the experimental group more successful than the control group in its performance on intelligence tests at the end of training? Second, can differences between the groups be detected within a year of the cessation of training?

Arithmetic means and standard deviations for all 28 variables and on all four occasions are presented in Table 1. Because we shall refer to Table 1 several times in this paper, it is worth mentioning the fact that, in the initial testing, the control group performed better than the experimental group on a number of variables. Over all variables, this difference was rather small (about 17% of *SD*) but, due to a large sample size, the differences on some variables were significant. This means that the two groups were not quite the same at the beginning of the experiment, with the control group being slightly superior on the average. On all other occasions, the experimental group, in general, had higher arithmetic means than did the control group.

Another feature of Table 1, which will be of some interest in the latter part of this paper, has to do with the observation that most variables have reasonably similar standard deviations across the occasions. The only exception is the second retest in the experimental group. Most variables on this occasion have a noticeably smaller standard deviation than on any previous occasions. Also, these standard deviations are smaller than corresponding standard deviations for the control group.

In order to understand the main findings of this study, it is convenient to present the results in terms of three separate sets of ANOCVA, with initial testing being treated as covariate in all cases. In this way some of the substantive findings will be made clear and it will also be possible to illuminate a methodological issue important for this particular quasi-experimental design.

ANOCVA with Subjects as Units of Analysis

The first analysis is the one carried out by Kvashchev, himself. Separate univariate ANOCVAs for each variable (with the same variable in initial testing defining the covariate) were obtained for the final test, and also for both retests. The first three columns of Table 2 display the F-tests and, as usual, asterisks indicate statistical significance.

The overwhelming impression one gains by the inspection of Table 2 is a clear preponderance of significant F-tests. Thus, out of 28 variables, 21 were significant at the .05 level and 17 at the .01 level at the end of training (final testing). One can conclude, as Kvashchev did, that training in creative problem solving improves the performance on intelligence tests in the experimental group over and above the performance of the control group. Furthermore, this improvement is about equal for verbal and for nonverbal tests and for fluid and crystallized intelligence markers.

TABLE 2
Differences between Experimental and Control Groups. ANOCVA with Pretest as Covariate. Units of Analysis: Subjects[a]

Variable	*F*-tests (df = 1,293)		
	Final	1st Retest	2nd Retest
1. Essential features test	24.52**	5.17*	37.98**
2. Word classification test	20.08**	7.50**	58.41**
3. Number series test	14.88**	11.11**	50.04**
4. Consequences test	3.82	.64	12.06**
5. Relations test	5.74*	9.76**	9.12**
6. Word meanings test	16.67**	.27	20.36**
7. Unbalanced structures test	8.18**	6.27*	12.13**
8. Numeric test	.06	4.40*	6.92**
9. Polyprofile test	18.18**	4.84*	22.89**
10. Meaningful memory test	6.69*	3.62	15.82**
11. Word classification "S" test	.91	4.71*	39.02**
12. Proverbs test	4.05*	8.96*	22.31**
13. Verbal analogies test	1.55	8.31*	16.22**
14. Disarranged sentences test	5.42*	3.64	2.64
15. Proverbs interpretation test	7.78**	39.63**	11.32**
17. Perceptual reasoning test	2.57	9.03**	11.96**
18. Pictorial unbalanced structures	7.04**	47.63**	38.20**
19. Combined solutions test	14.37**	50.35**	25.79**
20. Multiple solutions test	7.43**	17.30**	10.44**
21. Pictorial polyprofile test	1.85	15.25**	29.38**
22. Dominos test, D-48	7.93**	.07	3.30
23. Figure classification test	15.44**	12.12**	20.52**
24. Projections in water test	7.54**	8.65**	8.55**
25. Figure series test	13.34**	19.82**	24.74**
26. Matrices I	8.69**	7.45**	8.74**
27. Matrices II	8.96**	14.28**	33.33**
28. Matrices III	18.07**	48.29**	44.23**

[a]Asterisks indicate significant *F*-tests: Single asterisk indicates .05 significance; two asterisks indicate .01 significance.

It is also apparent from Table 2 that the benefits acquired by the experimental group remain not only during the 1-year follow-up period, but seem to increase. Thus, in the second retest (i.e., 1 year after the end of active training in problem solving), the experimental group retains its superiority. This is manifest in both the number of significant *F*-tests (25 out of 28) and in their magnitude. It is as if the experimental group continues, and perhaps accelerates, its superiority in the final year of high school—a rather important result for educational practice. We shall return to this finding again later in the paper.

ANOCVA with Classes as Units of Analysis

Older textbooks (those published in the 1950s and 1960s) on experimental design in educational research advocate the use of subjects as units of analysis and, as was the case above, reliance on rather high number of degrees of freedom for the within (or error) term. In the 1970s, it became generally known that the interpretation of ANOCVA results critically depends on the nature of the subject selection procedures. In particular, unless subjects represent a random sample from the population, one is not justified in assuming that the final F-test would be equal to zero when treatment exerts no influence. In other words, under the nonrandom selection procedures, the F-test is expected to be different from zero (see Reichardt, 1979).

Because random selection in this case involved school classes consisting of about 30 subjects each, it was decided to calculate arithmetic means for the whole class and treat that as a unit of analysis. For reasons outlined above, this approach should be preferred to the previous one.[4] This procedure ensures that the overall control and experimental groups' means remain the same but standard deviations change, that is, they become lower. The within-groups degrees of freedom change as well. The F-tests from this analysis are presented in Table 3. Kvashchev did not carry out these analyses.

The overall pattern of significant F-tests changes dramatically when one uses the whole class as a unit of analysis. For example, in the final testing, seven F-tests are significant at the .05 level and only two reach .01 significance. The second retest still provides a larger number of significant values (17 at the .05 level and 9 at the .01 level) but the effect is not as overwhelming as it was previously.

This analysis makes it apparent that the most pronounced improvements in the experimental group occur in the nonverbal and fluid intelligence area rather than in the verbal tasks. Thus, in the "Final" column of Table 3, significant differences exist for variables 18, 19, 23, 24, and 28—all known markers for fluid intelligence. The only crystallized intelligence measure showing significant difference between the groups is the Essential Features test (Variable 1). This is important in view of the presumed nature of these abilities and it certainly encourages the thought that the basic processes which underlie fluid intelligence may be affected through training of the kind employed here. Crystallized intelligence tasks are not affected by training to a significant extent because both groups attended normal educational activities.

Two-way Repeated Measures ANOCVA

This final set of analyses to be reported here is designed to check on the reasons for a larger number of significant F-tests on the last occasion (i.e., second retest). Two alternative reasons exist. First, it is possible that these differences derive from a genuine increase in the arithmetic means of the experimental group over and above the control group. Second, it is possible that significant differences derive predominantly from the previously mentioned lower variability within the experimental group.

[4]ANOCVA based on a hierarchical design with school classes nested within the groups factor would generate the same results as those presented in Tables 2 and 3.

TABLE 3
Differences between Experimental and Control Groups. ANOCVA with Pretest as Covariate. Units of Analysis: Classes[a]

Variable	F-tests (df = 1,7)		
	Final	1st Retest	2nd Retest
1. Essential features test	5.93*	1.00	4.94
2. Word classification test	3.06	1.45	29.75**
3. Number series test	1.12	4.80	6.63*
4. Consequences test	.87	2.14	.01
5. Relations test	1.26	12.01**	8.00*
6. Word meanings test	5.33	.26	4.60
7. Unbalanced structures test	.09	7.37*	6.79*
8. Numeric test	.60	3.96	3.41
9. Polyprofile test	2.08	1.41	6.19*
10. Meaningful memory test	2.24	1.55	4.38
11. Word classification "S" test	.91	.79	22.15**
12. Proverbs test	1.63	4.27	13.08**
13. Verbal analogies test	.54	5.78*	15.44**
14. Disarranged sentences test	1.90	.89	.78
15. Proverbs interpretation test	5.29	9.65*	2.77
17. Perceptual reasoning test	3.98	3.22	3.05
18. Pictorial unbalanced structures	5.62*	18.99**	37.00**
19. Combined solutions test	17.75**	43.44**	30.31**
20. Multiple solutions test	2.98	20.76**	7.04*
21. Pictorial polyprofile test	1.02	5.32	15.82**
22. Dominos test, D-48	1.99	.26	.57
23. Figure classification test	3.26	14.04**	10.43*
24. Projections in water test	7.30*	3.10	6.98*
25. Figure series test	5.73*	5.39	15.94**
26. Matrices I	3.85	6.20*	3.37
27. Matrices II	5.47	8.83*	7.46*
28. Matrices III	16.71**	20.04**	22.49**

[a]Asterisks indicate significant F-tests: Single asterisk indicates .05 significance; two asterisks indicate .01 significance.

A convenient way to find out about the likely explanation is to employ a two-way repeated measures ANOCVA by treating the final, first, and second retests as occasions. The second factor, of course, would be groups (i.e., experimental and control groups). The analysis to be reported again treats initial testing as covariate. Relevant information is contained in the occasions by group interaction. Thus, if this interaction is significant, there is evidence for an accelerated increase in performance of the experimental group. If this interaction is not significant, previously observed pronounced differences on the second retest (see Table 3) are due to smaller variability of the experimental group on the last testing occasion.

TABLE 4
***F*-tests from the Two-way ANOCVA on Class Means with Initial Test as Covariate**[a]

Variables	Groups $df = 1,7$	Occasions $df = 2,16$	Groups × Occasions $df = 2,16$
1. Essential features test	5.93*	53.11**	3.08
2. Word classification test	12.80**	23.77**	2.77
3. Number series test	6.07*	19.99**	2.20
4. Consequences test	0.01	15.11**	1.15
5. Relations test	6.09*	40.53**	1.05
6. Word meanings test	3.96	54.54**	3.74
7. Unbalanced structures test	5.90*	35.17**	0.13
8. Numeric test	0.76	2.70	0.47
9. Polyprofile test	3.46	41.59**	0.89
10. Meaningful memory test	3.31	62.43**	2.38
11. Word classification "S" test	3.22	22.62**	5.96*
12. Proverbs test	6.19*	90.76**	1.39
13. Verbal analogies test	5.52	63.57**	1.27
14. Disarranged sentences test	1.54	43.47**	0.02
15. Proverbs interpretation test	4.93	32.83**	0.01
16. Arithmetic test	6.71*	42.88**	2.85
17. Perceptual reasoning test	0.68	178.54**	9.99**
18. Pictorial unbalanced structures test	29.70**	29.22**	4.42*
19. Combined solutions test	102.80**	9.11**	1.94
20. Multiple solutions test	12.78**	54.01**	1.44
21. Pictorial polyprofile test	7.68*	50.44**	3.73*
22. Dominos test, D–48	1.27	20.41**	1.10
23. Figure classification test	13.47**	17.55**	0.16
24. Projections in water test	8.60*	18.84**	0.09
25. Figure series test	12.22**	22.16**	0.57
26. Matrices I	14.84**	8.71**	0.18
27. Matrices II	12.01**	5.44*	1.27
28. Matrices III	39.94**	14.63**	0.69

[a]Asterisks indicate significant *F*-tests: Single asterisk indicates .05 significance; two asterisks indicate .01 significance.

Table 4 displays the relevant data. It can be seen from this table that the occasions effect is rather pronounced—probably reflecting both the increased familiarity with the tests themselves and practice effects. It can also be noted that a somewhat larger number of variables show significant *F*-tests for the groups main effect and, again, the effect is more noticeable with the nonverbal fluid intelligence markers. This groups effect is at least in part due to the smaller variance of the second retest in the experimental group. In Table 4, 17 variables show significant differences between the groups at the .05 level and 9 are significant at the .01 level. We see again that the most pronounced differences obtain

TABLE 5
Variance Ratios (*F*-column) and Differences between Experimental and Control Groups (*d*-values) Expressed in Terms of The Standard Deviation in the Initial Session

Variable	F (df = 148, 145)	*d*-values Initial	Final	2nd Retest
1. Essential features test	2.21**	−.38	.25	.32
2. Word classification test	5.07**	−.30	.29	.55
3. Number series test	1.40*	−.74	.24	.54
4. Consequences test	.95	−.53	.00	.15
5. Relations test	1.89**	−.69	−.03	.10
6. Word meanings test	1.40*	−.30	.28	.26
7. Unbalanced structures test	1.45*	−.65	.10	.18
8. Numeric test	1.00	−.24	−.04	.12
9. Polyprofile test	1.47*	−.39	.33	.37
10. Meaningful memory test	.99	.17	.49	.83
11. Word classification "S" test	1.73**	.42	.10	.80
12. Proverbs test	1.88**	−.00	.20	.40
13. Verbal analogies test	1.46*	.11	.14	.28
14. Disarranged sentences test	1.06	−.08	.19	.13
15. Proverbs interpretation test	1.05	.09	.19	.19
16. Arithmetic test	.64	.09	.32	.49
17. Perceptual reasoning test	1.22	−.22	−.24	.20
18. Pictorial unbalanced structures test	2.31**	−.36	.09	.37
19. Combined solutions test	1.46*	.26	.25	.35
20. Multiple solutions test	1.69**	−.29	.16	.20
21. Pictorial polyprofile test	1.71**	−.11	.09	.47
22. Dominos test, D-48	1.14	.42	.52	.38
23. Figure classification test	1.06	−.09	.32	.42
24. Projections in water test	1.78**	−.13	.19	.19
25. Figure series test	.90	−.18	.22	.36
26. Matrices I	1.60**	−.27	.15	.17
27. Matrices II	2.13**	−.60	.13	.31
28. Matrices III	1.55**	.00	.75	.52

[a]Asterisks indicate significant *F*-tests of the variance ratios for the 2nd Retest: Single asterisk indicates .05 significance; two asterisks indicate .01 significance.

among the fluid intelligence tests. There is evidence in this data that the experimental group performs significantly better than the control group on a number of intelligence subtests. Finally, as can be seen from the third column of Table 4, a rather small number of variables show significant interactions (4 at the .05 level). Furthermore, the highest significant interaction (Variable 17—the Perceptual Reasoning test) is due to a crossover, rather than to a gradual increase in the experimental group's superiority.

The majority of variables, therefore, show nonsignificant interaction indicating that significant differences between the groups on the second retest are due largely to a smaller variability in the experimental group. In order to test the significance of the differences between the variances of the experimental and control groups, variance ratio test (McNemar, 1962, p. 247) was employed. The first column of Table 5 provides F-ratios obtained by dividing the experimental group's variance on the second retest by the corresponding control group's variance. It can be seen that 18 F-ratios are significant at the .05 level and that 12 variables are significant at the .01 level. It is also clear that only four variables show the opposite trend—i.e., the control group's variance is lower than the experimental group's variance.

Magnitudes of Group Differences

In many cases, the statistical significance of a finding obtained in the course of psychological research is the most important information for the evaluation of the outcomes. Occasionally, however, it may be useful to have a way to judge the psychological relevance of the obtained result. This is because large samples may lead to a rather small standard error of the mean, so that even a small difference may be statistically significant. In order to obtain an indication of the overall size of the difference, the following simple method was used.

First, standard deviations on all variables for the combined experimental and control groups in initial testing were obtained. Second, for every occasion and for every variable, arithmetic means for control group were subtracted from the arithmetic means for the experimental group. Third, these differences between the means were divided by the standard deviations from the first step. These are the d-values of Table 5. (Note that the d-values for the first retest are not presented in Table 5). The first column of the d-values contains a large number of negative signs which simply reflects the fact that the control group was superior to the experimental group initially. The most noticeable result of Table 5 is the absence of any values greater than one. This means that, for any variable, the difference between the two groups is not greater than one standard deviation of the initial test. It is also interesting to note that for two variables only (variables 3 & 10), the difference between the d-values on the initial testing and on the second retesting is greater than one. These are the only variables thât show an improvement in the experimental group which is greater than one standard deviation of the initial test.

In the fourth step, the obtained d-values were multiplied by 15, that is, the value of standard deviation employed in typical IQ tests. Thus, for every variable there was a number representing differences between the groups in terms of points similar to those of the IQ test. In order to save space, I will not present these values for all variables—arithmetic means over all variables will suffice. Thus, the average difference between control and experimental group over all 28 variables in the initial testing is -2.62 which means that at the beginning of the experiment the control group was 2.62 points better than the experimental group.

For the final testing, the average difference is 3.04 points in favor of the experimental group and, at the second retest, this difference was 5.17 points.

Two different interpretations can be attached to these data. First, initial average difference of −2.62 may reflect sampling error and, because 3.04 from the final testing is not much greater than absolute value of −2.62, the improvement is not greater than the sampling variation. Alternatively, we may calculate the overall improvement of the experimental group: 5.66 points (i.e., 2.62 + 3.04) in the final testing and 7.79 points in the second retesting. As a result, the overall improvement that can be attributed to the experimental manipulation (i.e., training in creative problem solving) is at best around 8 IQ points.

SUMMARY AND DISCUSSION

The main outcomes of the present study can be summarized in the following four points:

1. Training in creative problem solving can produce a small improvement in performance on intelligence tests. This improvement is not significant for a majority of tests and it is statistically significant only for some markers of fluid intelligence.

2. Although small, the effects of training are long lasting—they could be detected 1 year after the exercises had ceased.

These findings should be evaluated in conjunction with the following facts: (a) In general, it has been very hard to increase performance on IQ tests, (b) the sample of this study is not deprived in any reasonable sense of the word—if anything, it is above average on IQ; and (c) general, rather than specific, transfer is being implicated, that is, training was not devised to improve IQ test performance as such. Under these circumstances, even small effects are important from the educational point of view.

3. On the very last testing occasion, the experimental group showed smaller variability than the control group. If replicable, this could be an important finding for the educationists because it suggests that training tends to produce more equal (narrower) distribution of ability. However, in the present study this was not expected and there is no ready-made explanation for the outcome.

4. In the analysis of covariance, different outcomes obtain depending on whether one uses subjects or whole classes as units of analysis. The appropriate criterion for choosing between them is dependent on the nature of the selection process, that is, whatever was used to allocate experimental material to treatment groups on a random basis. It follows that, in this study, we should attach greater significance to the analyses based on class means.

It is necessary to comment here on the outcomes of two ANOCVA procedures that employed class means as units of analysis. Obviously, we can have different impressions of the data depending on whether we consider the results of the "Final" column of Table 4, the other two columns of the same table, or the "Groups" column of Table 5. All these columns provide strong support for the conclusion that fluid intelligence is the most affected and all of them, except the "Final" column, also suggest that some processes of crystallized intelligence may be affected as well. In my opinion, we should consider the "Final" column most seriously, and partly because of the third point mentioned above, attach smaller significance to most other columns. Also, it is hard to believe that some abilities are not affected after 3 years of intensive training (at the "Final" testing), and then, all of a sudden, show improvement within 1 year of no systematic training. It follows that the "Final" column reflects the real effects of training and all other columns reflect something that must have happened after the training ceased.

Even though it was expected that exercises will affect crystallized intelligence, these effects are obviously not pronounced in the data. This is probably because both the experimental and control groups were exposed to pretty much the same formal educational experiences and the exercises in creative problem solving were not designed for the purpose of inculcating knowledge. The main purpose of the exercises was to develop general cognitive schemas that could be employed in a variety of situations involving problem-solving activity. These schemas are obviously effective in helping to improve performance on at least some fluid intelligence tasks.

From one point of view, at least, we have a good reason to advocate the use of Kvashchev's exercises in our schools. Even though the effects are small, and the amount of training and effort employed in his work was far greater than what is typically found in similar studies, if we were to encourage teachers to emphasize the importance of these principles in working through cognitive tasks, society will benefit from it in the long run. Because there is no evidence that Kvashchev's actions were detrimental in any way, this view stems from the attitude that even a small improvement in this area is better than the status quo.

From another point of view, the increase was too small to warrant further steps along this path. Perhaps we should concentrate our energies in other directions. But what directions? One suggestion may be to proceed with the attempts to modify cognitive skills by focusing on particular tasks and by training people in executive processes related to working on these specific tasks. This line of attack was suggested by the contributors to Detterman and Sternberg's (1982) volume. My suspicion is that when these efforts are introduced on the same scale as Kvashchev had implemented his ideas, the outcomes will be similar—intelligence may change very little. But then, maybe, I am wrong. From this perspective the best policy would be to keep implementing the ideas of creative problem solving, while proceeding with the search for further and possibly more effective cognitive training procedures.

REFERENCES

Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist, Special Issue, 36* (10), 1086–1093.

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.* Norwood, NJ: Ablex.

Caruso, D.R., Taylor, J.J., & Detterman, D.K. (1982). Intelligence research and intelligent policy. In D.K. Detterman & R.J. Sternberg (Eds.), *How and how much can intelligence be increased.* Norwood, NJ: Ablex.

Cattell, R.B. (1971). *Abilities, their structure, growth and action.* Boston: Houghton Mifflin.

Detterman, D.K., & Sternberg, R.J., (Eds.). (1982). *How and how much can intelligence be increased.* Norwood, N.J.: Ablex.

Duncker, K. (1945). *On problem solving.* Washington, DC: Psychological Association.

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.* Norwood, NJ: Albex.

Guilford, J.P. (1967). *The nature of human intelligence.* New York: McGraw-Hill.

Guilford, J.P., & Hoepfner, R. (1971). *The analysis of intelligence.* New York: McGraw-Hill.

Horn, J.L. (1981). Concept of learning in relation to learning and adult development. *Intelligence, 4,* 285–317.

Horn, J.L. (1985). Remodeling old models of intelligence. In B.B. Wollman (Ed.), *Handbook of Intelligence.* New York: Wiley.

Kvashchev, R. (1980). *Mogucnosti i granice razvoja inteligencije* [The feasibility and limits of intelligence training]. Belgrade: Nolit.

Maier, N. (1945). Reasoning in humans III: The mechanism of equivalent stimuli and of reasoning. *Journal of Experimental Psychology, 35,* 349–360.

Maier, N., Julius, M., & Thurber, J. (1967). Studies in creativity: Individual differences in the storing and utilization of information. *The American Journal of Psychology, 80,* 492–519.

Maltzman, J. (1960). On the training of originality. *Psychological Review, 67,* 229–242.

McNemar, Q. (1962). *Psychological Statistics* (3rd ed.). New York: John Wiley.

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review, 69,* 220 232.

Nie, N.H. (1985). *SPSS^X user's guide.* New York. McGraw-Hill.

Reichardt, C.S. (1979). The statistical analysis of data from nonequivalent group designs. In T.D. Cook & D.T. Campbell (Eds.), *Quasi-experimentation: Design & analysis issues for field settings.* Chicago: Rand McNally.

Stankov, L., Horn, J.L., & Roy, T. (1980). On the relationship between Gf/Gc theory and Jensen's Level I/Level II theory. *Journal of Educational Psychology, 72* (6), 796–809.

Sternberg, R.J., Ketron, J.L., & Powell, J.S. (1982). Componential approaches to the training of intelligent performance. In D.K. Detterman & R.J. Sternberg (Eds.), *How and how much can intelligence be increased.* Norwood, NJ: Ablex.

Wertheimer, M. (1959). *Productive thinking.* New York: Harper.