

The Transfer Value of the Course in Logic,

by

John W. McGarvey

Submitted in partial fulfillment of the
requirements for the degree of


Master of Arts

in the College of Letters, Arts, and Sciences, of the

University of Arizona

1 9 3 4

Approved: _____


Major adviser

5-10-'34

Date

Table of Contents.

- I. Introduction.
- II. Statement of the Problem.
- III. History of the Problem.
- IV. Materials.
- V. Method.
 - 1. Subjects.
 - 2. Procedure.
 - 3. Scoring.
 - 4. Reliability.
- VI. Results and Discussion.
- VII. Summary.
- VIII. Conclusions.

I. Introduction.

The general problem of transfer of training is of considerable significance to the psychologist and teacher. It is important to the psychologist because of the theoretical interest in the psychology of learning, and to the teacher because of his practical interest in the transfer value of all curricular subjects. The term "transfer of training" means, according to Hunter (1), "...the effect of training upon the subsequent formation of habits, either "habit transfer," or "habit interference." In other words, transfer of training may be positive or negative or zero. The problem of transfer of training has been stated by James (2):

"Transfer of training refers to those changes (if any) in the facilitation and in the constituents and organization of an ability which result, not from direct training of that ability, but from some other ability."

Thus it may be seen that, among the ramifications of the problem of transfer, its importance for education is fundamental: without transfer of the training acquired in school to the subsequent behavior of the student, education would be to a large extent a futile process. Again, the efficacy of education will vary largely in proportion to the amount of trans-

(1). Hunter, W. S., in Foundations of Experimental Psychology (C. Murchison, ed.), 1929, 611.

(2). James, H. E. O. The transfer of training, Brit. J. Psychol., 1929, 20, 322.

fer which is made functional in the student's later life. It thus seems justifiable to attempt to develop a technique by which the transfer value of at least a few of the courses in the curriculum may be measured.

Concluding a review of the status of the problem of transfer, Hunter (1) states:

"We are still far from a satisfactory explanation of the facts of transfer. At present the theory of identical elements seems most adequate, and yet it is all but impossible to bring the theory to a rigorous experimental test."

Briefly stated, the theory of identical elements holds that transfer will take place between two abilities in proportion as the second ability contains elements which are identical with the elements contained in the first ability.

In his abstract, "The Acquisition of Skill, "McGeoch (2) concludes: "Clearly, an adequate theory [of transfer] waits on crucial experiments." The present study is not an attempt to perform a crucial experiment for a theory of transfer. Rather, it is concerned with the amount of transfer which occurs between the training acquired by students in the study of logic and the solution of non-academic problems.

(1). Op. cit., p. 614.

(2). McGeoch, J. A. Psychol. Bull., 1931, 23, 433.

II. History of the Problem.

A large part of the experimentation on transfer of training is based on peripheral or motor skills. An important study was made by Ewart (1) on mirror drawing. The subjects in the experimental group traced with a stylus a star-shaped diagram, which was seen via a mirror. One trial was made with the non-preferred hand, followed by fifty trials with the preferred hand; then a second trial was made with the non-preferred hand. The control group followed the same procedure, but substituted an hour's rest in place of the fifty trials with the preferred hand. Ewart's results showed that on the second trials with the non-preferred hand, the experimental subjects made 21% fewer errors and required 36% less time to trace the star, than did the control subjects, who had had no training of the preferred hand.

Of the studies of transfer of skills more strictly central in character, but few are germane to the present experiment.

Gates (2) shows that in the case of children, a part of transfer, as measured by tests, is due to increased familiar-

-
- (1). Ewart, P. H. Bilateral transfer in mirror drawing, Ped. Seminar, 1926, 33, 253-249.
 - (2). Gates, A. I. A critique of methods of estimating and measuring transfer of training, J. Educ. Psychol., 1924, 15, 545-558.

ity with and facility of taking the tests. He concludes that psychologists can judge fairly well the relative, but not absolute, amounts of transfer. That is to say, a relative comparison can be made between groups similar except in training in the particular ability under observation.

Using high school students as subjects, Nichols (1) put them through a number of tests of observation, comparison, enumeration and generalization. She found that previous training and emphasis in scientific method materially strengthened such abilities. Even the less intelligent students showed some improvement after training in scientific method. Nichols lays emphasis on the necessity of teaching science as a method as well as a system of facts, to obtain the maximum of transfer.

In a study of the efficacy of methods of teaching capitalization and punctuation, Leonard (2) found that special practice in proof reading, error correction and dictation, gave a clear advantage to the students in the experimental group as compared with the students not receiving such practice. Simpson (3) concludes that special training in evaluat-

-
- (1). Nichols, M. L. The high school and scientific method, J. Educ. Psychol., 1929, 20, 196-204.
 - (2). Leonard, J. P. The use of practice exercises in teaching capitalization and punctuation, J. Educ. Research, 1930, 21, 186-190.
 - (3). Simpson, R. G. The effect of specific training on ability to read historical materials, J. Educ. Psychol., 1929, 20, 343-351.

ing, outlining and summarizing historical materials, aids the ability to organize, in the study of history. Dorsey and Hopkins (1) showed that methods of study, knowledge of Latin, and skill in manipulation of special elements in descriptive geometry, transferred definitely to new situations - when the subjects were instructed to use the knowledge gained from the original material, in handling the new material. The control groups, not instructed to use the original knowledge and methods, showed less facility in meeting the new situations.

Studying the teaching of grammar, Symonds (2) discovered that the learning of correct English usage was more efficient when the methods of grammar were used, such as rules, definitions, and the analysis of grammatical constructions than when the training consisted of repeating correct forms. It was observed by Courtis (3), in a study of spelling ability, that such ability improved for words which were not taught, apparently as result of extrinsic, unspecified experience. Britt (4), in an experiment on the relation between the age and the strength of associations, concludes:

-
- (1). Dorsey, M. F., and Hopkins, T. The influence of attitude on transfer, J. Educ. Psychol., 1930, 21, 410-417.
 - (2). Symonds, P. M. Practice versus grammar in the learning of correct English usage, J. Educ. Psychol., 1931, 22, 81-95.
 - (3). Courtis, S. A. The measurement of the effect of teaching, School and Society, 1928, 28, 52, 54.
 - (4). Britt, S. H. The relationship between transfer of learning and age of previous associations, Psychol. Bull., 1933, 30, 543.

"If the associations for a learning problem are of the same functional strength but are of unequal age, the older association is more subject to positive transfer to a second problem than is the younger association."

In a later paper (1), Britt reports:

"From a theoretical standpoint, this conclusion does not imply that the amount of positive transfer is caused by the age of the association."

According to McGeoch (2), "Yum's test of the law of assimilation is a basic experiment on transfer." Yum (3) used double syllables paired with words, and, at recall, changed certain of the letters in the stimulus syllables. Yum also used pairs of meaningful words, the substitute stimuli being words of two different degrees of similarity to the original stimulus words. The results from these two sets of materials were verified in a similar experiment, in which visual patterns were used instead of words. Yum found that all alterations of the stimulus syllables, words, and figures reduced the amount recalled.

There are three experimental studies pertinent to the present problem, but which are not concerned with the transfer of training in that no special previous training is involved.

-
- (1). Britt, S. H. The relationship between transfer of learning, and age of previous associations, Amer. J. Psychol., 1934, 41, 113-116.
 - (2). Op. cit., p. 431.
 - (3). Yum, K. S. An experimental test of the law of assimilation, J. Exp. Psychol., 1931, 14, 68-82.

Bailor (1) investigated the effect of varying the content and form of tests of intelligence. Batteries of tests were given to 1039 school children, and the tests were repeated a year later. The effect of changing the content was measured by taking correlations between the individual standings in test groups in which the form was unchanged but the content varied. For instance, correlations were taken between the test results in grammatical, spatial and verbal analogies. Bailor concludes:

"the correlations throughout are positive, and that differences in the relative standings of pupils occur when they are given tests having differences either in form or content."

Thorndike (2), in a study of the effect of changed data on reasoning, had 97 graduate students solve nine simple problems in algebra. The same problems were then expressed in less usual ways, the amount of change varying. Thorndike's conclusion follows:

"...any disturbance whatsoever in the concrete particulars reasoned about will interfere somewhat with reasoning, making it less correct or slower or both."

An investigation by Wilkins (3) determined the effect of changed material on ability to do formal syllogistic logic.

-
- (1). Bailor, E. M. Content and form in tests of intelligence, Columbia Univ. Contrib. to Educ., 1924, No. 162.
 - (2). Thorndike, E. L. The effect of changed data on reasoning, J. Exper. Psychol., 1922.
 - (3). Wilkins, M. C. The effect of changed material on the ability to do formal syllogistic reasoning, Arch. Psychol., 1928, 16, (No. 102), 1-83.

The subjects represented a cross-section of the undergraduate body at Columbia University. The materials were sixty syllogisms, each expressed in four ways: familiar terms; symbolic terms such as a, b, x, and y; suggestive terms, familiar and concrete material with the meaning designed to be misleading; and unfamiliar terms, such as long words and specially coined scientific words, intended to be difficult. All four forms of the syllogism test had reliabilities above .900. The suggestive form was least reliable. This is accounted for by Wilkins on the ground that there is no means of measuring the relative suggestibility of the different test items; that the subjects might suddenly come on guard against suggestion; and that suggestibility varies with the individual. Wilkins concludes:

"Ability to do formal syllogistic reasoning is much affected by a change in the material reasoned about. The easiest material is the familiar and concrete... The suggestive material is more difficult than the familiar but not so difficult as the symbolic and unfamiliar."

III. Statement of Problem.

The present study is a preliminary attempt to discover the degree to which the training acquired by students in the course in logic is transferred to the solution of some questions of a more or less controversial nature which are met in everyday life and which involve choice and reasoning.

From a pedagogical standpoint, apart from purely theoretical considerations for psychology, we are interested in determining the value of the study of logic to the average student, insofar as such training enables him better to meet ordinary lay problems.

IV. Materials.

In order to measure the transfer of training which takes place between the study of logic and the solution of everyday problems, an achievement test was constructed. The majority of the items in this test were chosen because of their current interest and popular nature; the other items, mainly as indicators of logical ability, apart from suggestibility. Most of the test items were selected from passages in newspaper and magazine editorials, articles and advertisements; some were taken from practice exercises in logic text books. Several of the items concerned political, economic and social issues of a controversial character.

Two hundred items were submitted to a trained logician, who eliminated those least suitable. Fifty items were selected for a tentative test, which was taken by three senior students majoring in psychology. In the light of the criticisms of these three students, the test was again revised and the number of items reduced to 32, this being the number which it was anticipated could be solved by most of the students within the 45-minute time limit. The time limit was set at 45 minutes in order that the test might be administered during a regular class period.

Each item consisted of a quotation, which was to be evaluated by the student as "good" or "bad," according to whether

he thought the logic involved in the quotation was valid or invalid.' The subject was then instructed to check one of four reasons which were offered in support of the judgment "good" or "bad." Thus, there were eight possible answers to each question. It was believed that, since the chances of a guessed answer's being correct were only one in eight, guessing would be discouraged and the reliability of the test would thereby be increased. Half of the 32 questions were valid in their logic, and half were invalid. This provision was made in order to eliminate the possibility of the students' obtaining a clue as to the correct answers.

The test items were not equated into units of equally increasing difficulty, nor were they based on a point of zero difficulty, because of the exploratory nature of the entire experiment. The construction of a test meeting such requirements, plus that of validity, would require considerably more time than was available for the present study. The advisability of devising such a test would depend largely on the results of an exploratory study such as the present one.

The reasoning test devised for this study follows.

Directions: In each question below, a statement is made which is either logical or illogical. The procedure is to read the statement, and if you consider it logical, draw a line under the word Good. If you consider the statement illogical, draw a line under the word Bad. Beneath the words Good or Bad, there are four reasons offered as to why the statement is Good or Bad. Read these reasons, and draw a line under a, b, c, or d, according to which one you think best proves that your judgment is Good or Bad.

Read the following example:

"Only the choicest ingredients have ever gone into Blank's shaving cream - that's the reason for its cool, stay-moist lather."

Good, Bad - because :

- a. Blank's contains the best materials and must be the best.
- b. Shaving sticks cost less than shaving creams.
- c. Choice ingredients might be combined badly.
- d. The proof of the pudding is in the eating.

The statement is marked Bad because it is illogical. The best reason for its being Bad is c, and so reason c is underlined.

When the signal is given, proceed with the test, and work quietly and steadily until you have finished the test, Please wait until the end of the hour to hand in your paper.

1. "The claim is made that since intelligent persons sometimes become insane, this man, who is not intelligent, is in no danger of insanity. This claim is not justified."

Good, Bad - because:

- a. Not all intelligent persons become insane.
- b. Genius is not far removed from insanity.
- c. Insanity may not be confined to intelligent people.
- d. Morons frequently are feeble-minded.

2. "This poor fellow was very wrong in refusing to have a doctor simply because all of his acquaintances who died had doctors."

Good, Bad - because:

- a. Death comes to everybody.
- b. Doctors are sometimes powerless.
- c. Some doctors are unethical.
- d. A doctor might have cured this case.

3. "Education is futile because one is always hearing of fraud or forgery committed by someone who might have led an honest life if he had never learned to read or write."

Good, Bad - because:

- a. Education does little good considering the expense to the state.
- b. Educated criminals might have been dishonest without education.
- c. A clever thief might be more dangerous than a stupid one.
- d. Some fine people have been illiterate.

4. "X could hardly be expected to know the customs of polite society, educated as he was among savages."

Good, Bad - because:

- a. Well-bred men are innately polite.
- b. One can't be familiar with what he has never experienced.
- c. Civilized manners are only a veneer on the surface.
- d. Savages are often more happy than highly civilized persons.

5. "Had Jesus lived in 1917 he would have been the first to volunteer in the American Army, to wear a gas mask, and shoulder a rifle."

Good, Bad - because:

- a. God was on the side of America.
- b. One should fight for his country right or wrong.
- c. Jesus would have enlisted in the German Army.
- d. This is simply a patriotic assertion, without proof.

6. "Though you condemn a man's religion as superstition, you do not prove him any the less religious."

Good, Bad - because:

- a. What is superstition to one man may be real religion to another.
- b. There is no conflict between science and religion.
- c. It is dangerous to blaspheme by condemning religion.
- d. Religion is not all superstition.

7. "A Soviet flight into the stratosphere was called off because the balloon failed to ascend. This explains the failure of the engineers who designed the Russian tractors."

Good, Bad - because:

- a. The balloon must have been designed by poor engineers.
- b. Balloons fail to ascend in countries besides Russia.
- c. Communism isn't likely to succeed.
- d. Stratosphere flights don't prove much.

8. "Law may restrict liberty but not necessarily happiness."

Good, Bad - because:

- a. Any restriction of liberty reduces happiness.
- b. The happiest people are those obedient to the will of the state.
- c. Happiness is less important than law and order.
- d. Restriction of some liberty may increase the greatest happiness of the greatest number.

9. "Hospitals and doctors use liquid laxatives and always have. That's the reason that Dr. Blank's syrup pepsin is the ideal family laxative."

Good, Bad - because:

- a. Doctors and hospitals use liquids so they must be best.
- b. It is easier to measure liquids than solids.
- c. Children should be treated differently.
- d. Hospitals and doctors may use non-liquid laxatives also.

10. "If a man uses bribery in an election he deserves defeat, but that doesn't justify your saying that X deserved to win because he didn't use bribery."

- Good, Bad - because:

- a. Only honest men deserve election.
- b. All candidates use bribery in some form.
- c. X may not have used bribery.
- d. A man may be honest and yet lack ability.

11. "The positive action of Blank Tonic is demonstrated by modern scientific methods, and this is your guarantee of satisfaction."

Good, Bad - because:

- a. The guarantee should be in writing.
- b. The tonic may have positive action and yet be unsatisfactory.
- c. The tonic may not be good for everybody.
- d. The action is demonstrated by scientific methods and must be good.

12. "No philosophy can be taken as final, because scientists are continually bringing in new and strange facts about the world and man."

Good, Bad - because:

- a. Philosophy is only theoretical.
- b. A final philosophy would deal with all the facts of the universe.
- c. Science can't deal with spiritual facts.
- d. No philosophy claims to be the final or universal one.

13. "After a two-day search for the body failed, the widow put a crucifix on a piece of wood and placed it in the river. The frail craft came to rest over the spot where the body was found. Thus God's power is again proven."

Good, Bad - because:

- a. Divine power aided the search.
- b. The crucifix should have been used immediately.
- c. The crucifix and body may have stopped there by natural causes.
- d. If the husband were a good man, his life wouldn't have been taken.

14. "Motion picture stars are paid their \$5,000 or so a week because the directors find by experience that their popularity is worth this sum."

Good, Bad - because:

- a. Even the President of the U. S. isn't paid \$5,000 a week.
- b. If their popularity is profitable, the stars deserve their share.

15.

- c. Not all stars are paid \$5,000 a week.
- d. Some stars have 'pull' with the directors.

15. "The American Legion came out unanimously for a sound dollar. Now that proves that they weren't taking the easy way out."

Good, Bad - because:

- a. An unsound dollar would mean ruin.
- b. The Legion favors the best interest of the country at large.
- c. The Legion wants the presumptive clause restored.
- d. To favor a sound dollar may indicate no more than habit of thought.

16. "If T. had studied he would have been able to recite. But you can't say that since he recited he must also have studied."

Good, Bad - because:

- a. Some students get along without ever studying.
- b. Studying can be carried too far.
- c. One can sometimes recite without having studied.
- d. Some students are nervous and can't recite in class.

17. "When both the public and the experts agree, the verdict is certain. That's why Blank's is the best of the better beers."

Good, Bad - because:

- a. Perhaps the experts were paid.
- b. Brands of beer aren't classified as better or poorer.
- c. Since experts agree on Blank's, it is best.
- d. People may agree and yet all of them may be wrong.

18. "Since they buy millions of dollars worth of advertising space, the patent medicine makers have great influence with newspapers and magazines."

Good, Bad - because:

- a. Patent medicine advertising is dangerous and should be stopped.
- b. Physicians often prescribe patent medicines.
- c. Self-diagnosis and medication is the privilege of anyone.
- d. Any business generally tries to please good customers.

19. "A. M. Wilkins, air-mail ace, says: 'It's a steady grind, flying the mail. That's why I smoke Blanks - they never jangle my nerves.'"

Good, Bad - because:

- a. An air-mail pilot must have steady nerves.
- b. It isn't shown that other cigarets would jangle Wilkin's nerves.
- c. Some air-mail pilots don't smoke cigarets at all.
- d. Moderate smoking probably doesn't affect the nerves

harmfully.

20. "The head of the corset department in a famous store says: 'Blank's Flakes is the soap we advise. Careful observation of all brands shows us that Blank's keeps silk and elastic strong longest.'"

Good, Bad - because:

- a. Soap is more expensive in flakes than in bars.
- b. Soap flakes are all alike.
- c. Blank's may contain strong alkalies which are dangerous.
- d. Careful observation of all brands is a fair basis of judging.

21. "The Virgin Mary was responsible for Wiley Post's flight around the world. Wiley began and ended his flight on Saturdays, and Saturday is dedicated to the Virgin Mary."

Good, Bad - because:

- a. Wiley was lucky.
- b. Wiley flew on Sunday and violated the Sabbath.
- c. Honoring the Virgin Mary undoubtedly helped Wiley.
- d. Wiley might not have failed, starting on another day.

22. "He might not have been a Democrat even though all Democrats used to believe in Free Trade."

Good, Bad - because:

- a. Free trade would ruin American industry.
- b. One could advocate Free Trade without being a Democrat.
- c. World-wide free trade would enrich everyone concerned.
- d. This man must have been a Democrat to favor free trade.

23. "Recognition of Russia will not help the U. S. because nations which already trade with her are still in bad condition."

Good, Bad - because:

- a. Such nations might be in worse condition without Russian trade.
- b. Recognition will lead to revolution.
- c. Prosperity will be restored by recognizing Russia.
- d. Recognition of Russia is a betrayal of American ideals.

24. "Professors sometimes fail as politicians because their ideas are theoretical only and not burned in the flame of practical politics."

Good, Bad - because:

- a. Professors should stay in the class room.
- b. Government should return to basic principles.
- c. Professors often know more than party politicians.
- d. Inexperienced men may fail where experience is necessary.

25. "Bernard Shaw says the saving of the world lies in Commun-

ism.. Wonder to us that Shaw is permitted to run abroad. The man is a maniac."

Good, Bad - because:

- a. Those who favor Communism must be mentally unbalanced.
- b. The world can be saved only by Communism.
- c. Communists should be imprisoned.
- d. To favor Communism may not prove that one is crazy.

26. "Saying that one lynching is worse than another is at best splitting hairs."

Good, Bad - because:

- a. Some lynchings are fair.
- b. Lynching is necessary to keep the negro in his proper place.
- c. Every offender is entitled to a fair trial before being punished.
- d. Juries often give unjust decisions.

27. "Plagues and wars have killed millions of people down through the ages, and thus has 'Mother Nature' limited population better than birth control ever could."

Good, Bad - because:

- a. Birth control is morally wrong and is a sin.
- b. Plagues and wars do reduce population somewhat.
- c. Medical science has reduced the death rate.
- d. Plagues and wars take no account of the quality of people destroyed; birth control can do this better.

28. "These rugged, brilliant Blank cars appeared on college campuses almost over-night. Because Blank is all automobile, that's the reason."

Good, Bad - because:

- a. "Almost overnight" is indefinite.
- b. Students are often good judges of cars.
- c. Blank cars aren't shown to have appeared on all campuses.
- d. Students are no better judges of cars than other people.

29. "True, if a person is avaricious he will be unhappy. But you can't say that because L. is unhappy she is also avaricious."

Good, Bad - because:

- a. Unhappiness may be due to other causes than avarice.
- b. Rich people are often unhappy.
- c. Women are not always as grasping as men.
- d. Avarice is not always for money.

30. "The families of the radio stars can't be fooled about radio reception; for they're in a position to know. That's why the Blank radio is found in the homes of nine out of ten of the biggest stars in radio."

Good, Bad - because:

- a. A new radio is obsolete within a few months.
- b. The stars' families know their voices best.
- c. The stars' families may be no more competent than others.
- d. Liking for different radios is a matter of individual taste.

31. "Traitors are not to be trusted, but the fact that Colonel X is not one of the traitors isn't proof that he can be trusted."

Good, Bad - because:

- a. That X is not a traitor is in his favor.
- b. X might be dishonest without being a traitor.
- c. A traitor may betray and ruin his country.
- d. Benedict Arnold was not to be trusted.

32. "It is possible for other cities to be crime-free as Milwaukee is, because the people in Milwaukee can't be constituted very differently from other people."

Good, Bad - because:

- a. Gangsters are protected by politicians.
- b. People of the same ability may accomplish the same things.
- c. Milwaukee is largely Socialist and has a Socialist mayor.
- d. Other cities have tried to eliminate crime but failed.

V. Method.

Subjects. The subjects used in this experiment were one hundred fifty undergraduate students in the University of Arizona. The experimental group was the class in logic, numbering twenty-six students. The class in elementary psychology, one hundred twenty-four students, served as the control group.

It was intended to pair the experimental and control subjects according to age, sex, intelligence, and, as far as possible, training in science and philosophy preceding the time of the experiment. This requirement could not be fulfilled, however, because, contrary to past experience, two-thirds of the students in the logic class happened to be juniors and seniors, while the psychology class was about evenly divided between freshmen and sophomores. Another difficulty was the lack of intelligence scores for over half the subjects, who were transferred to Arizona from other universities or failed to take the freshman intelligence test when entering the university. The alternative taken was to base comparisons on the gross scores of the experimental and control groups.

Correlations taken between the scores on the reasoning test and the Thurstone Psychological Examination for High School Students and College Freshmen, for those students whose intelligence scores were available (forty-four students), shows a coefficient of .617.

Procedure. The test was mimeographed, each copy consisting of four standard size pages, clipped together. The time limit for taking the test was 45 minutes, in order that the directions could be read and the test administered during a class period 50 minutes in length. The test was administered to the experimental and control groups on the same day, during the first week of the semester. It was re-administered nine weeks later to both groups at the same time. Between the two tests, the logic class had completed the study of deductive logic. No attempt was made to prevent the students from talking among themselves about the test after they had taken it the first time, for the reason that it was considered that the memory factor would be negligible after an interval of nine weeks. The tests were administered to both groups at the same hour on the same day, to prevent any possible conferring between the students in the experimental and control groups.

Scoring. The tests were scored in terms of the number of questions correctly answered; it was unnecessary to score in terms of number of questions correct divided by the number attempted, as only two subjects failed to answer all the questions.

Reliability. The reliability of the test was computed first by using the split test method, and using the Spearman-Brown formula to obtain the reliability of the full length test. The odd-numbered questions correctly answered were correlated

with the even-numbered questions correct. This result was checked by correlating the questions incorrectly answered. After the test had been administered the second time, the reliability was computed by the test and re-test method, using the control group only.

VI. Results and Discussion.

In Table I the reliability computations of the reasoning test are presented:

Table I.

Reliability of the Reasoning Test.

Method.	r.
Split test (questions correct)	.685
Split test (questions incorrect)	.674
Test and re-test	.695

The reliability of the reasoning test is extremely low, due to several factors. As pointed out by Wilkins, test items which contain an element of suggestibility are likely to be unreliable because: there is no means of measuring the relative suggestibility of the different items; the subjects might suddenly come on guard against suggestion; and suggestibility varies with the individual.

Furthermore, the reasoning test items are not equated in a scale of difficulty nor based on a point of zero difficulty. It was found that most of the students finished the test within half an hour; thus the test could have contained a greater number of items and still have been performed within a time limit of 45 minutes. The average number of test items marked incorrectly by the control group at the first administration of the test was 8.57. The average number of questions missed by the experimental group was 4.40. It is obvious that many of the

test items were too easy, and the majority of the questions were of no diagnostic value and might have been omitted.

In Table II are shown the results of the first administration of the reasoning test to the experimental and control groups.

Table II.

Results of First Test Administration.

	N. (subjects)	Mean (errors)	Range (errors)	S.D.	S.D.av.
Control group	98	8.47	0-19	3.75	.399
Experimental group	20	4.40	0-11	2.82	.530

It is quite evident from an examination of the mean scores in Table II that the experimental and control groups are not equivalent; the logic class appears superior at the outset. However, this superiority may be due to chance. To determine whether this difference was real or apparent, the coefficient of reliability was computed. The reliability of the difference between the two means was calculated by means of the following formula (1):

$$S.D.(diff) = \sqrt{S.D.m_1^2 + S.D.m_2^2}$$

The S.D. of the difference, obtained by use of this formula, is .655. To determine the chances in one hundred that the

(1) Garrett, H. E. Statistics in Psychology and Education, p. 129.

obtained difference is also a true difference, the actual difference, 4.07 errors on the reasoning test, was divided by .655, the S.D. of the difference. The result obtained was 6.21. By referring to Garrett's table showing the chances for a true difference greater than zero (1), it was established that the chances for a true difference are one hundred in one hundred. This indicates extremely high reliability.

It is possible that a control group equivalent to the experimental group would have improved more than did the present control group; or, an experimental group equivalent to the present control group might have improved less. Due to absences and changing of courses, only ninety-eight students in the control group took the test and re-test, and only twenty in the experimental group took both test and re-test. The experimental group is thus perilously small.

The results of the second application of the test are shown in Table III.

Table III.

Results of the Second Testing.

	N.	Mean.	Range.	S.D.	S.D.av.
Control group.	98	7.38	1-22	3.54	.357
Experimental group.	20	2.90	1-10	2.44	.545

The reliability of the difference between the means of the experimental and control groups shown in Table III were

(1). Garrett, H. E. Op. cit., p. 134.

computed by means of the same formula used on the results of the first test. The formula yielded S.D. of the difference between the two means, of .651. The actual difference, 4.43 errors, when divided by .651 gave a quotient of 6.89. By reference again to Garrett's table, as explained above, the chances for a true difference were found to be one hundred in one hundred.

In Table IV a comparison is made between the results of the first and second testings for both groups.

Table IV.

Comparison of First and Second Testings.

	Control (1st test)	Control (2nd test)	Exp. (1st test)	Exp. (2nd test)
Mean score	8.57	7.38	4.4	2.9
S.D.	3.57	3.54	2.37	2.44
S.D. _{av.}	.399	.357	.530	.545

The percentage of improvement for the control group, computed by dividing the difference between the mean scores for the first and second tests by the mean score of the first test, was found to be 13.8%. The percentage of improvement for the experimental group, computed in the same manner, was 34.0%.

The S.D. of the difference between the two means of the control group was .534. The obtained difference divided by the S.D. of the difference, was 2.2. The chances for a true difference between the two means of the control group, taken from Garrett's table mentioned above, were ninety-nine in one

hundred. For the experimental group, the S.D. of the difference between the means for the first and second testings was ,760. The obtained difference between the two means was 1.5 errors, which, divided by the S.D. of the difference, was 1.97. Referring again to Garrett's table, the chances for a true difference were found to be ninety-eight in one hundred.

It is thus seen that the experimental group shows a greater improvement on the second test than does the control group. In both cases, however, the improvements are unreliable, and the improvement of the experimental group is less reliable than that of the control group.

Attributing part of the improvement of both groups to practice in taking the test, the advantage, though less reliable, appears still to be with the experimental group.

VII. Summary.

The present paper describes a preliminary attempt to measure the amount of transfer of training which takes place between the study of deductive logic and the solution of some ordinary problems of everyday life.

To measure the transfer of logical training, a test of logical reasoning ability was devised. Each item in this test consisted of a quoted statement, which was to be evaluated by the student as either logical or illogical. The logical validity of each quotation was judged beforehand by a competent logician. Half the test items contained logical statements, and the other half contained illogical statements. The student, after judging a quotation as logical or illogical, checked one of four reasons which were offered in support of his judgment.

Most of the quotations used concerned political, social or economic issues of somewhat controversial nature. The thinking of the average layman about these issues is especially apt to be rather confused, or perhaps based on prejudice.

The coefficient of reliability of the reasoning test was about .68, varying slightly with the manner of computation. This reliability was low, but was deemed sufficient considering the exploratory purpose of this experiment. The test has a coefficient of correlation of .617 with the Thurstone Psychological Examination for College Freshmen.

The experimental group consisted of twenty students in the course in logic. This group was too small to give results of high statistical reliability, but a larger group was not available. Ninety-eight students in elementary psychology served as the control group. The reasoning test was administered simultaneously to experimental and control groups during the first week of the semester. Nine weeks later, after the logic class had completed the study of deductive logic, the test was again administered simultaneously to both experimental and control groups.

The results show that the experimental and control groups were not equivalent. It was impossible to pair the subjects in the two groups according to age, sex, intelligence, etc.

The improvement of the experimental group, as shown on the second test, was 34%. The improvement of the control group was 13.8%. In both cases, however, the improvement was too small for complete statistical reliability. While the experimental group showed the greatest amount of transfer, the improvement has less statistical reliability.

VIII. Conclusions.

1. There seems to be a moderate but unreliable degree of transfer from the study of deductive logic to the ability to solve everyday questions of a non-technical nature as represented in the reasoning test here employed.

2. As a corollary to the above conclusion, the study of deductive logic appears to be of moderate practical value to students, this conclusion being subject to the above qualifications.

3. It is evident that the students electing the logic course are somewhat more logical reasoners than the students in elementary psychology who acted as the control group, using the present reasoning test as the criterion.

4. The reasoning test herein described has a coefficient of correlation of .617 with the Thurstone Psychological Examination for High School Graduates and College Freshmen.

5. In spite of its defects, the reasoning test has some reliability. The results of this experiment, though inconclusive, seem to warrant the effort necessary to reconstruct the test, eliminating the above-mentioned defects, and a repetition of the experiment using a larger experimental group equated with a more suitable control group.

It is conceivable that a technique applied thus crudely in determining the transfer value and practical utility of

logic might, after the requisite refinement and adaptation, be applied in ascertaining the transfer (and hence, practical) value of other courses in the curricula of colleges and perhaps high schools. Proposed new courses for these curricula might be evaluated by variations of the present technique.

References:

1. Asker, W. The Reliability of Tests Requiring Alternative Responses, J. Educ. Res., 1924, 9, 234-240.
2. Brinkmeier, J. H. Sentence Length as a Specific Determiner in True and False Statements, J. Educ. Res., 1930, 22, 203-205.
3. Cuff, N. B. Scoring Objective Tests, J. Educ. Psychol., 1932, 23, 681-686.
4. Freeman, F. N. Mental Tests, 1926.
5. Garrett, H. E., and Schneck, M. R. Psychological Tests, Methods and Results, 1933.
6. McCall, W. A. How to Measure in Education, 1922.
7. Pintner, Rudolf. Intelligence Testing: Methods and Results, 1923.
8. Thorndike, E. L. The Measurement of Intelligence, 1927.
9. Turney, A. H. The Concept of Validity in Mental and Achievement Testing, J. Educ. Psychol., 1934, 25, 81-95.
10. Whipple, G. M. The Transfer of Training, Nat'l. Soc., Study of Ed., 1928, 27, 179-209.
11. Wood, Ben. D. Measuring in Higher Education, 1923.