



eCOMMONS

Loyola University Chicago
Loyola eCommons

Master's Theses

Theses and Dissertations

1942

The Persistence of Erroneous Assent in True-False Tests

Francis J. Sweeney
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses

 Part of the [Psychology Commons](#)

Recommended Citation

Sweeney, Francis J., "The Persistence of Erroneous Assent in True-False Tests" (1942). *Master's Theses*. 699.

https://ecommons.luc.edu/luc_theses/699

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#).
Copyright © 1942 Francis J. Sweeney

THE PERSISTENCE OF ERRONEOUS ASSENT IN TRUE-FALSE TESTS

BY

FRANCIS J. SWEENEY

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF MASTER
OF ARTS IN LOYOLA UNIVERSITY

JUNE

1942

TABLE OF CONTENTS

CHAPTER	PAGE
I. THE STATEMENT OF THE PROBLEM.....	1
II. THE PRESENT STATE OF STUDIES AFFECTING THE PROBLEM	10
III. THE METHOD OF INVESTIGATION	16
IV. THE RESULTS OBTAINED	26
V. SUMMARY AND CONCLUSIONS	41

VITA

Francis J. Sweeney was born in Chicago, Illinois, January 31, 1909.

He attended St. Patrick's Grammar School, on the far South Side of Chicago and Bowen Public High School. From the latter he was graduated in June, 1927.

During the next eight years he worked, mostly at electrical work in industrial plants. During part of this time he attended evening school at Armour Institute of Technology for six semesters where he studied electrical engineering.

In October, 1935 he entered the University of Chicago and in June, 1939 was granted the Bachelor of Arts degree with a major in Psychology.

In September, 1939, he began graduate work in Psychology at Loyola University and was appointed an assistant in the Department of Psychology.

He was a teaching fellow in the Department of Psychology for the school years 1940-41 and 1941-42.

CHAPTER I

THE STATEMENT OF THE PROBLEM

Examinations of some kind have been given as long as schools have been in existence. The chief purpose of examinations is to measure the achievement of the student in the course of study. Indirectly, of course, the examination also measures the student's ability, at least his ability to do the type of work required for success in the course. Besides this main purpose, other purposes served by examinations are also widely recognized. Among these may be mentioned the diagnosis of students' disabilities, the measurement of teaching efficiency, the determination of goals or standards of attainment, and the provision of motivation and opportunities for effective learning.

The relative merits of various types of examinations with regard to the realization of these purposes has been the subject of a vast amount of research since the beginning of the present century and the periodicals dealing with educational topics have published numerous reports and discussions concerning the findings of the research, their validity, reliability, meaning, importance, etc. Much of this discussion is controversial and involves philosophical and other theoretical assumptions the validity of which it is difficult to determine. Fortunately, it is not necessary to attempt any such validation here. However, it is appropriate to point out that from the controversy concerning examinations there has, in some quarters, crystallized the conviction that the traditional type of examination

that calling for essay answers to questions, is valueless, or almost so, because of the subjective factors which are likely to influence the grading of answers.

During the time the traditional type of examination was being evaluated, and, in the opinion of many, being found wanting in validity and reliability, other types of examinations were being developed by researchers on educational problems. It was widely recommended that the essay examination be replaced by the "scientifically" constructed tests being developed to meet research needs. Teachers were influenced to regard these new-type tests as scientific and to adopt and use them. Often no distinction was made between types of tests suitable as measuring devices in research and those more fitted for achievement of practical school purposes.

It was perhaps natural that the protagonists of new-type tests should, in their enthusiasm over the success they met, come to overestimate and overstate the values of these tests. But careful consideration indicates that the traditional examination and new-type tests each has individual points of strength and weakness peculiar to it. Therefore, there can be no question of which is better in itself. The only legitimate question is which is better for a particular purpose. Conversely, it appears likely that some types of tests should be more ill suited than others to achieve a given purpose.

In the light of these convictions it seemed advisable to undertake still another investigation on new-type tests—this time to determine whether these tests were actually achieving the general purposes of examinations. It was thought best to limit the problem to one of the new-type

tests, namely the True-False type and to seek to discover whether or not it meets the purpose of providing opportunities for effective learning. The investigation was further limited to tests in only one subject, general psychology as taught at Loyola University.

The main problem of this investigation is, therefore, whether repeated presentation of false statements in True-False tests interferes measurably with the realization of the purposes of a course in general psychology. The objectives of such a course, as stated in the syllabus prepared by the Department of Psychology, are: (1) to acquaint the student with the chief facts, laws and principles of mental life; (2) to impart the theoretical basis and practical rules of self-knowledge and self-discipline; and (3) to lay the foundation for further studies in psychology and related disciplines, such as logic, ethics and social studies. The acquisition by the student of these skills and items of information should result from successful completion of the course. They may be regarded as relatively objective and capable of measurement by appropriate tests. Other educational outcomes such as attitudes, appreciations and interests are more subjective, difficult to define or even to express, hence difficult to measure by means of paper and pencil tests.

But the problem here is not whether True-False test results are a valid and reliable measure of educational outcomes. Correctly constructed and used, True-False tests are commonly held to be sufficiently valid and reliable for practical purposes, that is, for distinguishing those students who have mastered the minimum essentials, or met any other standards set in the course of study, from those who have not. Rather, the problem here considered is whether the True-False test interferes measurably with the

acquisition of the information and skills the course in general psychology is designed to impart. This problem is suggested by the fact, shown in many studies, that students tend more often to mark false statements true, than they do to mark true statements false. This tendency seems to result from the apparent plausibility and the grammatical correctness of the false statements. It is necessary that the test constructor select false statements having great appearance of truth; because if a statement is too flagrantly false no one will judge it to be correct and thus the item will be invalid since it will not differentiate those who know the matter of the statement from those who do not.

A priori it seems probable that students' experience of judging and marking false statements as true should to some extent result in what might be called negative learning, that is the assent to false statements should leave the students with erroneous ideas concerning the matter of the statements. This statement of the problem leads to the hypothesis here advanced: if there is any deleterious effect resulting from assent to false statements presented in True-False tests, this effect should be reflected in lower average final examination scores for the specific treatises on which students have been tested by this device.

It is here assumed that the proportion of false statements again marked true on a second presentation is a measure of the extent to which the first presentation of these statements had a deleterious result. This assumption is justified only if other factors are held constant, or the variability of these other factors be correctly accounted for otherwise. The efforts made to control variable factors will be indicated in the chapter on the method of this investigation.

But, it might be objected that the use of the same True-False items in the final examination to measure the effect these items had on their first presentation is an unwarranted procedure. It is true that new and unknown accidental variable factors may, to an indeterminable extent, influence the students' judgments of these items in the final examination. Moreover, the same influences that operated on the first presentation, namely the apparent plausibility and grammatical correctness of the items, may be active in the final examination and so make for erroneous judgments again. The proportion of items again judged incorrectly on the second presentation would then not be wholly due to persistence of erroneous assent. To escape this difficulty it is necessary on the final examination to test students' knowledge of the separate treatises of the course by other devices than the True-False test, the assumption here being that those who took a True-False test on a specific treatise should have less actual knowledge of the matter of this treatise than those who were not exposed to false statements concerning this matter. This inadequacy of knowledge should be measurable by means of other types of tests. Therefore, those who took a True-False test on a certain treatise should do worse on that part of the final-examination dealing with the matter of that treatise, even when tested by Completion, Multiple Choice and Yes-No types of tests.

It is also to be expected that in answers to the essay questions of the final examination some echo, in the form of either direct or paraphrased quotations of the false statements, is to be discovered. This would of course be the strongest type of evidence for the thesis that negative learning results from the taking of True-False tests. In short, by whatever

device tested, the students who were exposed to false statements should know less about the matter of the course, if there is any persistence of erroneous assent.

The objection that negative learning may result from taking tests does not apply in other than True-False tests as can be easily understood from an analysis and comparison of the mental processes involved in taking the various types of tests. If the hypothesis is correct, and students who took True-False tests on a specific treatise make lower average scores on the final examination than those who took alternate types of tests this would indicate either: (1) that the group which took True-False tests did not learn as much from these tests; or (2) that the group which took True-False tests persisted in error. In other words, either less positive or more negative learning results from taking True-False tests.

The second alternative is the one here advocated. This can be shown to be correct by the following line of reasoning. There is more opportunity for positive learning in completion than in True-False test items since in the former only one term of a proposition is presented and in order to complete the statement the student must recall and recognize the other term of the proposition or else he must construct a term and judge its appropriateness for completing the statement. Either of these processes is of benefit to the student in that either makes for retention of the judgment expressed by the completed proposition. But there is no opportunity for negative learning in Completion test items since the way the proposition is to be completed is in no way suggested. In Yes-No and Multiple Choice items the same opportunity for positive learning obtains as for True-False items, but

there is less opportunity for negative learning. This is due to the fact that while in Multiple Choice and Yes-No items the same amount of material is presented as in True-False items, namely the two terms of a proposition, the nexus between the terms, the formal element of judgments, is only dubiously presented in Multiple Choice and Yes-No items; but in True-False tests affirmation is suggested in every item due to its grammatical construction and apparent plausibility. Hence this suggestive effect of True-False items results in negative learning, but its absence in Yes-No and Multiple Choice items does not induce more positive learning because these types of items contain nothing over and above the presentation of material elements of judgments and this is common to all three forms. Therefore, differences in scores must be due to negative learning, induced misinformation, since opportunity for positive learning is equal in True-False, Yes-No and Multiple Choice forms, and in favor of the completion over the True-False form.

The thesis, then, that True-False tests result in negative learning appears to be very probable when the importance of suggestion is considered. Virtually all psychologists are agreed that suggestibility is a general tendency of the human mind. Everyone is prone to think and act in any manner suggested by others, especially when one is fatigued, in doubt or under any other kind of strain. Students confronted with an objective test item naturally react to it with a mental state of doubt. If the test item happens to be one calling for completion of a judgment the student must recall and recognize the bit of information which is the other term of the judgment. The Completion type test is therefore "a recall-type test" and

no erroneous answer is implied or hinted at by the form of the item. Only recall of the required information and not suggestion can resolve the students' doubt in this case. If a Multiple Choice item is presented, each of the two or more answers to it is only dubiously suggested as the correct one. This type of test is primarily "a recognition-type test" and the student is called upon to recognize the term which correctly completes the judgment.¹ No one answer is more strongly implied to be the right one and so the doubt of the student is not resolved by suggestion but by recognition. The Yes-No type of item, since it is merely a definite question calling for an answer of yes or no, does not by its form imply the correctness of either answer. The student can resolve his doubt concerning such a question only on the basis of his knowledge of the subject matter. However, the True-False item does suggest an affirmative answer by its very form. Each such statement is grammatically correct and its falsity of content is very apt to go unrecognized just because of this correctness of form. Therefore, the students' knowledge of the subject matter is not the only factor operating to resolve his doubt. Due to the suggestive effect of the form of the True-False item the student is led to make false judgments and thereby to acquire misinformation. The fact that students generally mark more false statements as true is good corroborative evidence of this suggestive effect. No one could hold that the student is in any way benefitted by the acquisition of false information, nor that the objectives of a course of study are in this way

¹ The terms "recall type-test" and "recognition type test" are currently used in educational literature. In normal human memory, of course, both the mental processes of recall and recognition are functioning during the taking of either type test, but in the case of the "recognition-type test" the necessity for spontaneous recall, on the part of the student taking the test, is obviated by the form of the test.

attained.

So then, the problem of whether false statements presented in True-False tests interferes with mastery of subject matter is answered affirmatively on the basis of reason. But, to test experimentally this answer, and to attempt to determine to what measurable extent the suggestive effect of false statements operates, the present investigation was undertaken.

CHAPTER II

THE PRESENT STATE OF STUDIES AFFECTING THE PROBLEM

Careful search of the standard textbooks on educational measurement, The Psychological Abstracts and Wilson's Education Index indicates that no report of an investigation of any problem exactly similar to the problem of the present study exists. Consultations with members of the Department of Education at Loyola University confirm this conclusion.

However, the reports of two studies, the titles of which suggested that they might throw some light on the present problem, were looked into. One study made by H. H. and E. M. Remmers and entitled "The Negative Suggestion Effect of True-False Examinations Questions"¹ is mentioned by at least one author² of a textbook on educational measurements as tending to show that true-false tests do not suggest false information to such an extent as to make this type of test injurious.

In the opinion of the present writer this investigation of the Remmers is not concerned with the precise problem of this thesis, nor is the method employed by them the same as the one used for this study, and so their conclusions do not render the present investigation superfluous. But in view of the resemblance between the problems of the Remmers and of the present investigation it was thought best to discuss briefly the former study.

1 H. H. and E. M. Remmers. "The Negative Suggestion Effect of True-False Examination Questions." Journal of Educational Psychology, January, 1926, 52.

2 C. W. Odell. Traditional Examinations and New Type Tests. The Century Co., New York, 1928.

The report of the Remmers' study may be briefly summarized as follows. The problem as they state it is: "Do true-false tests tend to leave a residue of false associations?" In seeking an answer to this problem an experiment was conducted on one hundred and thirty-six college sophomores. The subjects were divided into two equal groups of sixty-eight students each, and the groups were equated as to intelligence on the Otis Self-administering Higher Examination. All the subjects were given to read a passage from Tacitus' "Customs of the Germans". From the matter of this passage a one hundred and twenty-one item true-false test was constructed and also a completion type test of the same number of items and of the same content as the true-false test. The day after the students had read the passage one group was given the true-false test and the other group the completion test. Four weeks later the subjects were retested with the same tests, but the type of test given each group was reversed so that those who took the true-false test the first time took the equivalent completion test the second time, and vice versa. The scores each group made on both tests were averaged. The results indicated that there was no significant difference between the average score on both tests of the group that took the true-false test first and the average score of the group that took the completion test first. From this the authors conclude: "With the kind of material and subjects used there is no evidence of negative carry-over from the taking of true-false examinations."³

The nature and setting of the problem and the method of investigation in the Remmers' experiment were not precisely the same as in the present

³ Remmers, op. cit., 56.

investigation as is clear from the following considerations. First, their subject matter was limited; the students were allowed to read it only once and no efforts were made to teach it. Whereas in the present study the subject matter of the test was the regular course matter and the subjects were given every help and encouragement to master it. Second, their basis for the comparison of the groups was the sums of the mean scores made by each group on both the true-false and the completion test. In the present study the comparison of the groups is made on the basis of the average percentage of correct responses on both the true-false and alternate type test items of the final examination that correspond in form and content to each of the subtests. It is not evident that the method of the Remmers yields an adequate measure of the effects of true-false tests in general, and so their conclusion cannot be interpreted as arguing against the persistence of erroneous assent under all circumstances. Whether the method of the present experiment yields any better results remains to be seen.

In commenting on the disadvantages of true-false tests Odell⁴ admits that false statements are more confusing than other types of test exercises but he adds that this effect is often overestimated. He cites the study of Remmers and Remmers as evidence that the after effects of true-false tests need not be feared. Enough has been said of the differences between the study of the Remmers and the present one to suggest that this conclusion of Odell need not be considered as a final answer to the problem. Another method may reveal that the persistence of erroneous assent is measurable otherwise than by the devices the Remmers employed.

⁴ Remmers, op. cit., 335.

Another study made by McClusky⁵ was concerned with the immediate negative suggestion effect of the true-false test. He presents data to show that in the final examinations of four different courses at the University of Michigan false statements contributed a total of 1796 errors whereas only 1174 errors were made on the true statements. This indicates that false statements present much greater difficulty than true statements on the true-false test. This fact makes relevant an investigation of the negative suggestion effect of the true-false examination. His method was to administer a true-false test followed immediately by a multiple-choice test identical in subject matter. Two forms of the true-false test were used each containing eighty items, half of which were true and half false. The two forms differed in that the forty true items of form A were made false for form B and the forty false items of form A were made true for form B. The multiple choice test likewise consisted of eighty statements with four choices to complete each statement. All three tests were identical in content. The tests were administered to ninety-five students in the first course in educational psychology at the University of Michigan. The group was divided into four sections and sections one and three took form A of the true-false test while sections two and four took form B of the same test. Immediately afterwards all sections took the same multiple choice test. This method was used to insure that any differential responses obtained would be due to the form only and not the content of the test items. Answers to each statement on the true-false test were compared with the answers to the corresponding statement

⁵ Howard Y. McClusky. "The Negative Suggestion Effect of the False Statement in the True-False Test." The Journal of Experimental Education, March, 1934, 2, 269.

on the multiple choice test. The results of this comparison showed that the false statements of the true-false test were the sources of more errors on the multiple choice test than were the true statements. This held regardless of the form of the true-false test and the intelligence of the subjects. The author concludes that "The evidence clearly indicated that the immediate effect of the false statement was negative and detrimental."⁶ The reason for this, he holds, is that "the false statement itself...apparently introduces a 'confusional element' in the performance of the subject."⁷ This is the same idea that is at the basis of the present investigation, except that McClusky was concerned only with the immediate effect of this "confusional element", whereas this study is concerned with its persistence. Indeed, he adds the warning that since his "study has dealt only with the immediate effect of the false statement; great caution must therefore be exercised in extending the conclusions to the problem of delayed performance."⁸

However, it seems likely that since the immediate effect of false statements was detrimental this effect should persist; especially since in this experiment no direct means were used to counteract it. The students who served as subjects of this study were not informed of the numbers and natures of the errors they made on the sub-tests, nor were the correct answers to the items on any of the tests discussed with them, either in class or in private. Each student was merely informed of the grade he had merited on each of the sub-tests, but he was not told of the error equivalence of his grade. But, since McClusky's problem and the present one do

⁶ Ibid., 273.

⁷ Ibid.,

⁸ Ibid.

differ in the important respect indicated his results are not an adequate answer to the present problem, although they are suggestive. Therefore it seemed worthwhile to undertake the present study.

CHAPTER III

THE METHOD OF INVESTIGATION

The entire group of students of General Psychology during the first semester of the school year 1941-1942 were the subjects for an experiment to determine whether repeated presentation of false statements in True-False tests did, in fact, interfere measurably with mastery of the subject matter of the course. The total number of subjects whose test results are used was one hundred and thirty. The test results of a few others are not used in the final tabulation because such students failed to take one or more tests at the proper time, due to absence on the days the tests were given. This procedure seems legitimate since the results of a test taken by a student at a later time, as make up work, are not comparable with the test results of other students who took the test earlier; because the student taking the test later has had more time for study or review, or may have acquired, from those who took the test previously, information concerning the nature of the test, the detailed matter covered by it and possibly the correct answers to some of the items. All of these factors could operate to defeat the purpose of a test and there is no way to predict or control their operation.

The total number of subjects were divided into six groups on the basis of their "quiz" sections. These groups are designated by the letters A to F inclusive and were made up of the following numbers of students respectively: twenty, fifteen, thirty-one, twenty-three, twenty-five, and sixteen, or a total group of one hundred and thirty. This was done because

it was deemed advisable to give the tests in the "quiz" classes where the testing could be more carefully supervised and the chances for students copying from each other, or from their notes or books, could be eliminated.

Approximately one half of each "quiz" class was given a fifty item True-False test on the first two treatises of the course matter. Twenty-five of the statements on this test were true and the other twenty-five were false. The criterion of truth was the presentation of the matter in the text used in the course and in the lectures. The other half of each class was given an alternate test consisting of twenty completion items, thirteen Multiple Choice items and seventeen Yes-No items; but since some of the Completion items contained more than one blank and each such blank was counted as an item this gave twenty-eight completion items or a total of fifty-eight items for the test as a whole. The matter of this test was the same as that of the True-False test, namely the introductory chapter and the chapter on emotions as presented in Fr. Herr's textbook: Introduction to General Psychology. Twelve true and thirteen false statements were derived from the first chapter and thirteen true and twelve false statements concerned the matter of the second chapter. On the alternate test sixteen Completion, six Multiple-Choice and eight Yes-No items dealt with the introductory matter of the course; while twelve completion, seven Multiple-Choice and nine Yes-No items were constructed from those true and false statements based on the second chapter of the text.

The method of constructing the tests was to make up first a true or false statement concerning a definition, principle or fact deemed to be of importance for a basic knowledge of general psychology. Fifty such state-

ments, half of them true and half of them false, comprised the first True-False test. Then each of these statements was converted into a Completion, Multiple-Choice or Yes-No item on the basis of the ease and readiness with which each statement could be adapted to one of these forms. Only true statements can be converted to Completion items without change of meaning. The process of conversion consists only of dropping one or more essential words or phrases from either term of the statement, or of dropping one term completely. Blanks are substituted for the missing word, phrase or term, and the student is required to write in the blank spaces any appropriate word, phrase or term which, in his judgment, makes the statement complete and true. Other statements for which several plausible second terms can be found can be easily converted to Multiple-Choice items. In constructing such items, one term of a statement is given as true and three or four second terms, all but one of which are specious, are provided as possible completions of the proposition. From among these second terms the student is directed to select the one which he thinks best and most truly completes the proposition. Statements, mostly false ones, that did not readily lend themselves to conversion to either the Completion or Multiple-Choice form were changed into Yes-No items by prefixing some such phrase as: "Is it true that," or otherwise changing the statement into a direct question that could be answered by yes or no.

The first tests were given in the "quiz" classes during the week of October 6, 1941. The test results of sixty-seven students who took the first True-False test and of sixty-three who took the first alternate test are used in the final compilation. The chief reason why there were four more who

took the True-False form of the test was that four of the six "quiz" groups contained odd numbers of students and since the forms of the test were passed out alternately according to the way in which the students were sitting in the class room, it happened that this distribution resulted. No effort was made to divide the whole group into two equal sub-groups, but merely to divide each "quiz" class into two sub-groups as nearly equal as possible. However, since all comparison between groups are in terms of percentages of errors and of correct responses the fact that one group contained four more students than the other is not significant. Then too, since the test results of those who took the tests later, due to absences on the days the tests were first given to the classes, were not to be counted in, and since it could not be foreseen how many would miss one or more tests it was not considered essential to have the groups absolutely even numerically as long as they were approximately so. In each subsequent test on the later treatises it was, of course, made certain that each student took the form of the test different from the one he took on the previous treatise.

On the second treatise the types of tests given each half of each "quiz" class were interchanged, so that those who took the True-False form on the first treatise took the Alternate test on the second and those who took the first Alternate took the second True-False test. On the two subsequent treatises the same kind of an interchange was made.

The reason for this interchange was to cancel out chance variations and individual differences within and between the two groups. Differences in intelligence and in study methods or habits of the students composing the groups; as well as differences in instruction, quizzing, emphasis on various

facts and principles of the subject matter and similar accidental variables that naturally result from the way the course was conducted, will not operate to invalidate the comparison of groups when this interchange is made. Each group is thus compared with itself with regard to the effects each type of test had on its learning of specific parts of the course matter. This is apparent when it is noted that the ultimate comparison is between each student's performance on two parts of the final examination; that part upon the matter of which he had been previously tested by True-False tests and that upon whose matter he had taken alternate type tests. The accidental variables due to differences in instruction and so forth are overcome by comparing each half of each "quiz" class with the other half of the same class which had the same opportunities for learning. These two considerations together with the method of taking the percentage of the average number of correct responses to test items as the basis of comparison lead to the assurance that in this experiment the type of test each student took on each treatise is the only independent variable.

The matter of the second tests was Chapters Three and Four of the textbook, those which treat of Instincts and of Innate Equipment. These tests were made up in the same manner as were the first tests. The second True-False test consisted of twenty-five true and an equal number of false statements. Twelve of the true and thirteen of the false propositions were based on the discussion of instincts, while thirteen true and twelve false statements concerned the subject of innate equipment prerequisite for mental life. The second Alternate test was composed of twenty-three Completion items, counting each blank as an item. Eleven of these were on instincts

and twelve on innate equipment. Also, eleven Multiple-Choice items, six on instincts and five on innate equipment, and twenty-one Yes-No items, ten from Chapter Three and eleven from Chapter Four, were included in this test. These second tests were given during the week of November 3, 1941. The test results of sixty-three students who took the second True-False test and of an equal number who took the second Alternate test are utilized.

The third treatise concerned the Nervous System and Sensation. Test items on this matter were derived from Chapters Five and Six of the textbook. Twenty-eight true and thirty-two false, or a total of sixty statements comprised the third True-False test. Of these, ten true and twelve false propositions concerned the nervous system, and eighteen true and twenty false propositions were about the process of sensation. To make up the third Alternate test these statements were converted into fifty Completion, fourteen Multiple-Choice and thirteen Yes-No items. Twenty-three of the Completion, two of the Multiple-Choice and seven of the Yes-No items were based on the discussion of the nervous system; while twenty-seven Completion, twelve Multiple-Choice and six Yes-No items were on the topic of sensation. Scores of sixty-two students on the third True-False test and of the same number on the third Alternate test are used as data for this experiment.

The fourth tests were based on Chapters Seven, Eight and Nine of the textbook and dealt with Complex Sensations, Perception and Imagination. Twenty-four true and twenty-six false propositions made up the fourth True-False test. Five true and seven false statements were on complex sensations; seventeen true and sixteen false were on perception; and two true and four false were on imagination. The sixty-two items of the fourth Alternate test

were so distributed as to yield thirty-two Completion, fifteen Multiple-Choice and fifteen Yes-No items. Of the Completion items seven concerned Complex sensations, twenty-two were on perception and three about imagination. The Multiple-Choice group consisted of four on complex sensation, eight on perception and three on imagination. Seven of the Yes-No items were based on the matter of complex sensations and eight on perception. Test results of sixty-two students who took the fourth True-False test and of fifty-eight who took this Alternate test during the week of December 15, 1941, are made use of in the final comparison.

The final examination consisted of seventy-seven True-False items, thirty-one true and forty-six false; forty-eight Completion, fifteen Multiple-Choice and nine Yes-No items, or a total of seventy-two objective items other than True-False. All of these items had been presented previously in the various sub-tests. Twenty-one items from the first True-False test were used in the final examination; from the second True-False test eleven items, from the third True-False test twenty-four items and from the fourth True-False test twenty-one items were again utilized. From the first Alternate test ten Completion and four Multiple Choice items were adopted. Seven Completion, two Multiple-Choice and four Yes-No items were taken from the second Alternate test. Twenty-one Completion, four Multiple-Choice and three Yes-No items from the third Alternate test were utilized; while from the fourth Alternate test ten Completion, five Multiple-Choice and two Yes-No items were used.

In all cases the test items to be presented in the final examination were selected on the basis of their validity, that is according to how well

they differentiated students who knew the matter from those who did not. In accordance with current practice those items which not less than twenty-per cent nor more than eighty per cent of the students failed on were deemed valid. This is admittedly an arbitrary criterion, but any other would be equally so, and this one has the merit of being sanctioned by wide usage. It does seem reasonable to hold that if less than twenty per cent of an average class of students miss an item it is too easy, while if more than eighty per cent miss one it is too hard. An item that nobody or everybody misses does not help to distinguish those who have learned the matter from those who have not.

In addition to the objective items the final examination also contained ten short essay questions calling for direct answers. Of these two questions concerned the mental process of perceiving, two had to do with the differences and likenesses between a percept and an image, two dealt with the process of sensation and one question was about the process of imagination. The answers given to these questions are to be scanned to discover whether they contain any misinformation that can be traced to false statements on previous tests. The last two questions, on learning, are not thus evaluated because they do not deal with matter that was subject to previous false statement.

The entire group of a hundred and thirty students took this final examination on January 19, 1942. They were directed to answer the essay questions first, then these were collected so as to preclude, as far as possible, their getting false information from reading the True-False items.

The results of all tests and of the final examination are to be

presented so as to reveal the average percentage of correct responses made by each half of each "quiz" class on each test they took and also the average percentage of correct responses each such sub-group made on the parts of the final examination corresponding to each treatise of the course. The sums of the average percentages of correct responses made on all True-False tests and on all Alternate tests will be computed and compared with those made on the parts of the final examination corresponding to each treatise.

If there is persistence of erroneous assent, on the corresponding part of the final examination, the average percentage of correct responses for all those who took the True-False test on the first treatise should be lower than the average percentage of correct responses made by all those who took the Alternate test on this treatise. Also, on the parts of the final examination corresponding to the other treatises those who took the True-False tests on the matter of these should make lower scores than those who were not exposed to false statement concerning the matter of these parts of the course. In addition, the average of the sum of the percentages of correct responses on the final examination as a whole should be lower for the entire group that took True-False tests than for the group that took the Alternate tests. Finally, it is to be expected that some traces of misinformation, acquired from exposure to false statements, will be revealed in the answers to the essay questions, given by those who had previously encountered and assented to the erroneous notions contained in the false statements of the True-False tests.

The technique outlined above seems to be adequate to test experimentally the hypothesis. All conceivable accidental variables due to individual

differences among the students, inequality of groups, differences in instruction and so on, were eliminated by interchanging the groups to which True-False tests were given and by taking the percentage of the average number of correct responses as the basis for comparison.

CHAPTER IV

THE RESULTS OBTAINED

The results obtained on the four sub-tests and on the parts of the final examination corresponding to each of these tests can be most orderly presented in the form of a series of tables: Tables I, III, V and VII set forth the data concerning the four sub-tests. They reveal the composition of the groups that took each form of the sub-tests, the average number of errors made, the average percentage of errors and the average percentage of correct responses made on each form of each sub-test. Tables II, IV, VI and VIII are the comparison of the average number of errors, percentage of error and percentage of correct responses on the parts of the final examination corresponding to the sub-tests, made by those who previously took True-False tests and by those who took Alternate type tests on the matter covered by these parts of the final examination. These, and Table IX are the significant tables for this study. Table IX summarized Tables II, IV, VI and VIII and thus compares, in terms of percentages of correct responses, the group that took all True-False tests with the group that took no True-False tests. The groups designated by the symbols A 1, A 2 etc. are the quiz groups which were so divided that each half could be given a different type test on each treatise of the course matter. The composition of these groups remains unchanged throughout, but the type of test given each group was alternated for each treatise.

Sub-Test I

T-F 50 items. Alternate 59 items.

Group	N	Test	Avg. Err.	% Err.	% C.R.
A 1	10	T-F 1	10.6	21.2	78.8
B 1	9	"	11.0	22.0	78.0
C 1	15	"	12.26	24.52	75.48
D 1	12	"	10.83	21.66	78.44
E 1	12	"	10.41	20.82	79.18
F 1	<u>9</u>	"	<u>10.0</u>	<u>20.0</u>	<u>80.0</u>
Totals	67		65.10	30.20	469.90
Averages			10.85	21.7	78.31
A 2	10	Alt. 1	19.6	33.79	66.21
B 2	6	"	18.83	32.55	67.45
C 2	16	"	19.5	33.62	66.38
D 2	11	"	17.18	29.62	70.38
E 2	13	"	20.0	34.48	65.52
F 2	<u>7</u>	"	<u>13.85</u>	<u>23.87</u>	<u>76.13</u>
Totals	63		108.96	187.93	412.07
Averages			18.16	31.32	68.68

TABLE I

C.R. = Correct Responses

Corresponding Parts of Final Examination

Group	N	True-False Test 21 items			Alternate Type Tests 14 items			
		Avg. Err.	% Err.	% C.R.	Avg. Err.	% Err.	% C.R.	Avg. % C.R.
A 1	10	5.3	25.23	74.77	4.6	32.85	67.15	70.96
B 1	9	6.33	30.12	69.88	5.77	41.21	58.79	64.33
C 1	15	7.93	37.76	62.24	6.4	45.71	54.29	58.26
D 1	12	6.0	27.61	72.39	5.66	44.28	55.72	64.05
E 1	12	5.41	25.76	74.24	5.33	38.07	61.93	68.08
F 1	<u>9</u>	<u>5.11</u>	<u>24.33</u>	<u>75.67</u>	<u>5.33</u>	<u>38.07</u>	<u>61.93</u>	<u>68.8</u>
Totals	67	36.08	170.81	429.19	33.09	240.19	359.81	394.48
Averages		6.01	28.47	71.53	5.51	40.03	59.96	65.74
A 2	10	7.2	34.28	65.72	6.2	44.28	55.72	60.72
B 2	6	6.33	30.12	69.88	5.5	39.28	60.72	65.30
C 2	16	6.68	31.81	68.19	5.87	41.92	58.08	63.13
D 2	11	5.0	23.81	76.19	4.36	31.12	68.88	72.53
E 2	13	7.76	36.95	63.05	5.3	37.85	62.15	62.60
F 2	<u>7</u>	<u>4.57</u>	<u>21.76</u>	<u>78.24</u>	<u>4.57</u>	<u>32.64</u>	<u>67.36</u>	<u>72.80</u>
Totals	63	37.54	178.73	421.27	31.8	227.09	372.91	397.08
Averages		6.25	29.78	70.21	5.3	37.84	62.15	66.18

TABLE II.

TABLE III

Sub-Test II

T-F 50 items. Alternate 55 items.

Group	N	Test	Avg. Err.	% Err.	% C.R.
A 2	10	T-F II	8.5	17.0	83.0
B 2	6	"	6.0	12.0	88.0
C 2	16	"	6.12	10.24	89.76
D 2	11	"	5.63	11.26	88.74
E 2	13	"	10.14	20.3	79.7
F 2	<u>7</u>	"	<u>4.57</u>	<u>9.14</u>	<u>90.86</u>
Totals	63		40.96	79.94	520.06
Averages			6.82	13.32	86.64
A1	10	Alt. II	11.4	20.72	79.28
B 1	9	"	11.09	21.06	78.94
C 1	13	"	11.0	20.0	80.0
D 1	12	"	11.66	21.2	78.8
E 1	11	"	12.72	23.12	76.88
F 1	<u>8</u>	"	<u>10.75</u>	<u>17.72</u>	<u>82.28</u>
Totals	63		68.62	123.82	476.18
Averages			11.43	20.63	79.36

Corresponding Parts of Final Examination

Group	N	True-False Test 11 items			Alternate Type Tests 13 items			
		Avg. Err.	% Err.	% C.R.	Avg. Err.	% Err.	% C.R.	Avg. % C.R.
A 2	10	3.4	30.9	69.1	3.3	25.38	74.62	71.86
B 2	6	4.5	40.9	59.1	2.66	20.46	79.54	69.32
C 2	16	4.06	36.9	63.1	3.62	27.86	72.14	67.62
D 2	11	2.63	23.9	76.1	3.18	24.44	75.56	75.83
E 2	13	3.46	31.45	68.55	4.46	32.0	68.0	68.27
F 2	7	2.28	20.72	79.28	1.57	12.07	87.93	83.6
Totals	63	20.33	184.77	415.33	18.79	142.21	457.79	436.50
Averages		3.38	30.76	69.22	3.13	23.7	76.29	72.75
A1	10	3.6	32.72	67.28	3.3	25.38	74.62	70.95
B 1	9	3.22	29.27	70.73	3.44	26.46	73.54	72.13
C 1	13	4.76	43.27	56.73	3.61	27.76	72.24	64.48
D 1	12	2.41	21.9	78.1	2.75	21.15	78.85	78.47
E 1	11	3.18	28.9	71.1	4.27	32.84	67.16	69.13
F 1	8	3.25	29.54	70.46	2.5	19.23	80.77	75.61
Totals	63	20.42	185.60	414.40	19.87	152.82	447.18	430.77
Averages		3.4	30.93	69.06	3.31	25.47	74.53	71.29

TABLE IV

TABLE V

Sub-Test III

T-F 60 items. Alternate 77 items.

Group	N	Test	Avg. Err.	% Err.	% C.R.
A 1	10	T-F III	15.5	25.0	75.0
B 1	8	"	15.87	26.45	73.55
C 1	14	"	18.64	31.06	68.94
D 1	12	"	19.33	32.21	67.79
E 1	12	"	18.75	31.25	68.75
F 1	<u>9</u>	"	<u>18.77</u>	<u>31.28</u>	<u>68.72</u>
Totals	65		105.86	177.25	422.75
Averages			17.64	29.54	70.45
A 2	10	"	32.6	42.33	57.67
B 2	6	"	30.66	39.81	60.19
C 2	16	"	29.68	38.54	61.46
D 2	11	"	35.27	45.8	54.2
E 2	12	"	39.75	51.55	48.45
F 2	<u>7</u>	"	<u>34.0</u>	<u>44.15</u>	<u>55.85</u>
Totals	62		201.96	262.18	337.82
Averages			33.66	43.70	56.30

Corresponding Parts of Final Examination

Group	N	True-False Test 24 items			Alternate Type Tests 28 items			
		Avg. Err.	% Err.	% C.R.	Avg. Err.	% Err.	% C.R.	Avg. % C.R.
A 1	10	8.1	33.75	66.25	9.4	33.92	66.08	66.16
B 1	8	8.25	34.37	65.63	11.87	42.39	57.61	61.62
C 1	14	9.28	38.66	61.34	13.78	49.1	50.9	56.12
D 1	12	11.25	46.87	53.13	12.5	44.64	55.36	54.24
E 1	12	9.41	39.21	60.79	14.25	50.89	49.11	54.95
F 1	<u>9</u>	<u>12.33</u>	<u>51.37</u>	<u>48.63</u>	<u>14.77</u>	<u>52.75</u>	<u>47.25</u>	<u>47.94</u>
Totals	65	58.62	243.23	355.77	76.57	273.69	326.31	341.03
Averages		9.73	40.53	59.29	12.76	45.61	54.38	56.83
A 2	10	7.7	32.08	67.92	10.3	36.78	63.22	65.57
B 2	6	8.66	36.08	63.92	10.66	38.07	61.93	62.92
C 2	16	9.12	38.0	62.0	12.08	43.07	56.93	59.46
D 2	11	8.9	37.08	62.92	11.45	40.89	59.11	61.01
E 2	12	11.0	45.83	54.17	14.91	53.25	46.75	50.46
F 2	<u>7</u>	<u>10.14</u>	<u>42.29</u>	<u>57.71</u>	<u>13.14</u>	<u>46.92</u>	<u>53.08</u>	<u>55.39</u>
Totals	62	55.52	231.36	368.64	72.52	258.98	341.02	354.81
Averages		9.25	38.56	61.44	12.08	43.16	56.82	59.13

TABLE VI

TABLE VII

Sub-Test IV

T-F 50 items. Alternate 62 items.

Group	N	Test	Avg. Err.	% Err.	% C.R.
A 2	10	T-F IV	11.3	22.6	77.4
B 2	6	"	9.83	19.86	80.14
C 2	16	"	10.75	21.5	78.5
D 2	10	"	12.8	25.6	74.4
E 2	10	"	13.6	27.2	72.8
F 2	<u>7</u>	"	<u>8.14</u>	<u>16.28</u>	<u>83.72</u>
Totals	59		66.42	133.04	466.96
Averages			11.07	22.17	77.82
A 1	9	Alt. IV	20.88	33.67	66.33
B 1	8	"	16.5	26.61	73.39
C 1	14	"	22.92	36.96	63.04
D 1	10	"	27.6	44.51	55.49
E 1	9	"	17.0	27.42	75.58
F 1	<u>8</u>	"	<u>27.27</u>	<u>44.16</u>	<u>55.84</u>
Totals	58		132.27	213.33	386.67
Averages			22.04	35.55	64.44

Corresponding Parts of Final Examination

Group	N	True-False Test 21 items			Alternate Type Tests 17 items			
		Avg. Err.	% Err.	% C.R.	Avg. Err.	% Err.	% C.R.	Avg. % C.R.
A 2	10	6.5	30.95	69.05	9.0	52.94	47.06	58.05
B 2	6	7.5	35.71	64.29	6.5	38.23	61.77	63.03
C 2	16	6.5	30.95	69.05	9.31	54.76	45.24	57.14
D 2	10	7.2	34.28	65.72	8.6	50.59	49.41	57.56
E 2	10	7.9	37.61	62.39	11.7	69.35	30.65	46.52
F 2	<u>7</u>	<u>7.57</u>	<u>36.04</u>	<u>63.96</u>	<u>6.28</u>	<u>36.35</u>	<u>63.55</u>	<u>63.80</u>
Totals	59	43.17	205.54	394.46	51.39	302.22	297.68	346.10
Averages		7.19	34.25	65.74	8.56	50.37	49.61	57.68
A 1	9	5.77	24.47	72.53	8.22	48.35	51.65	62.09
B 1	8	6.0	28.57	71.43	7.87	46.29	53.71	62.57
C 1	14	8.21	39.09	60.91	9.5	55.88	44.12	52.51
D 1	10	6.1	29.09	70.91	9.2	54.11	45.89	58.40
E 1	9	6.11	29.09	70.91	6.88	44.70	55.30	63.10
F 1	<u>8</u>	<u>6.75</u>	<u>32.14</u>	<u>67.86</u>	<u>8.62</u>	<u>50.70</u>	<u>49.30</u>	<u>58.58</u>
Totals	58	38.94	185.45	414.55	50.29	300.03	299.97	357.25
Averages		6.49	30.90	69.09	8.38	50.00	49.99	59.54

TABLE VIII

TABLE IX

Parts of Final corresponding to Sub-Test:	Group that took all True False Tests			Group that took all Alternate Type Tests		
	True-False % correct Resp.	Alternate % correct Resp.	Average % correct Resp.	True-False % correct Resp.	Alternate % correct Resp.	Average % correct Resp.
I	71.53	59.96	65.74	70.21	62.15	66.18
II	69.22	76.29	72.75	60.06	74.53	71.29
III	59.29	54.38	56.83	61.44	56.82	59.13
IV	65.74	49.61	57.68	69.09	49.99	59.54
Totals	265.78	240.24	253.00	269.80	243.49	256.14
Averages	66.44	60.06	63.25	67.45 <u>66.44</u>	60.87 <u>60.06</u>	64.03 <u>63.25</u>
			Differences	1.01	.81	.78

Since this method of comparing the performances of the two groups on the objective parts of the final examination did not reveal differences as large as were expected it was decided to try another method of comparison. It seemed likely that the average number of errors made by all subjects on those parts of the final examination that corresponded to the true-false tests each took should be greater than the average number of errors made on those parts of the final examination that corresponded to the alternate type tests each took. Accordingly a count was made of the number of errors each student made on the parts of the final examination corresponding to each sub-test he had taken. The average numbers of errors made on all parts of the final examination corresponding to the two types of sub-tests were computed and found to be 27.18 for those parts of the final examination corresponding to all true-false sub-tests and 26.82 for those parts corresponding to all alternate type sub-tests. The standard deviations were found to be 9.9 in the first case and 9.41 in the second. The probable errors of the means were .67 and .58 respectively. The probable error of the difference was .88 and the critical ratio .4.

So small a critical ratio cannot be regarded as revealing significant differences between the groups compared.

Answers to the questions on the essay part of the final examination were marked by a person other than the present writer. Each of the ten questions was given an equal weight of ten points and points were subtracted for major errors made in each answer. It was expected that students who had been presented with false statements concerning the matter of a particular treatise would make a larger number of errors in their answers to questions

about this same matter than would be made by students not exposed to the false statements. This expectation was not verified to a statistically reliable extent. Those who had taken true-false tests did not make enough more errors than those who took the alternate type tests. It can be seen from the following table that the average difference between the groups is only 1.19 points.

Treatise on:	Took True-False Tests Avg. No. of points deducted	Took Alternate Tests Avg. No. of points deducted
Sensation	2.76	2.72
Perception	5.14	4.24
Imagination	<u>2.14</u>	<u>2.16</u>
Total	10.31 <u>9.12</u>	9.12
Difference	1.19	

So this method of counting the points deducted from the grades on students' answers may not be regarded as a good measure of the persistence of erroneous assent due to taking true-false tests. This does, however, indicate that the groups were about equal in ability and had mastered the subject matter of the courses to about an equal degree.

Another approach was to determine the average number of actual false statements found in the answers to the essay questions given by each group. This comparison showed that the group which had taken the true-false sub-tests made an average of 5.15 false statements in their answers; while the group which had taken the alternate type sub-tests made an average of only 4.74 false statements. This difference of .41 does not at first sight appear to be significant. But, if the character rather than the number of false

statements be analyzed very significant differences are revealed. The false statements made by those who had taken true-false tests consisted, to a large extent, of echos or equivalences of the false statements they had encountered on these tests. On the other hand, no exact quotation or equivalent paraphrase of any false statement from a true-false test was found among the false statements made by those who had taken the alternate type sub-tests.

The false statements made by the latter group can be easily explained as due to natural confusions between those products and processes of mental life which are very much alike and hence difficult for the beginning student in psychology to distinguish. For example, one of the essay questions was: "State exactly all the differences between percept and image." In answering this question many in both groups stated that the percept is present while the image is absent. The correct answer is, of course, that the percept is a mental representation of a present object; while an image is a mental representation of an absent object or sensible quality. Confusing the presence or absence of the mental representation with the presence or absence of the object is quite natural to a beginner and since no false statement embodying this confusion was contained in any true-false test used it cannot be said that those who did make this false statement were helped to do so.

But on True-False test IV appeared the false statement: "Images differ from percepts because they do not represent sensory qualities." Seven students who had taken this true-false test quoted this false statement exactly in their answers to the essay question concerning the differences between percept and image. Moreover, four other students of the same group gave equivalent paraphrases of the same false statement in their answers.

No one who had not taken this true-false test made any statement which could be construed as an echo or equivalence of the false statement from the true-false test. This is the best evidence that false statements did have a detrimental effect.

Another example of the same thing is found in the answers to another of the essay questions. The question was: "How would you prove that man possesses a special power of perceiving, or synthetic sense?" In answering this question eight students who had taken true-false test IV included in their answers partial equivalences of a false statement that had been used in this true-false test. The statement was: "Perception involves the synthesis of various sensory qualities in such a way that we become aware of objects existing only in the mind." These students would advance arguments similar to the following: "Man is aware of objects that exist only in his mind; so he must have the power to synthesize them." On the same true-false test there was another false statement having a connection with the matter of the same essay question. This statement was: "Synthesis in perception consists of separating present from past sensory impressions." In their answers to the essay question, ten students made partial use of the false notion contained in this statement. They would argue that since man does separate past from present sensory impressions he must possess a synthetic sense power. A total of eighteen students, then, who had taken true-false test IV revealed that two false statements contained thereon did result in erroneous learning; while not one student who had not taken this true-false test made any error that was similar to either of the false statements on the true-false test.

Other examples could be cited if need be of the deleterious effects of false statements from true-false tests, that were discovered in the answers to the essay questions given by those who had taken these true-false tests. In all, forty such clear instances were discovered. Perhaps more could be found if one were more liberal in his judgments of equivalence. At times it is extremely difficult to judge whether a false statement made by a student in his answer to an essay question is really meant to express the same notion contained in a false statement on a true-false test to which the student had been exposed. There are, at times, so many ways of saying the same thing that one cannot always be certain that his interpretation of the meaning of a student's statement is correct. Then too, the student may have had in mind an idea equivalent to one of the false notions in a true-false test, but not having the required vocabulary at his command was not able to express his false idea in such a way that it would be apparent. The reader, therefore, even though recognizing the student's statement as false could not be certain that it was equivalent to a false statement on a true-false test.

Nevertheless, even forty certain echoes and equivalences of false statements from true-false tests made by a group which was exposed to the tests and not made by the group not so exposed, is at least suggestive evidence that the false statement does interfere with the students' mastery of subject matter and that assent to false statements does persist.

Chapter V

SUMMARY AND CONCLUSIONS

This investigation was undertaken to determine experimentally whether assent to false statements on true-false tests persisted to a measurable extent. One hundred and thirty students taking their first course in general psychology served as subjects. The total body of subjects was divided into two groups as nearly equal as possible in numbers, ability, etc. and an attempt was made to insure that each group had the same instruction and opportunities for learning. Four sub-tests, each concerned with a particular portion of the course matter, were given both groups. Each sub-test was made up in two forms, a true-false form and an alternate form consisting of completion, multiple choice and yes-no items. The type of sub-test given each group was alternated every time so that those who took the first true-false test took the second alternate type test, and so for each subsequent test. This was done to insure that the type of test used was the only independent variable. The average percentage of correct responses was used as the basis of measurement throughout.

At the end of the course a final examination consisting of the most valid true-false and other type items used in the sub-test and containing besides ten essay questions was given to both groups. Six attempts were made by different methods to measure the effects the false statements of the true-false sub-tests had upon the students' answers to the items of the final examination.

As measured by the comparison of results on the true-false parts of

the final examination the difference between the group which had taken all true-false sub-tests and the group which had taken all alternate type sub-tests was not significant. The former group made a total average percentage of 66.44 correct responses; while the latter group made a total average percentage of 67.45 correct responses. This difference of 1.01 is hardly significant.

On those parts of the final examination made up of objective type test items other than the true-false, the subjects who had taken the true-false sub-tests made a total average percentage of 60.06 correct responses; while those subjects who had taken the alternate type sub-tests made a total average percentage of 60.87. Again, this difference of .81 does not appear to be significant.

The question naturally arises as to why these differences between the final examination results of the two groups were not larger. There are two possible answers. It may be that the erroneous assent to the false statements did not leave a measurable effect upon the subjects exposed to these statements; or it may be that the effect left is not measurable by the devices employed. Evidence from actual false statements made by students in their answers to the essay questions indicates that the latter is the more probable alternative.

In comparing the average number of points deducted for errors in the answers to the essay questions of the final examination, it was found that those who had taken the true-false sub-tests lost an average of 3.43 points each, and those who had taken the alternate type sub-tests lost an average of 3.04 points. This difference of .39 merely indicates that each group

made approximately the same number of errors. It tells nothing concerning the kind of errors made. Therefore, it was decided to determine the average number of actual false statements made by each group. The group which had been exposed to true-false tests made an average of 5.15 false assertions in their answers to the essay questions. Those who had not taken the true-false sub-tests made an average of 4.74 false statements in their answers. This difference of .41 is not statistically significant. But, when the actual false statements made by each group were examined it was found that the group which had taken the true-false sub-tests made a total of forty false statements which were clearly quotations and paraphrases of false statements they had encountered on the true-false sub-tests. On the other hand, those who had not taken the true-false sub-tests made no false statements which could be construed as echoes or equivalences of any false statements appearing on a true-false test.

It is not reasonable to suppose that the students who made false statements the same as or equal to those on the true-false tests could have independently constructed these statements. The ideas and terminology in the two cases were too much alike. And, the fact that the students who had not taken the true-false sub-tests did not make the same false statements in their essay answers suggests strongly that the false statements on the true-false tests did have a detrimental effect on those exposed to them.

To the question as to why there were not more than forty echoes and equivalences of false statements found it may be answered that the essay questions were not designed to bring them out. Not one essay question concerned the introductory matter of the course nor the treatises on emotions, instincts, innate equipment and the nervous system. All of these topics had

been subject to false statements on the earlier true-false tests. The essay questions were intended primarily to cover the latter portions of the course matter on which students had not been tested by the previous sub-tests. That so many false statements like those on the true-false sub-tests did show up in the answers to these essay questions was surprising as well as significant.

The only remaining difficulty is the question of why those not exposed to false statements on true-false tests made as many, albeit different, false statements in their answers to the essay questions. From the facts available concerning the students and from the test data no ready answer to this question emerges. It does not seem likely that the other types of objective tests used did, like the true-false tests, interfere with the mastery of subject matter by the students; so, the question must be left open.

It may, then, be concluded that the main problem of this investigation is not definitely solved by the methods employed. The differences between the groups in the average percentages of correct responses on the objective parts of the final examination, although positive and in favor of the group which did not take the true-false sub-tests, are too small to be significant. Perhaps better designed sub-tests and final examination would reveal greater differences. But, as seems plausible from this study, another method in which the final examination consisted entirely of essay questions; or, at least, contained many more essay questions than the one employed in this experiment, would better reveal the persistence of erroneous assent due to taking true-false tests.