

Summer 1963

## Cheating: A Comparison of its Incidence in Self-Scoring and on a Paper and Pencil Test at Fort Hays Kansas State College

Norman Simmons  
*Fort Hays Kansas State College*

Follow this and additional works at: <https://scholars.fhsu.edu/theses>



Part of the [Psychology Commons](#)

---

### Recommended Citation

Simmons, Norman, "Cheating: A Comparison of its Incidence in Self-Scoring and on a Paper and Pencil Test at Fort Hays Kansas State College" (1963). *Master's Theses*. 822.  
<https://scholars.fhsu.edu/theses/822>

This Thesis is brought to you for free and open access by the Graduate School at FHSU Scholars Repository. It has been accepted for inclusion in Master's Theses by an authorized administrator of FHSU Scholars Repository.

105

CHEATING: A COMPARISON OF ITS INCIDENCE IN  
SELF-SCORING AND ON A PAPER AND PENCIL TEST AT  
FORT HAYS KANSAS STATE COLLEGE

being

A Thesis Presented to the Graduate Faculty  
of the Fort Hays Kansas State College in  
Partial Fulfillment of the Requirements for  
the Degree of Master of Science

by

Norman Simmons, B.A., B.D.  
Southwestern College, Winfield  
Southern Methodist University

Date

7-18-63

Approved

William F. Jensen  
Major Professor

Approved

Ralph L. Cochrane  
Chairman, Graduate Council

## Table of Contents

	Page
INTRODUCTION . . . . .	1
Survey of Literature . . . . .	3
PURPOSE. . . . .	11
Hypothesis . . . . .	11
PROCEDURE. . . . .	12
Subjects . . . . .	12
Testing Instruments. . . . .	14
Experimental Procedure . . . . .	14
An Operational Definition. . . . .	19
Analysis . . . . .	20
RESULTS. . . . .	22
Problem I. . . . .	22
Paper and Pencil Test. . . . .	22
Table 1. . . . .	23
Self-scoring . . . . .	24
Problem II . . . . .	24
Other Analysis . . . . .	26
DISCUSSION AND CONCLUSIONS . . . . .	28
SUMMARY. . . . .	33
REFERENCES . . . . .	35
APPENDIX A . . . . .	37

Page

APPENDIX B . . . . .	38
TABLE 2. . . . .	41
TABLE 3. . . . .	42

## Introduction

The perennial problem for students of every generation is the problem of cheating or cribbing on assignments and examinations. Many people consider themselves capable of understanding and partially solving the problem, hence a great flood of literature is continually being written over the years. Much of the literature is repetitious as persons from different localities and different college generations say essentially the same things. However there has been much said which has been significant. It is of crucial importance that continuing research in this area be done with as many new techniques as can be devised.

Despite the mass of articles on cheating, the majority pertain to the opinions of persons as to the causes and solutions of the problem. There is a shortage of competent research published which either demonstrates that cheating takes place or is able to study its many motivational forces. It is proposed in the present investigation to study the responses of college students when given the opportunity to cheat. One reason for the lack of responsible research is the difficulty in determining when cheating is actually taking place. New and better methods of studying the problem are in need of discovery. It is proposed in this experiment to utilize one method of studying the incidence of classroom cheating which has by now become almost standard among researchers. This method will be compared with a new test which has not, to this writer's knowledge, been formally published in the literature. However, the

fundamental idea of this new method has been utilized in a previous study. This will be a new innovation.

## Survey of the Literature

One might group the host of literature on cheating into three broad categories. First, there are those articles which are concerned primarily with the motivational factors in cheating, which consider the question of why students cheat. The second category included in the literature is that which raises the question of what can be done about cheating as seen by various persons. Often these are "popular" articles that do little more than moralize and point the finger of responsibility. We shall exclude them from consideration here. Finally, there are a number of studies which attempt to examine the prevalence of cheating. The consideration of to what degree students cheat will be the prime focus of the present study.

Since the primary concern in this study is the incidence of classroom cheating among college students, the major part of this survey will be limited to studies in this area. But first a brief statement should be made regarding the first category.

There are many articles on motivation for cheating. One of the pivotal studies is that of Hartshorne and May (1928) who studied hundreds of children experimentally. It was found that honesty seems largely to be a function of and dependent upon the actual situation rather than a generalized moral trait which some people have and others do not. Hartshorne and May presented the children with a problem so difficult - a complicated finger maze to be traced with the eyes closed - that a child had to peek or "cheat" to obtain a high score on it. This is the same basic principle to be employed

in using the new paper and pencil test. Carlson (1935) among others, places much responsibility for cheating on the teacher's attitudes. In a study done by Campbell (1933,) he suggests that cribbing is prevalent due to the ease with which cheating can be accomplished. This is demonstrated in several experiments where students grade their own exams. Chidester (1958) and Brownell (1928) observe that students of lower intelligence have greater incidence of cheating. Chidester (1958), Fowler (1960) and McQueen (1957) found that social status affects cheating.

In a typical article Crawford (1952), a professor of education at the University of Southern California, gives his views on the causes of student cheating. The reasons given are desire for approval, recognition, praise for brightness, admission to college or graduate school, athletic eligibility or just plain laziness. Contributing factors were seen as lack of interesting and vital curriculum, over emphasis on marks and grades, and the fact that adults cheat. In typical style, the author places much responsibility for cheating upon the teacher.

One of the problems with measuring motivational factors in cheating is that ordinarily many of these motivations operate concomitantly. To say, for instance, that social status or intelligence affects cheating perhaps does not fully take into account the whole realm of forces at work motivating the student. It is virtually impossible to isolate separate aspects to see which are primary for



all factors are present simultaneously. Gross (1946), in an interesting study of the effect of self-competition and group-competition on achievement and honesty found that motivation did not appear to increase dishonesty nor did one kind of motivation give evidence of superiority to the other in promoting achievement. In brief, it can be said that there are many opinions as to the motivations of cheating. It shall be left to other studies to verify the authenticity of these causes. It is an area of research all of its own.

It has been stated that one reason for the lack of competent research in the area of cheating is the difficulty in determining when cheating is actually taking place. It is proposed in this study to make some contribution to the research in this area. To begin with a review of some of the methods commonly used in ascertaining the extent of cheating will be helpful.

Campbell (1933) was successful in using a "spy system." Spies interspersed in the classroom recorded cheating when they heard fellow students exchanging questions and answers or when they saw a student use materials during the examination. With this method, 96 of 173 college students (56%) were discovered cheating or later admitted to it.

Weinland (1948) used two forms of an examination to study cheating. Students in alternate seats received Forms A and B. The two forms were partially alike and partially different. For example, questions 1-4 were the same on both forms but questions 5-8 were different.

The assumption was that when a student looked on his neighbor's paper, cheating would be easier on the questions that were the same and the semester grades should be higher on the easy-to-crib questions. Weinland found, however, that for a class of thirty eight college students, the average score for those numbers where the questions on the two forms were the same was slightly lower than when the questions were different. He offered several possible conclusions, the main one being that if cheating via copying took place, it was not effective in raising grades.

Saupe(1960) has developed a statistical method which involves determining the extent of correspondence between a pair of answer sheets concerned in a case of suspected cheating. He calls it "an empirically based model for the corroboration of suspected cheating on multiple-choice tests." This is a useful study, however it is limited to situations where cheating is suspected and is hardly practical for widespread application to all students simultaneously in large college classes.

Another method used to study student responses to opportunities to cheat is intentional grading errors. Campbell (1931) at Stanford University returned test papers which had been graded too low. 97% of the students reported every error. When the test papers were returned with the grades intentionally too high, 65.7% kept six or more grade points in their favor. It was found that the number of points kept varied inversely with the students' mastery of the content of the course.

In a similar study, Krueger (1947) intentionally made grading errors on student papers. When the grades were to the students' advantage 90% of them left the grade too high while 10% lowered their grade. When the grading score was to the students' disadvantage 95% raised their grade while 5% left their grade too low. After the students had been informed concerning the purpose of the experiment nearly 99% made all changes required.

A most common method of checking cheating is the questionnaire. Adams (1960) and Greaves (1953) both made use of the questionnaire asking students if they did cheat and if so to give their reasons. In the latter study the questionnaire sent out to students revealed a large group who did not cheat and resented those who did, a small group of inveterate cheaters, and a large middle group who gave assorted excuses to justify occasional dishonesty.

Another questionnaire was developed by Anderson (1957) and used again by Frymier (1960) asking students to rate hypothetical cheating situations as to the degree to which they feel the situation is cheating. This study indicated that a hierarchy of what constitutes cheating and what does not exists in students' minds. Females are more strict than males, graduate students more strict than undergraduates. Frymier found that faculty members were more severe than students in their views of what constitutes cheating.

In an interesting study Freeman and Ataov (1960) ranked thirty eight students in terms of observed cheating. They were then

questioned both directly and indirectly about cheating and the results were correlated. Since all correlations were insignificant, the results of this study cast doubts upon the validity of indirect and direct measures of attitudes toward cheating. Thus it calls into serious question the use of questionnaires for determining attitudes toward cheating. What people practice and what people profess seem to be two different things.

One of the commonly accepted notions among teachers is that if you give students their own examination papers to score, many of them will deliberately alter their answers or misscore them in order to get a better grade. This notion holds in it the central thesis of the final method used to measure classroom cheating. It is proposed to use this method in the present experiment. This technique, in which the test is given and then the answer sheets are copied by the teacher and returned to the students for self-scoring, seems to be the most persistently used method for detecting cheating.

This method has been criticized by Wrightsman (1959) on two counts. First, it does not study cheating "on the test itself." Secondly, the users of this technique often generalize their results to say that cheating on the test itself is as prevalent as in the self-scoring. Wrightsman contends that such a generality has yet to be demonstrated to his satisfaction. Both of these criticisms seem just and cannot be ignored.

Many different percentages of students cheating have been discovered by those using the technique of students scoring their own previously graded exams. Campbell (1933) found 45% changed scores in a large state university in the South. Chidester (1958) discovered 54% of her grade school children changed answers. Atkins and Atkins (1936) using the same method among college students who were prospective teachers found 50.9% made one to fourteen alterations, 42.9% made one or two alterations, and 7% made more than ten. They observed that the brighter students were not likely to cheat. Perhaps the most realistic study was done by Yepsen (1927). He found 29.5% and 24.5% of the students in two different studies accepting the opportunity to raise their own grades.

There are several possible explanations for the discrepancy in percentages between the Yepsen study (29.5% and 24.5%) and that done by the two Atkins (50.9%). The Yepsen test was only a five minute test of literacy and thus in terms of total items and time spent grading gave less opportunity to cheat. The Atkins and Atkins study was a General Information Test in which everything possible was done to enhance the opportunity to cheat. Resistance to altering questions was decreased by placing stress on the importance to a teacher of a large fund of correct general information. Nothing was said regarding honesty. Students were instructed during the test period to leave blanks if they were in doubt giving opportunity to fill them in during grading. The students were also directed to circle every blank and

every incorrect response - the idea being to get them to use their pencils. Their papers were on their desks the entire class period during grading. Self-checking took place from a list of correct answers which had been placed on the blackboard.

Another important aspect of this study was that a paraffin sheet duplicating the work was attached beneath the exam and was removed before self-scoring so as to prevent grading errors on the part of the experimenter. Campbell (1933) used a control group for his study to control grading errors. The students in this group were told before the self-grading that an accurate copy of their papers had been made. Since an earnest appeal for accuracy had been made to the experimental group, Campbell says it is questionable whether the motivation for accuracy was any greater between groups. In any case actual changes in scoring by the student was judged unethical behavior.

Thus we have seen at least five methods to ascertain the extent of classroom cheating. These have included using a spy system, two separate forms of the exams, intentional grading errors, questionnaires, and self-scoring of previously graded exams. It is proposed to use this last technique as one measure of cheating in the present study. The advantage in the plan to be used here is that actual exams counting toward final semester grades will be employed.

## Purpose

Problem I The problem is to measure the incidence of cheating among college students.

Hypothesis Students given the opportunity to cheat will do so.

Problem II To determine whether a new method of testing the extent of cheating is a valid assessor of classroom cheating.

Hypothesis Given adequate motivation as many students will cheat on a simple performance test as will cheat on an exam counting toward a semester course grade.

## Procedure

### Subjects

All Ss came from college classes at Fort Hays Kansas State College during the summer session. A General Psychology class was used as the experimental group. This consisted of forty two students with a median age of nineteen. The group included thirty women and twelve men. Since all Ss are required to take General Psychology this group was a representative sample of the school population. Two sets of control groups were utilized. The original plan was to use two Human Growth and Development classes as one control group. One class met in the morning, the other in the afternoon of the same day. These classes would have employed fifty four students with a median age of 22.8. It would have included eighteen men and thirty six women.

However, the plan to have just one control group had to be changed on the day the tests were administered when it was discovered that there was inadequate control in the performance test<sup>\*</sup> given to the first class of Human Growth students. Cheating was taking place despite planned controls. An additional control factor had to be introduced. This meant that additional subjects had to be secured. It was felt at this point that homogeneity of subjects might need to be sacrificed for the sake of good control.

\* Paper and Pencil Test and Performance Test will be used interchangeably in this study.



After the additional control factor was introduced for the performance test, the following subjects were secured as a control group for this phase of the study. The afternoon Human Growth class was used in the formation of this new control group. This class consisted of eighteen students. Two of these students had to be disqualified because of professed knowledge of the paper and pencil test and one student was disqualified because he was a member of the experimental group. A class of eleven students in Atypical Child was secured and one of these students was disqualified because of previous knowledge of the test. Still another class in Applied Statistics was utilized. This consisted of fourteen students. Two of these students were disqualified on the basis of a professed understanding of the nature and purpose of the test.

These three classes together comprised a control group of thirty seven students of whom eighteen were women and nineteen were men. The median age for this group was twenty five, compared with that of nineteen for the experimental group. The range in age in the control group varied from nineteen to fifty, with thirteen students over thirty years of age. Some of these subjects are functioning teachers returned for additional college work. More than half of these students are then not representative of the typical college student population.

Since all tests were administered in one day and at no time during this period was an explanation made concerning the purpose of these tests, there is little possibility of any subjects becoming newly

informed about the study. The reason five students were aware of the study was because a pilot study was done the previous summer.

### Testing Instruments

The regular multiple choice tests covering a portion of the textbook and counting toward the final course grade were administered. Multiple choice tests are given regularly in both General Psychology and Human Growth. These exams were utilized in that part of the study where the tests were previously graded by the experimenter and then handed back for the students to grade and alter as they chose. Students in Atypical Child and Applied Statistics did not participate in this phase of the study.

The second instrument was a standard sheet of ditto paper with seven circles distributed on the page in an irregular manner (see Appendix A for sample.) These were dittoed and administered to all Ss. The Ss were asked to place an X in each of the circles either blindfolded and in a darkened room (control group) or with their eyes shut (experimental group.) The experimental group was not prevented from opening their eyes a bit to assist them in the task to be performed if they wished.

### Experimental Procedure

Multiple choice examinations were administered to all Ss in the General Psychology and Human Growth classes. On the day of the test the professor or his proctor promised the class he would have the papers graded the following class period. The students were also

given notice that a short performance test would be given with no preparation being necessary. This was to insure attendance at the class sessions.

The experimenter took the exam answer sheets on the day of the test and photostated every one. The photostating worked so clearly that even the erasures could be seen on the photostat copy. No marks were placed on the papers to be returned to the students for self-scoring the following day. The experimenter graded every photostat twice making every effort to prevent grading errors. If there had been some grading errors they could have been discovered later during additional checkings, however no errors were found after the initial grading. The use of photostatic copies provided a "built in" control for the examination papers. No changes, alterations, or failures to report errors could go undetected once a copy of the original was made.

On the day following the examination each student received his own paper to grade with the explanation "I just didn't get time to grade your papers from yesterday as I promised so you can grade them yourselves. It will save me some time."

The professor then wrote the correct answers on the board with his back to the class. This gave the student considerable opportunity to alter some answers if he desired. The student was instructed to place a large X by each wrong answer. Upon completion of writing the answers on the board, the professor told his class that he had

forgotten to bring the other test to be given that day and would have to go to his office to get it. (This was actually true in the case of the professor of the experimental group.) It had been agreed to give the paper and pencil test first but this had to be reversed when the tests were forgotten. It was perhaps fortunate it occurred this way for it gave the professor an excuse to leave the room. It had been planned that he should leave the room but it had not been planned as smoothly as it actually occurred. No problem was presented by this change because the experimental group was the first group to take the test and exam and thus the change could be made easily for the other two classes.

Upon the professor's return to class (he was to be gone five minutes) he answered any questions the students had about the test. Then he instructed them to count the number wrong and place it circled on the front of the exam answer sheet. At no time were the students to be rushed in this part of the experiment.

Another aspect of the study not yet mentioned is that pencils were provided the day of the examination with the explanation that answers could be more easily changed if they were written in pencil. Pencils were also available the day of the self-scoring ostensibly for the paper and pencil test, but actually in order that answers could be changed on the exam. It was felt that without a pencil the chances of a student changing an answer to his benefit would be considerably narrowed. On the day following the examination the student was given every opportunity to cheat.

It might be mentioned here that cheating on the actual examination the first day was cut to a minimum by the use of plastic cover sheets over the top of the exam answer sheets. The use of these plastic cover sheets is standard procedure in the Fort Hays Psychology Department on multiple choice questions and is a deterrent to possible cheating. These plastic sheets are about  $\frac{1}{4}$  of an inch thick and are perforated in such a manner that the student can mark any of five possible answers while preventing his neighbors from viewing his answers.

The second phase of the study took place the same class period as did the self-grading so there would be little time for suspicions to be aroused. The professor handed out the performance test face down of the Ss desks. Blindfolds were distributed to all in the class in the control group. Both experimental and control groups were asked to write on the back of the tests their name, grade average, major, I.Q. if known, age, and occupation if other than student. Their I.Q. and grade average were requested to convey the idea that there was an interest in relating the results of the test to these measures.

Then the following statement designed to enhance motivation was read by the professor:

Prior evidence indicates this is a valid test of perceptual, visual imagery. Thus it tests this phase of one's intellectual ability and correlates with one's overall I.Q. or Intelligence Quotient. It has been shown that there is a high level of correlation between the functioning on this perceptual task and one's academic grade. For example, I will find a close relationship between the ability to properly perform this task and the grade you would receive in this class.

Following the reading of these instructions the professor asked the students to notice carefully the position of each circle on the page. Then members of the experimental group were asked to take up their pencils, turn the page over, place their hands in their laps, and begin to study the position of the circles on the page. They were instructed to have their hands in their laps so they could not use their fingers to mark the place of the circles or some other such method of measurement. After two minutes of study had elapsed, the professor said "Tims is up, shut your eyes and place an X in each circle" (a pause of 6-8 seconds). "You may hand in the test when you are through." (See Appendix B for exact instructions given to the professor.)

Those in the control group were asked to turn their papers over, study the position of the circles on the page while keeping their hands in their laps. Following two minutes of study the control group was instructed to put on blindfolds. They were to hand in their tests before they took off their blindfolds.

A statement should be made here regarding the problem of inadequate control with the morning Human Growth class and the introduction of a new control factor for all subsequent classes. It became apparent when the results of the test were quickly appraised that several students had achieved an impossible number of Xs inside the circles and that many more students got Xs in the circles than could be realistically expected in a good control situation. It was clear

then that cheating was occurring in the control group. This necessitated the introduction of an additional control factor.

The problem was solved very simply. The tests were being administered in two basement rooms. One of these rooms had heavy drapes and the other had blackout curtains, for they were both used on occasion for audio visual purposes. By turning out the lights and pulling the shades the rooms were darkened so that it was impossible to see the circles on the page without a blindfold. However, as an added precaution blindfolds were used as well as the darkened room. The test papers were then collected before the lights were turned on in the room. Thus the more reliable control factor used was the darkened room.

After all the testing was completed the experimenter checked the exam answer sheets three times for alterations or discrepancies in student grading. The performance tests were separated and a panel of two faculty members served as judges reviewing all questionable near hits or misses. The results were then compiled for analysis.

#### An Operational Definition

An operational definition of cheating had to be established so some criteria for measuring the cheating could be set. In the paper and pencil test every S who had two or more Xs inside the circles on the test were considered to have cheated. This criterion was first established from a pilot study where only two or twenty two subjects got as many as two hits inside the circles, under control circumstances.

The present study seemed to further verify this judgment with only four of thirty seven controlled subjects each making two hits in the circles. Any X whose axis was in the circle or clearly on the line was viewed as a hit by the judges.

The operational definition of a "cheater" on the examination was originally to have been every S who altered one answer while grading the exam. However the ingenuity of the students was not fully taken into account. For example, three students changed several answers in their favor but they also failed to check several answers wrong. In fact ten students missed checking wrong answers which would have lowered their grades. One girl out of the ninety six students grading the exam actually checked two too many wrong thus grading her test too low. It could be argued then that these misses were a result of student grading errors. However, of the ten students who missed checking wrong answers, six cheated on the performance test. This makes it seem highly unlikely that their errors in grading were only "accidental."

In the light of these facts it seemed reasonable to make more definitive the criterion of cheating on the exam. Thus the operational definition used was: any discrepancies in grading which were in the students' favor were considered cheating. These criteria enabled the experimenter to evaluate and make an analysis of the results.

### Analysis

A Chi Square was used to compare the results of these two groups. All chi squares obtained in this study were corrected with a modification



known as Yate's correction for continuity. This has the effect of reducing the amount of discrepancy between observed and expected frequency to the extent of .5. This correction is needed because chi square varies in discrete jumps whereas computation by formula gives more continuous variations.

## Results

Problem I

Two measures for ascertaining the incidence of cheating were used in this study. It is now time to consider the results of cheating on the paper and pencil test and on the examination. Table 1 on the following page will give a summary of these results.

Paper and Pencil Test. There were fourteen persons in the experimental group who cheated and four in the control group. The analysis yielded a chi square of 4.76. With one degree of freedom, the obtained chi square is larger than 3.841 which has the probability of .05. It can therefore be regarded as significant and the null hypothesis of no difference between the groups can be rejected. In brief, this reveals a significant difference in the number of cheaters between the experimental and control groups on the performance test. Table 2 in the Appendix presents the contingency table for the above data.

A computation of chi square for the experimental group and the control group, which used only blindfolds, revealed more clearly why more stringent controls were needed. The chi square obtained was only .0102. Thus even though these two groups were not treated similarly, the effects revealed very little difference in the number of cheaters produced. The blindfolded group had eleven of thirty seven cheaters (29.7%) while the experimental group had fourteen of forty three (33.3%).

Table 1  
 Chi Squares of All Comparisons Made  
 in this Study with Probability Levels

Comparison	Chi Square	Probability with 1 df
Control Group and Experimental		
Group on Performance Test. . . . .	4.76	.05
Blindfolded Group with Experimental		
Group on Performance Test. . . . .	.01	not significant
Control Group with Blindfolded		
Group on Performance Test. . . . .	3.01	.10
Experimental Group results on		
Exam and Performance Test. . . . .	7.09	.01
Blindfolded Group		
on Exam and Performance Test . . . . .	.28	not significant
Experimental Group and Human Growth		
classes on the Exam. . . . .	3.32	.10

Another dimension to the problem is revealed when an analysis is made of the results of the two control groups. It is remembered that the first control group was merely blindfolded and these results (29.7% cheating) called for the additional control group (12% cheating) using a darkened room. Both groups had thirty seven subjects. A chi square value of 3.01 was obtained which is only significant at the .10 level. This suggests that the difference between the two treatments approaches an acceptable level of significance.

Self-scoring. There were ninety seven students who took the examinations. One student was in both General Psychology and Human Growth so he was disqualified on his second exam. Of the ninety six students who graded their own papers, there were twenty two persons who had discrepancies in grading in their favor. By this criterion roughly 23% of the students who graded their papers were considered cheaters. Twelve persons or 12.5% of the students made actual changes in their favor. Only 7% of the students had two or more discrepancies in grading in their favor. The largest amount of cheating done by one student was three wrong answers not checked and four changes made.

#### Problem II

The second problem was to ascertain whether the new innovation in testing cheating (the paper and pencil test) was a valid assessor of cheating in the classroom. The way in which this was to be determined was to compare the more established method of self-scoring with this newer technique. If the students who cheated in self-scoring also

cheated on the paper and pencil test, and if those who did not cheat on one did not cheat on the other, then it could be assumed that the new method devised was able to measure cheating in a classroom situation.

A contingency table was utilized which would show the relationship of cheating and non-cheating by the experimental group on both tests. Actually, it enabled the experimenter to study whether the same subject cheated on both, none, or only one of the two tests of cheating. Table 2 in the Appendix shows how the data was studied. A chi square was used to analyze this data.

There were forty two subjects in the General Psychology class who took both the exam and the performance tests. Nine of these persons cheated on both the exam and the performance test. There were five who cheated only on the performance test and five others who cheated only on the exam. This made a total of fourteen instances of cheating on each task. It also makes nineteen persons involved in cheating or roughly 45% of the class.

In computing chi square and correcting for continuity a value of 7.09 was received. With one degree of freedom this is significant at the .01 level for it is larger than 6.63. Thus the chi square is very significant and it can be said that the preponderance of persons in the experimental group responded the same way on both the exam and the performance test. Those who cheated or did not cheat on one tended to respond the same on the other. Nearly half of those who cheated did so on both test and exam. A contingency coefficient was

computed directly from chi square to measure the association of these two tasks. The obtained value was .38 which suggests a moderate correlation. However according to Guilford (1956) the maximal value attainable for a coefficient of contingency with two categories in both variables, which is the case with the exam and performance test, is .707. This being true, it appears that the results here point to a substantial relationship between the two tasks.

Another interesting sidelight of the analysis of results was to use the above method to study cheating in the morning Human Growth class. There was an interest here in comparing cheating on the exam and cheating among those blindfolded on the performance test. In this instance a chi square yielded a value of .28. With one degree of freedom the obtained value is not significant.

It has been predicted that given adequate motivation as many students would cheat on the simple performance task as would cheat on an exam counting toward a semester course grade. In terms of the criteria set up earlier this occurred precisely as predicted in the experimental group. There were fourteen instances of cheating on each of the two tasks.

#### Other Analysis

One of the apparent differences in cheating on the examination was between the number of persons cheating in the General Psychology class and those cheating in the Human Growth classes. Only eight of fifty four (14%) students cheated on the exam in the Human Growth

classes while fourteen of forty two students (33.3%) cheated on the General Psychology exam. A chi square of 3.32 was obtained when a comparison was made between these two findings. With one degree of freedom this is not significant at the .05 level for it is not larger than 3.841 but it does approach significance. It is significant at the .10 level. There are some possible explanations for this which will be discussed later.

Another interesting comparison was the difference in cheating by sex. Perhaps by accident the results figured to 26% cheating by both men and women. There was no difference in cheating by sex in this study.

The number of erasures on the multiple choice exam was counted. Some persons argue that one should not change any answers on tests of this kind for first impressions are ordinarily right. This was not found to be true in this case. Of one hundred and twenty two erasures on the examination, seventy six changes were to right answers and forty six changes were from right answers to wrong answers. The student probably remembers more bitterly those he changed to wrong, thus such a misconception exists.

## Discussion and Conclusions

The reader should be reminded again at this point that what has been measured in this study is not what can be called classroom cheating. What is meant here by "classroom cheating" is that which exists during the taking of a classroom test. Here we have measured cheating in a classroom situation. One must draw an operational distinction between classroom cheating and cheating in the classroom. The motivations are similar. The circumstances may be much the same, but it was mentioned at the outset the difficulties in measuring cheating during the actual classroom exam. One should be careful in making inferences as to the validity of these results for cheating while taking a classroom exam.

This study has attempted to show the prevalence of cheating among college students when opportunity for cheating is given. The results of this study ought to give ample warning to those teachers who may sometimes be careless in the administering of exams. Many students will cheat if the opportunity is offered.

In measuring the incidence of cheating among students, it was discovered that there was a significant difference between the two groups involved in taking the performance test. It seems possible then to account for this difference in terms of the varied treatment given the groups. The purpose was to make the task so hard they would have to cheat to be successful. The independent variable was the amount of opportunity given to cheat. In the control group the



opportunity to cheat was cut to a minimum. In the experimental group the students could open their eyes surreptitiously to cheat. The dependent variable was cheating. Since these students came from the same general population we can say with a reasonable degree of confidence that our treatment effect was successful. Since chance alone cannot account for this difference it must be considered significant that three times more cheating took place in the experimental group than in the control group. Again, many students will cheat when given the opportunity. In this instance one third of them did so. Thus our first conclusion in this study is that many students will cheat when given the opportunity.

Likewise cheating on the examination was just as prevalent. Here again one third of the students in the General Psychology class cheated. However, the overall cheating percentage of all those taking the exam was 22.6%.

One of the questions still unanswered is to account for the difference in cheating between the Human Growth and General Psychology classes on the exam. Were they simply more honest? There are three possible explanations for this. Indirectly some are probably a criticism of the homogeneity of the population. First, the General Psychology classes had more test questions and thus more chances to cheat. There were only thirty two questions on the Human Growth exam while there were forty six General Psychology questions. Secondly, the smaller Human Growth classes could have had some effect

on the amount of cheating. It would be interesting to do further research on whether it is easier to cheat in a large class than in a smaller class where the possibility of detection might be greater. And thirdly, the median age for the Human Growth was twenty five, while the median age for the General Psychology was nineteen. One could speculate that the older students were less naive, more likely to think something "fishy" was going on when the professor asked them to grade their own papers. But whatever the reasons, it is not possible within the framework of this study to give the answers, one can only guess.

The second important facet of this study was to ascertain whether the new innovation in measuring cheating involved in the performance test was a valid assessor of cheating in a classroom situation. The reasoning is that if the students who cheated in self-scoring also cheated on the paper and pencil test and if those who did not cheat on one did not cheat on the other, it could be assumed they were both able to measure cheating similarly. The results as shown in Table 2 in the Appendix seem to bear out this hypothesis. The chi square was significant at the .01 level. It then appears the majority of subjects who cheated or did not cheat did so on both tasks. The second conclusion in this study is that the newly devised paper and pencil test seems as good a determiner of cheating as is the more established method of self-scoring.

Table 2 further demonstrates another prediction made earlier. It was suggested that given adequate motivation, as many students

would cheat on a simple performance task as would cheat on an exam counting toward a semester course grade. There were fourteen students in the experimental group who cheated on each of the two tests.

Another interesting facet worth mentioning is concerned with the relationship of cheating on the exam and performance test where the performance test was given with the students only blindfolded. It is recalled that there were eleven persons who cheated on the performance test because of the inadequacy of the blindfolds. There were seven persons who cheated on the exam. Of these eighteen instances of cheating only one person cheated on both test and exam. Why is this true especially since there was such a consistency in cheating on both tasks in the experimental group? A possible explanation is that those who cheated on the blindfolded task did so because after they put their blindfolds on it was placed in such a manner that they could see a bit at the bottom of the blindfold. Perhaps they did not even plan to cheat. At any rate, it appears this was an instance of cheating being more "situational" and perhaps not even intentional. Again this is only a matter for speculation. This would however, agree with Gross (1946) who found that two-thirds of the children who were dishonest on self-corrections of papers were so on only one of the two days they were given opportunity to cheat. There is much other evidence which points to cheating being "situational." Our present study did not reveal this, however.

Perhaps the last question to be asked is why did the students cheat? What was their motivation - was it just to succeed? As we mentioned at the outset, this was not a concern of this study. However it does seem like a fascinating and challenging area for further research.

### Summary

The literature contains at least five methods to ascertain the extent of classroom cheating. These include using a spy system, two separate forms of the exam, intentional grading errors, questionnaires, and self-scoring of previously graded exams. It was proposed to use this last technique to measure cheating in the present study and compare it with a new method which gave the student a problem so difficult that he would be motivated to cheat to obtain a high score.

The problem as stated was twofold. First, to measure the incidence of cheating among college students and secondly to determine whether a new method of testing cheating is a valid assessor of cheating in the classroom. The students were given an exam counting toward a final semester grade. The experimenter graded these exams without placing any marks on the answer sheets. The students were asked to grade their own papers and then discrepancies were checked. The students were also given a performance test requiring them to place Xs in seven circles with their eyes closed or in the case of the control group, blindfolded and in a darkened room. The students in the experimental group who were to close their eyes were allowed to cheat if they so desired.

It had been hypothesized that students given the opportunity to cheat would do so. This was conclusively confirmed when it was discovered that one third of the students cheated on the performance test and about one fourth of all students taking the exam cheated.

These were different individuals in some cases. Roughly 45% of the experimental group cheated on one of the two tasks. One of the important statistical findings was that there was a significant difference between the number of cheaters in the experimental and control groups taking the performance test.

A second important conclusion was that the new method of testing cheating was a valid assessor of cheating in the classroom. The comparison between the two methods yielded a chi square which had a probability of .01. This indicated that those who cheated or did not cheat on one task did the same on both. The hypothesis that as many students would cheat on the performance task as cheated on the exam was confirmed precisely with fourteen students in the experimental group cheating on each task.

This has been an interesting and challenging study. Perhaps it lifted up some possibilities for further research.

## References

- Adams, H.R. Cheating--situation or problem? Clearing House, 1960, 35(4), 233-5.
- Anderson, W.F. Jr. Attitudes of university students toward cheating. Journal of educational Research, 1957, 50(8), 581-8.
- Atkins, B.E. and Atkins, R.E. A study of the honesty of prospective teachers. Elementary school Journal, 1936, 36, 595-603.
- Brownell, H.C. Mental test traits of college cribbers. School and Society, 1928, 27, 704-764.
- Campbell, W.G. Measurement in determining the personality and behavior of the college cribber. Education, 1933, 53(7), 403-408.
- Campbell, W.G. Student honesty as revealed by reporting of teacher's errors in grading. School and Society, 1931, 33(838), 97-100.
- Carlson, H.S. Teacher's attitude in relation to classroom morale and cheating. Journal of experimental Education, 1935, 4, 155-213.
- Chidester, Louise. Why do they cheat? Ohio Schools, 1958, 36(8), 25.
- Crawford, C.C. Why children cheat. Education Digest, 1952, 17(6), 9-11.
- Fowler, Mary E. Are we teaching our children to cheat? National Parent-Teacher, 1960, 54(8), 28-30.
- Freeman, L.C. and Ataov, T. Invalidity of indirect and direct measures of attitude toward cheating. Journal of Personality, 1960, 28(4), 443-7.
- Frymier, J.R. Faculty and student perceptions of cheating. Journal of educational Research, 1960, 54(3), 118-20.
- Greaves, J. The honor you save. National education association Journal, 1953, 42(5), 292.
- Gross, M.M. The effect of certain types of motivation on the honesty of children. Journal of educational Research, 1946, 40(2), 133-140.
- Guilford, J.P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill, 1956.
- Hartshorne, H. and May, M.A. Studies in Deceit. New York: Macmillan, 1928, 3 vols.

- Krueger, W.S.F. Students' honesty in grading errors. Journal of applied Psychology, 1947, 31(5), 219-21.
- McQueen, R. Examination deception as a function of residual, background, and immediate stimulus factors. Journal of Personality, 1957, 25(5), 643-50.
- Saupe, J.L. An empirical model for the corroboration of suspected cheating on multiple-choice tests. Educational and psychological Measurement, 1960, 20(3), 475-89.
- Weinland, J.D. Attempt to measure cribbing in objective examinations. Journal of educational Sociology, 1948, 21, 97-100.
- Wrightsman, L.S.Jr. Cheating-a research area in need of resusciation. Peabody Journal of Education, 1959, 37(3), 145-9.
- Yepsen, L.N. The reliability of self-scored measures. School and Society, 1927, 26(673), 657-60.



## Appendix B

(Copy of the instructions to professor as was actually given to him.  
For any changes see procedure.)

I Paper and Pencil Test

1. "First of all, we'll take this paper and pencil test."  
Hand out circle test face down on desks.
2. Pass out pencils to all who need them.  
Pass out the blindfolds to all in the control group.
3. Ask to place on back of test:  
name  
grade average if known  
major  
I.Q. if known  
age  
**occupation if other than student**

Then the following statement to enhance motivation:

Prior evidence indicates this is a valid test of perceptual, visual imagery. Thus it tests this phase of one's intellectual ability and correlates with one's overall I.Q. or Intelligence Quotient. It has been shown that there is a high level of correlation between the functioning on this perceptual task and one's academic grade. For example, I will find a close relationship between the ability to properly perform this task and the grade you would receive in this class.

Now when I say begin, you should notice carefully the position of each circle on the page.

## Experimental Group      General Psychology

1. "Now take your pencils."  
"Turn the page over."  
"Place your hands in your lap."  
"Begin your study of the position of the circles on the page."
2. (Time them two minutes.) "Time is up." "Shut your eyes and place an X in each circle." (Pause 6-8 seconds.) "You may hand in the test when you are through."

## Control Group      Human Growth

1. "Now turn your paper over and study the position of the circles on the page - keep your hands in your lap. (Time them two minutes.)"
2. "Time is up." "Place your blindfolds on. Turn lights out. Now take your pencils and place an X in each circle." (Pause 6-8 seconds.) "Please hand in your test papers before you remove your blindfold."

II Grading Test (Classroom type)

1. "I just didn't get time to grade your papers from yesterday as I promised so you can grade them yourselves." "It will save me some time."
2. Hand out papers to the owners.
3. Professor will write answers on the board with back to the class.
4. Instruction: "Place a large X by each wrong answer."
5. Then tell class you have to go to office for about 5 minutes-- (professor told them he forgot the other test to be given that day and he would have to go to his office to get it.) --they should spend this time checking answers and looking over their papers.

6. Tell class to count number wrong and place it circled on the front of the exam answer sheet. Questions and answers over any of the questions.
7. At no time rush the students in this part of experiment.

Table 2  
Contingency Table for  
Performance Test

	Cheaters		Non-cheaters		Totals
	Observed	Expected	Observed	Expected	
Experimental	14	9.57	28	32.43	42
Control	4	8.43	33	28.57	37
Totals	18		61		79

$$\chi^2 = 4.76 \text{ and Probability of } .05$$

Table 3

## Contingency Table for Exam and Test

## Exam

	Cheaters		Non-cheaters		Totals
	Observed	Expected	Observed	Expected	
Cheaters Performance Test	9	4.67	5	9.33	14
Non-cheaters	5	9.33	23	18.67	28
Totals	14		28		42

$$\chi^2 = 7.09 \text{ and Probability of } .01$$