

# EXAMINATIONS

## Their Proper Design and Use



A Teaching - Learning Symposium for Faculty

College of Agriculture

Kansas State University

*Rock Springs Ranch*

*September 4-5, 1968*



DEAN, COLLEGE OF AGRICULTURE, WATERS HALL 117  
DIRECTOR, AGRICULTURAL EXPERIMENT STATION, WATERS HALL 113  
DIRECTOR, INTERNATIONAL AGRICULTURAL PROGRAMS, WATERS HALL 14  
DIRECTOR, KANSAS EXTENSION SERVICE, UMBERGER HALL 122

October 14, 1968

To: Faculty, College of Agriculture, and Other Interested Faculty and Administrators

Dear Colleagues:

Much too frequently in the last few years, U.S. newspapers have carried headlines reporting mass student rioting on our campuses. Students report that they riot because of being alienated by faculty and administration and because quality instruction and student advising is nearly disregarded. It is quite possible that dissent is justifiable on many campuses. But, so often the news media fail to report genuine efforts by faculty and administrators to improve the quality of classroom instruction, to improve student advising programs, and to involve students in policies affecting the quality of their educational experiences. These proceedings are a good example of such efforts.

The attached proceedings exemplify the kind of concerns and genuine interest of the faculty of our College of Agriculture as they critically examined our examinations, their design and administration. Cognizance of the need for student participation was demonstrated by including a lead-off panel of five College of Agriculture students who spoke frankly to the question: "What is wrong with the design and administration of examinations at Kansas State University?" The discussion following the student statements exemplifies a mutual concern by KSU faculty and students in this topic.

The remainder of the symposium dealt with the questions of "Why Test," "Kinds of Tests," "How to Design Better Tests" and finally, "Psychological Impacts of Tests on Students." Students, faculty and resource persons interacted freely in both small-group and plenary sessions.

This is the second such symposium. In September, 1967, approximately 80 of our agriculture faculty retreated off campus for a two-day symposium entitled, "New Insights into Your Performance as a Teacher." Appropriately, the topic this year centered on examinations, a question that troubles both faculty and students.

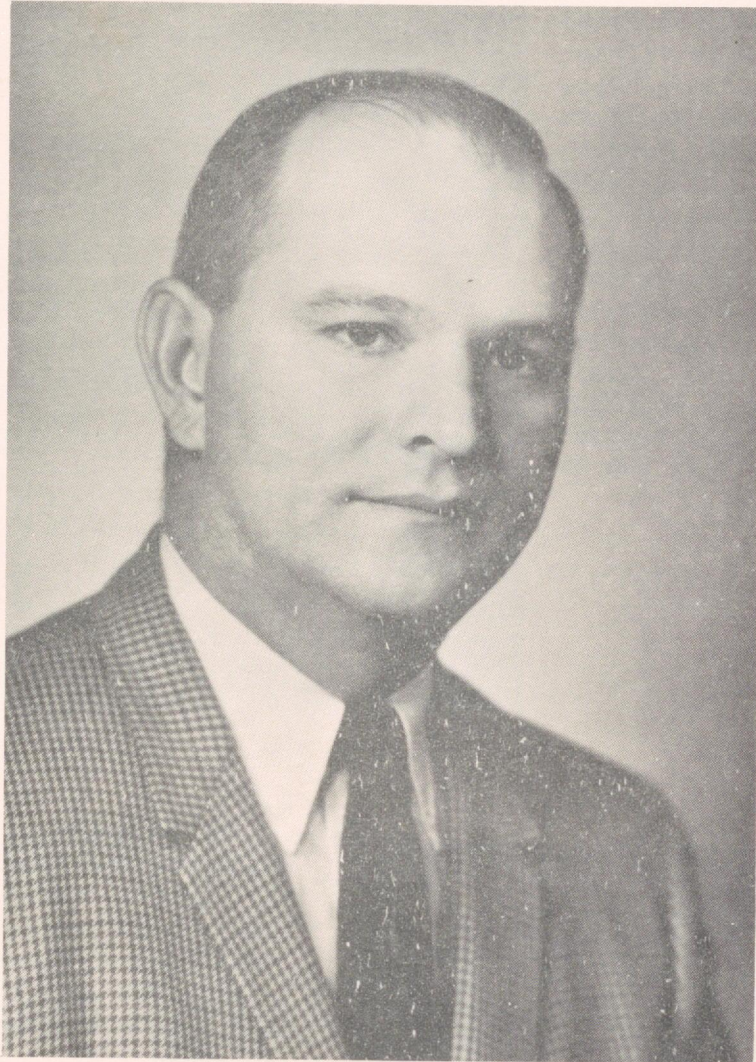
I commend our faculty for their continued interest in seeking additional ways to improve teaching. My only wish is that more of the "general public" and students could be made aware of these and similar efforts in other institutions. In my judgment, much of the present student unrest would subside if students, faculty and administrators could interact more freely and cooperate on efforts as demonstrated by this symposium.

I consider it both an honor and a privilege to have participated with the faculty in this endeavor.

With deepest appreciation, I remain,

Respectfully yours,

Carroll V. Hess, Dean  
College of Agriculture



Dr. Miles McKee, Department of Animal Science and Industry, received the Gamma Sigma Delta "Outstanding Teacher Award" in 1968. He is pictured on the cover in an informal discussion with a group of his students.

The Symposium Staff



Dale P. Scannell received three degrees from the University of Iowa. He spent eight years at the University of Kansas becoming Professor of Education and Associate Dean of the Graduate School. In 1967 he returned to the University of Iowa to become Professor of Education and Director, University Evaluation and Examination Services. Dr. Scannell has authored more than 20 publications including one book in the testing area.

Lawrence M. Aleamoni received his formal education at Westminster College, Salt Lake City, University of Utah and Michigan State University. He has spent several years studying item and test analysis programs. Currently, he is compiling results of a questionnaire on Selection, Placement, and Proficiency Procedures in American Colleges and Universities. Dr. Aleamoni is Assistant Director, Measurement and Research Division, Office of Instructional Resources, University of Illinois.



Program and Contents

<u>Wednesday, September 4</u>		<u>Page</u>
8:30-9:45 am	Registration	
10:00	Plenary Session I	
	R. Kenneth Burkhard, presiding	
	Student Panel to discuss "What is Wrong With the Design and Administration of Examinations at Kansas State University?"	1
	Panel Members	
	Ken Jorns, Moderator, President, Ag. Student Council	
	Michael Burns, Senior, Animal Science & Industry	
	Robert Dobson, 1968 Graduate, Dairy Production	
	Gale Fuller, Graduate Student, Horticulture	
	Brad Kerbs, Senior, Agricultural Economics	
11:00	General Discussion	15
12:00	Lunch	
	Arlin Ward, presiding	
	"Rock Springs Ranch: History and Facilities" - Merle Eyestone, Director, Rock Springs Ranch	
1:00	Plenary Session II	
	J. David Mitchell, presiding	
	"Why Test?" - Dr. Dale Scannell, Director, University Evaluation of Examination Services, University of Iowa	18
2:00	Refreshments	
2:30	"Kinds of Examinations" - Dr. Larry Aleamoni, Assistant Director, Measurement and Research Division, Office of Instructional Resources, University of Illinois	26
3:30	Discussion	32
4:30	Recreation: - Swimming, Canoeing, Archery, Horseback Riding, Volley Ball, etc.	
	Arrangements - Robert Mills	

6:00 Steak Fry

7:30-9:30 Evening Session

Charles Hall, presiding

Laboratory in Test Construction and Evaluation - 36  
Dr. Dale Scannell and Dr. Larry Aleamoni

Thursday, September 5

8:30 am Wilbur Ringler, presiding

"Small Group Discussion" - Each topic was discussed by a different group and results are reported by topics.

1. How might students be involved in formulating policies on testing and grading?
2. What should be the role of examinations?
3. What criteria and standards should be followed for determining course grades?
4. Should the College of Agriculture adopt an honor system of taking examinations?
5. Some universities adjust final course grades for differences in grading standards among faculty, departments, and colleges. Discuss the merit of such a system for KSU.
6. What factors should be included in a "composite evaluation" of a college graduate?
7. Some universities have an office of instructional resources to help faculty in evaluating test construction, test procedures, help design instructional aids, and provide an overall consulting service for improving teaching effectiveness. Discuss the merit of such an office at KSU.

10:30 Refreshments

11:00 Reports from Discussion Groups

45

12:00 Lunch

Ernest Mader, presiding

Introduction of Guests

<u>Thursday, September 5 (cont.)</u>		<u>Page</u>
1:30	Plenary Session III	
	Carroll V. Hess, presiding	
	"Psychological Implications" - Dr. Dale Scannell and Dr. Larry Aleamoni	56
2:30	Refreshments	
3:00	Discussion	57
3:30	Concluding Remarks, Carroll V. Hess, Dean, College of Agriculture, KSU	60
4:30	Recreation	
	Robert Mills, in charge	
6:00	All Ag Faculty - Family Picnic	
	David Mugler, in charge	
	Symposium Participants and Guests	62
	Suggested References	64

WHAT'S WRONG WITH THE DESIGN AND ADMINISTRATION OF EXAMS AT KSU?  
Student Panel Opening Statements

Brad Kerbs, Senior, Agricultural Economics

May I begin by saying I am most honored to be selected for this panel. Whether or not this gathering is a success will be difficult to determine, however, it will be less difficult if we keep a few things in mind.

The important thing here today is that I hope you instructors have a sincere and true interest in improving your course, improving the quality of education and instruction, and in the welfare of undergraduates. The symposium will be a success if these things can be done. Your future is somewhat dependent on the quality of instruction because of feedback from students on your teaching.

I hesitate to say what is wrong with examinations, because I feel in our college we have more things right with the examinations than wrong. This is not to say that all examinations, both good and bad, cannot be improved to a certain degree. I feel that talking about "what is wrong with exams," is a very good place to start to improve the quality of education, but it seems to be approaching the problem from the wrong end. It is somewhat like going to the freezer, opening the lid and finding the ice cream melted. There is a reason for the "bad ice cream." There is also a reason for "bad exams." In comparing both we find that "something is wrong here." In both cases everything looks fine, all equipment seems to be functioning properly, so you can then assume the electricity has been off, in both the freezer and the classroom. I believe this is a good parable. Most classrooms lack that electricity, that creative stimulation to the student. There will be no "melted exam" if there is some electricity in the classroom. I have yet to take a poor exam from a good, fair, enthused instructor.

Try something different. I think one of the big holdbacks of our present collegiate education, is the often worn out, abused lecture. It seems to me that with the new types of visual aids available that something a bit more exciting and interesting could be used. I realize there are some economic drawbacks and the layout of buildings in some instances hamper this, but something different should be done. Have more student feedback, not a few questions, let them enter into the course.

Last semester I had an instructor that didn't even stand to lecture to us. He simply came into class, sat down among the students, and the course was begun and ended that way. Something different was done. Students actually did most of the teaching. It was the most enjoyable course and most stimulating one I have had. I realize all courses couldn't be taught that way but something new and stimulating can be tried.



The student panel L. to R. Robert Dobson, Mike Burns, Ken Jorns, Gale Fuller and Brad Kerbs each presented ideas on examinations then answered questions from the faculty participants.

Student Panel Statements (cont.)

To get a bit specific on tests, I agree with the rest of the panel to concentrate on the important material. First things first and forget irrelevant and trivial details. I think students are as tired of trivia on tests as you faculty probably are, but nothing is ever done.

I feel a combination of all types of questions is the best and the fairest to all students. However, do not weight one section of the combination.

I may be completely out of line, but I feel that after the second semester of the sophomore year we should throw away the normal curve in grading. It seems to me that after three semesters the bottom of the original curve is cut off. The reason I say this is because on a junior level, if the class gets C's and above, most instructors feel they gave a poor test. Perhaps the test was concentrated upon the main things and the students absorbed them. Is it wrong that none of the students flunk? Most instructors felt the curve was too high so they threw in some insignificant trivia next time to lower the curve. All the students suffered by not knowing what was important.

There is a difference between learning and memorization. I have had a test where I was asked to list 13 different items, four or five pages of listing principles, objectives, etc. This is memorization. In high school they taught me that you cannot learn over three big items and if you add to them the three lose their importance.

Actually I think you are very fair in our college. You are more open minded than other instructors on the other parts of the campus. This is evident by your presence today. Basically you are doing a fine job but I hope that all of you keep in mind that improvement is the key word.

Gale Fuller, Graduate Student, Horticulture

What's wrong with exams at K-State? I find it difficult to discuss this negative point of view. I have found the great majority of examinations that I have taken here at Kansas State were the result of careful and conscientious preparation by the faculty member. In these exams, I found the questions clear and well organized so that incorrect answers on my part were a result of lack of knowledge of the subject, not misunderstanding. However, I have found a few points I would like you to consider.

One type of test question which I find often difficult to interpret and understand is the true-false question. I realize this

## Student Panel Statements (cont.)

type of test question is widely used and many students prefer this type. It takes less time and effort for the instructor to prepare and the student to take. For the student who is not willing to study for the exam, these true or false statements greatly increase his odds for passing the exam. On the other hand, the student who has put forth a great deal of effort to study thoroughly often finds the T or F statement a source of confusion. He may be familiar with exceptions or contradictions to a statement which would make it false to him when it was intended to be true. Tricky wording and ambiguous phrasing seem to creep into the questions no matter how hard the instructor tries to avoid them. I feel that T and F questions should not be used on major exams and should be used sparingly on short quizzes when time is an important factor in taking and grading the quizzes.

Multiple choice questions are probably one of the better types for an exam even though difficult to prepare. However, one modification of this type can be nearly impossible from my point of view. This is the selection of 1, 2, 3, 4, 5, any, all or none. Although it can be used effectively, it usually is the source of confusion and misunderstanding. It can become a compound T and F question. There are so many combinations of possible answers that scoring these questions presents another problem. If one error is recorded for each correct answer not marked and another error recorded for each incorrect answer marked, then the number of errors could exceed the total possible number of correct responses. If a large number of this type of question is included, it then becomes very time consuming. It is possible to select the best or most correct answer and the least correct but the statements that fall in between may be on the borderline whether to accept or reject.

My comments on these types of test questions might be combined into the one basic criticism that students express so often when taking exams. "The prof picks out all those insignificant little facts and gives them to us on the test." It is true that involvement in detailed facts can hinder the learning process. It is more important for us to learn concepts and ideas, the methods of approach to the understanding of a problem, and the tools with which to solve it rather than collecting vast series of facts which will soon be forgotten.

## (Examples of Application Rates of Growth Regulating Chemicals)

Recently I had a course where we were presented a lecture on a number of different growth regulators on plants, both stimulants and retardants. On the test over this subject we were asked to reproduce the application rates on a number of different plants in parts per millions, amounts in 10 gallons of water and 1000 gallons of water. These are all confusing and one cannot trust memory for this kind of detailed information.

Student Panel Statements (cont.)

Large amounts of factual material cannot be retained by most students. If these detailed facts and figures are called for on the exams, then the student is faced with the task of memorizing for many hours material which will be forgotten soon after the exam and hopefully not during the exam.

Our present testing methods often put so much emphasis on these details that the student who finds memorizing and the retention of these factual materials easiest is usually most apt to do well or score high on the exam. Students who have difficulty retaining materials of this nature will probably do poorly on the test even though their understanding of the subject and ability to apply it may be as extensive as the higher scoring student.

Most professors are conscious of this problem and are working to construct their lectures and examinations covering these lectures to emphasize the principles and concepts they are setting forth. I believe learning experiences building to understanding and knowledge are developed by thinking and interpreting more than reiterating.

Ken Jorns, President Ag. Student Council  
Junior, Ag. Mechanization

I want to comment on the information supplied to us concerning tests. In the material from one, a statement on the quality of tests says, "Does the question emphasize long run achievements rather than small incidental details?" This is important to me. I took a course, organic chemistry, in which they required us to give three different ways of preparing methane in the laboratory. I remembered that there were about four ways to make methane in the laboratory. I did okay on the exam, but now I could not tell you one of them. I feel that if I want to know how to do this I can go back to the book and find the method. However, I am not planning to make methane in my laboratory back home. I do believe it is important to know that there are several ways to make methane. None of these were commercial methods, but rather laboratory methods.

Another important point made in the instructions was, "Does the test require repeating details or applications of these?" I believe you should know some application to everything learned. I am not knocking knowledge for knowledge's sake but think you should be able to apply such things as history and philosophy to your life rather than just learning the material as facts. An example of this is like the little moron who was taking a course in scientific methods. He had this all figured out. He caught a frog one day and was going to run a test on it. He set the frog down and said, "jump frog"; the frog jumped. He noted in his book, a frog with four legs jumps ten feet. He then

## Student Panel Statements (cont.)

proceeded to cut off one of the frog's legs and said, "frog jump." This time he jumped eight feet, so he noted that a frog with three legs could jump eight feet. The same procedure was repeated by cutting another leg off. A frog with two legs jumps six feet. A frog with one leg jumps four feet. Finally he cut the fourth leg off and said, "jump," but the frog did not jump. So he noted in his book, a frog with no legs cannot hear. This is a good example of how important it is to be able to apply knowledge to what you are doing.

Generally, the best tests come from the best instructors and in the best classes. There is the human element involved. I note in some of these instructions for preparing tests, especially the one from Wisconsin, all of the details for preparing tests are present but they leave out the human elements. In what I consider a good course, the instructor does not have to worry about holding anyone's interest. Naturally, some students will sleep even at a basketball game or some of our football games. However, there is no discipline problem in these classes. Most students will sit up and listen. You must have a spark of enthusiasm in the course. If the instructor is not interested, how does he expect the students to be interested? He should be able to relate this enthusiasm to his students by applying his knowledge to interesting things.

Most students favor more tests than just a few, especially if they can drop one at the end of the semester. Even though the grades may not change the students are happier. This influences your reputation on the campus as an instructor before the students have ever seen you and in turn influences enrollment in the class. Your reputation gets around, if you are easy, tough, or as one who is interesting, who excites students and makes the subject matter real to them. I know in large classes it is tough to give a large number of tests and instructors have a problem with them.

Grades and tests are all a part of courses and I want to comment on grading. I feel that if a student comes to class regularly and does a half way job of keeping up, tests should be easy enough for him to earn a C in the course or at least expect to pass. In most classes where the normal curve with 7 percent A and 7 percent F is used, the F's can come from students who do not even go to class. I wanted to say that tests should be easy but in discussing this with my roommate last night we concluded that we both really do better when we have a real stinker of a test with an average of about 42. Even in this kind of situation the student who has been in class and halfway keeping up will still get his C or pass out of the course.

I do not think it is fair to all students to have only one kind of test all semester (all objective, multiple choice, true false or discussion). Some people do better on one kind of test than another. A change in format of tests will keep students on their toes. Even a quiz each week or taking up home work assignments helps stimulate interest and keeps students on their toes. You may only check names of those turning in home work without grading the assignment and still stimulate students' interest.

Student Panel Statements (cont.)

Most of the ideas and suggestions for preparing tests leave out the human element and application of subject matter. Both of these are important.

No matter how a professor decides to run a class (and, in the final picture it must be his decision) he should at least tell the students what will happen. Lay out a course outline and system of grading (fair or unfair); then stay with it.

Allow a student, at all times, to know how he stands with the rest of the class. Even if the stimulation of grade improvement is the only stimulation in the class, it's still better than no stimulation at all.

Robert Dobson, 1968 Graduate, Dairy Production

The first thing that came to mind when I was invited to speak was that the philosophy of the instructor, in regard to his course, is an integral part of his exams. I think the instructors can have two basic philosophies. They can either present the material, give the student a grade and send him on or they can have the philosophy that they are going to try to teach their students as much as they know and as much as the student can accumulate about a particular subject. I think that the instructors who have the attitude that they are going to try to teach the student will have a lot less trouble with the exams than the instructors that are just trying to get a grade on the student and get him out of there. I want to make this point very clear at the outset. I think if the instructor can keep uppermost in his mind that he is trying to teach students he usually will not have too much trouble.

The objective of tests may be looked upon as hurdles, instruments of teaching, or for placing a grade on a student. I think tests should measure the student's knowledge about the subject and also teach him at the same time. It takes an instructor that is both an artist and a good instructor to do this.

I think that tests should disregard material that really does not make too much difference. I don't feel that at any time material should be given on tests that is not covered by the instructor at least briefly. I realize that it may be necessary to just touch on subjects because you want to find out if the student is reading the book, but I feel that if it isn't important enough to mention in class, then it isn't important enough to have on a test. I personally like essay and recall tests because I have a fair command of the English language. However, I do not think they are really fair to students who have a lot of trouble with composition, etc., since these are subjective tests. To get a grade on a student you need to get a balance of different types of tests, both objective and subjective, before you really know your student's knowledge. If you want to make a test easy,

## Student Panel Statements (cont.)

have the same number of blanks to be filled in as you have answers. That should give every student about two points. I think it a unanimous agreement that true and false tests are the worst kind of tests. I think this is one of the few kinds of tests where you can actually be penalized for being knowledgeable in your field because there probably are not any completely true or false statements. If you look hard enough there is just a little inkling that maybe it was 99 percent true and one percent false. If you happen to be the poor student who knows it's one percent false you don't know what to do. You can forget it like you think you should, but you really can not have a clear conscience by putting it down either. As far as I am concerned the true and false test may be easy to make and easy to grade but they are sure not easy to take and they don't really tell you what the student knows.

Another thing which we have common agreement on is that frequency of tests should be such that we do not have a test every day that we go to class. On the other hand, I think that we should have enough tests for you to know that a student did not just have a bad day on one of these tests but for you to determine what the student knows about the subject. It takes me about one test to learn the instructor and I usually do rather poor on the first test. Then I go ahead and overcome this and get the grade I feel I really deserve in the course. If you have only two tests this is really bad for the student. If one does poorly on the first test, then it is really too late to do much about it on the second test. I think that the overall test quality should be such that the students have an opportunity to tell their instructor what they know about a particular subject. I think this is the most important thing. If he does know the subject and is not given an opportunity to tell the instructor of this, then I think the instructor has failed. I think that to a certain extent tests should allow students to demonstrate that they can think. By the same token, if you are taking a chemistry course, I do not think that you should be overly graded on English. I am the first to advocate good communications because I think this alone is responsible for a large part of the major problems that we are facing today in the United States and across the world. By the same token, a student should receive his English grade in English. I think that if a chemistry course is being taught to the utmost ability of the instructor, that should be what is required of the student at that particular time. Making sentences long and difficult to read and requiring lengthy explanations from the student puts more emphasis on English ability rather than on chemistry. I think this should be watched.

Grading cannot be overlooked from our standpoint either when you are talking about tests, the course, or college. I think that an instructor has to use the old term "common horse sense" when it comes to placing grades on students, both on individual tests and on the overall course. In the first place, I am very much against the

## Student Panel Statements (cont.)

A,B,C system. I think this is right back there with the horse drawn plow and all the other things in that period of time. I think, that for instance, when a student has received a high B on four out of five tests and an A on the other test that if the raw scores were added he would get an A in the course but if each time he was given a letter grade his grade point would probably be a B. I think this can be true throughout college and it can influence one's overall grade average. I think they could come up with a better way to grade students. I think that 94-100 is an A and a good guide to go by but by the same token if you are a good instructor and have a normal distribution of students in your class and do not have any A's, this is a mark against you. If students are trying, it is very demoralizing to the whole class and to students individually, particularly those students that are A students in other courses. I do not see anything wrong if you are a good enough instructor, if the material is presented well enough, and if the students are trying hard enough, to give all A's. By the same token if the students are not trying, I do not see anything wrong in giving all D's and F's. In a lot of classes 50 percent would be B's or C's whereas another instructor for the same course would give all the students F's. I do not think this is fair.

I do not like straight curves; I think they have to be tempered. The instructor has to use common sense. You can use the standard curve for a guideline but you can not use it 100 percent of the time.

These are a few final things I would like to leave with you. I do not think trivial matters should be included. I think students should have emphasized and drilled, if necessary, the main broad aspects of the course. I think they should get out of the course the main aspects -- the things that will stay with them. I had the opportunity to apply much of the knowledge I gained in dairy science. I found that you can see what courses you are weak in, because you do not know the broad aspects. Now I do not remember any of the specific details on this. The memorization of  $A + B = C$  are not the things you are going to remember. You will remember how to use  $A + B = C$ . I think this is going to be true in nearly all undergraduate education. Finally, I think that we need to give the students an opportunity to tell the instructor what the students know and at the same time incorporate into the course the test, a teaching device in itself.

Mike Burns, Senior, Animal Science & Industry

After listening to everyone else talk I really do not have any formal statement prepared that will add to what has already been said. I would like to give my opinion on certain types of tests that I feel more or less measures the ability of the student. Perhaps I could give

## Student Panel Statements (cont.)

a little criticism on the formation of these tests and how the tests measure up to the actual student ability.

First of all, I would like to say that I am in favor of and like to take essay tests. They give more personal contact between the instructor and the student. Essay tests can have short or extensive answers. From my point of view the essay test allows the student to put forth his ideas on the subject along with what he has learned in class. However, he may also disagree and have his own ideas, or different ways of presenting the material presented in class. His particular idea possibly is as good or better than the one presented to him. If he can show that he knows what he is talking about and he can convince the instructor that he knows the material presented in class, it allows more personal contact between the student and instructor. The instructor must design the test so that the student can supply answers he wants. The student can supply the answer the instructor is looking for if he really knows what he is talking about. The problem the instructor is faced with here is in making the essay type questions more specific. In other words, do not say discuss, discuss all the points about a particular subject, or something like that. Let the student answer it in his own way. Maybe he can not remember every little detail which is a natural problem. You either have to remember every single little detail or worry about what the instructor is going to give on the exam. You should be able to apply these details in general concepts or in problem solving rather than repeating them specifically. For the past several weeks I have had contact with people from other parts of the country. They were able to express and throw out to everybody all these specific ideas, very specifically, but I have not seen one of them who was involved in anything where he could use all these. He was able to rattle off all the equations in physics and chemistry. Students that I associate with may not know all the specifics but they know where they could go to get them and they know where they can use them. Now we should design a test that is going to measure the student's ability to use these and show how to do something worthwhile with them. If you just know the material it usually does not mean anything. I have had the same material in two courses yet one disagrees with the other. The literature I had first was more recent and I thought I knew this material pretty well until I took the other course where I had to write down something different than what was in the book. This was an outdated book we were using for class and I knew this was the case. The questions asked on the test pertained to this particular book. I know different and I told him, "I know what you want but this is not correct." Fortunately he gave me credit for knowing what I was talking about. It is a problem to determine if the student knows what he is talking about or if he is just trying to fill up space. The problem throughout testing is how to get the students to express themselves directly to the instructor and then for the instructor to put the proper evaluation on the students' knowledge. I do not believe tests are really the proper means of evaluation but they are the only thing we have to work with in school. We should

## Student Panel Statements (cont.)

try to put some kind of evaluation on the knowledge that a person obtains.

One point I want to make is that not every student is going to be particularly concerned with chemistry. They may need chemistry as part of their background to go into their field of study, however, they should not have to know every detail as the chemistry major does. Our fields are very much diversified and we can not be expected to know everything. As a student in chemistry, we have to know a lot of different areas of chemistry. A student in agriculture has to know, besides chemistry, something about entomology, plant science, and animals in general. It should be taken into consideration that the students in the class have diversified interests and therefore in an essay test they can express themselves on how this particular course or this particular field of study can apply to what they are doing. This I find makes it more interesting to the student and enhances learning. If I have to take the course and it does not apply at all to what I am doing then I am going to lose interest. Take chemistry I and chemistry II, I could not get with it at all, but when I got to organic chemistry it was more specific to my interests. I have not gotten to biochemistry yet but will next semester. I hope it is going to be the most interesting part of chemistry.

It has been said before, there are other people who can not really sit down and express themselves and of course this has to be taken into consideration as a difference between students. Each student is an individual and has his own way of expressing himself. Therefore, each instructor should attempt to understand the student, what he is trying to get across, and if he knows what he is talking about. I think our whole educational system has become so large and so extensive that I kind of look backward to the old Greek System where a student worked with his teacher.

#### Questions and Discussion

- Q. What is fair, or desirable for make-up exams?
- A. Jorns - Depends on the circumstances under which the test was missed. If you are sure he was sick, I would let him make it up.
- Burns - If the student has a legitimate excuse he should be allowed some form of make-up or given his basic average for a grade.
- Q. Support I am real honest and frank and say, "I simply overslept, Prof," what would you do with me? This is the second time. You have already given him one make-up before.

## Student Panel Statements (cont.)

A. Kerbs - This is difficult. I would prefer about five tests and in this way he could be given an average score from all exams. You people are too busy to use your time this way. Give him an average over four rather than five.

Q. What do you think of the pass-fail system?

A. Kerbs - I took a pass-fail course last semester and it was the hardest course I have ever had. This is contrary to reports that students only try for a C or D and these are easy courses. I would take it again because you do not worry about A or B grades lowering your average. I was worried about failing the course. I feel that I put more into the course than any of the others.

Q. Did you know how you stood during the course?

A. Kerbs - Yes, I knew I was on the brink of failing, so I really do not know how I would have reacted if I knew I had a B average. About 55 percent of the students were flunking the course going into the final, so the pass-fail really did not make the course easier. I believe there is a great potential for this kind of course. Only about 25 percent of the students in the class were on pass-fail. The instructor let us know we were not going to slip through the course easy because we were on pass-fail.

Q. What is your reaction to oral examinations?

A. Dobson - I took a class where this method was used for a small number of students and liked it very much. About 90 percent of our grades came from oral tests. These were administered in class as a part of discussion and were very effective. However, there were only about eight students in class and the method would not work in large classes. I did gain a lot from the class. No one got below a B in class.

Comment from Jorns - One thing that concerns me is keeping the students informed about progress in a course. Some professors take up exams, file them away, and the student never sees them again unless you requested to see the exam. Then you did not know where you stood with the rest of the class.

Q. How do students feel about having their final grades posted by their names at the end of the semester?

A. Jorns - This really does not bother me. However, I am usually not on the bottom of the scale and I do not know how you would feel if

## Student Panel Statements (cont.)

you were. It would probably depend on your grade in the course.

Comment from audience - Grades can be posted by assigned student numbers which takes care of this problem.

Q. What do you think about professors making old exams available to students?

A. Jorns - Good idea. Most organized houses have a file on each course and these can be an advantage.

Fuller - I feel that each semester a course is taught different enough that the tests will not be identical. New information should be added and each test should be different enough so that this would not be a problem. Students have old tests and there are files, so it is important to give different questions. There may be similarities which will help but not a great deal. From the instructors standpoint, he should use the old exam to improve his exams and to avoid giving the same exam over again.

Q. Should the Ag Council establish and maintain a test file in the Ag Reading Room for all agriculture students?

A. Discussion of pros and cons with no specific recommendation.

Q. Do students have sufficient protection or recourse when professors give all low grades, say 50 percent D's or lower?

A. Burns - This seems to be up to the conscience of the individual instructor in the College of Agriculture. I feel that there should be more personal or individual evaluation of students rather than saying so many will get A, B, C, etc. The important thing is that instructors should have some personal feeling for the students rather than just doing his job or pay check.

Dobson - I don't think we have very much protection at all from this kind of teacher. Neither do I think we have enough protection from poor instructors. Often good instructors are discouraged because they are not rewarded for doing a good job. I wish students were permitted to help in deciding who should and should not be rewarded financially for doing a good job. More reward should be given to good instructors on basis of this as well as on research accomplishments.

Kerbs - I do not believe students have enough protection against poor practices and that is why we have so much unrest across the country. Much of the problem with student demonstrations today is for publicity, but there are sincere students who have

## Student Panel Statements (cont.)

legitimate problems and are sincere in seeking solutions. I believe a faculty evaluation conducted by the Dean's Office which would be unpublished could help eliminate some of the problems. I do not agree with a student published evaluation or rating system. The instructor could profit from an evaluation by students. Then the superior instructors should be awarded for their efforts.

- Q. Are grades, evaluations, etc., relics of the past as claimed by the NSA\* at the convention here? Are most kids capable of complete evaluation?
- A. Burns - We can be highly critical of this group, especially their steering committee, because they do have many undesirable motives. However, this does not mean we cannot have some means of self-evaluation. I think we are more capable of evaluating ourselves than instructors are. However, I do not believe that the NSA group in any way represents the general thinking of students. Many of them are undesirable people, yet others are well meaning students. I do believe student evaluation should be supervised by instructors who have had experience and training in these areas.
- Q. Unless a person is to be self-employed, isn't self-evaluation rather illogical?
- A. Burns - No, regardless of what you do you are going to have judgements placed on you, but everyone has a different kind of judgement. It is up to me to decide where I want to go and decide what I am capable of doing.

Comment from audience - You are going to be judged throughout life in terms of salary incomes, work efficiency and many other things and the student should become adjusted to this and make the best of it.

Jorns - I would agree that grades and other student evaluation procedures are good in helping students know where their weaknesses and strengths are and should be used in that way. We have to make adjustments and should do this gladly if needed rather than oppose the evaluations. It is to our benefit.

- Q. Will a large percent of students be objective enough to evaluate themselves?
- A. Burns - We have to look at our background. We have had grading systems where we compare ourselves with other people. We have not been evaluating ourselves, our own personal strengths and weaknesses and would probably not be prepared to do this right away.

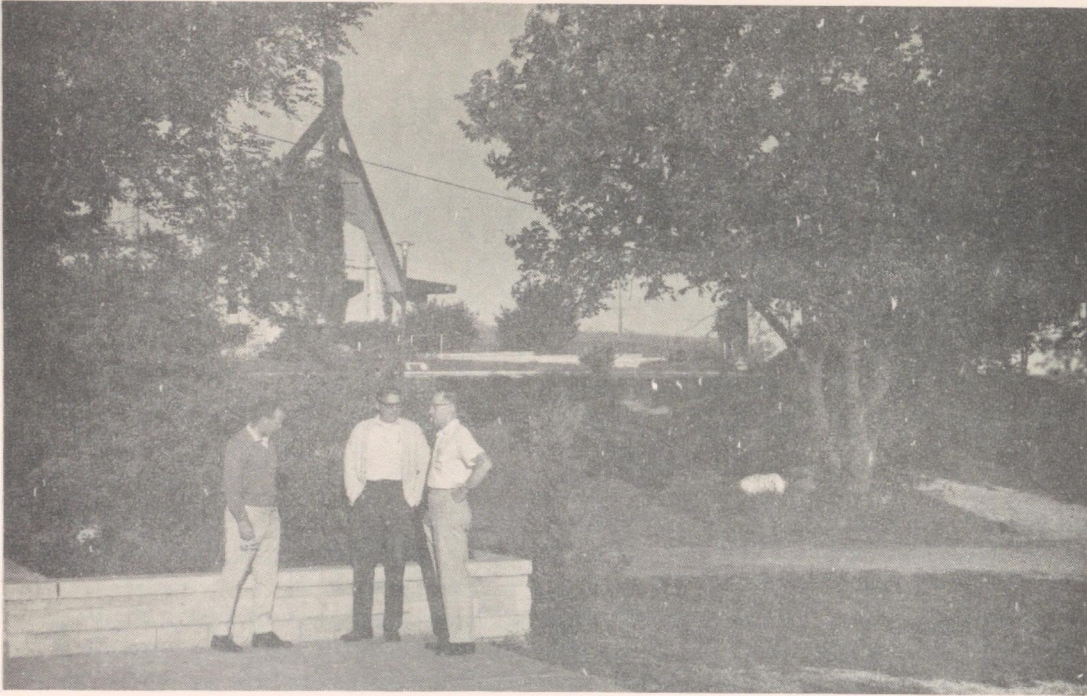
\* National Student Association

## Student Panel Statements (cont.)

- Q. Would student evaluation be more accurate with essay, or objective exams?
- A. Jorns - This is again up to the instructor. He can grade a discussion question any way he chooses and he may grade it differently on two different days. Accuracy for grades is probably less on discussion exams, especially when applied to a point system.

Discussion

- Q. What part should students play in determining the number and kind of exams?
- A. Scannell - It is my own personal conviction that they do not have a right to say anything. If I am not professionally competent to know how many tests I have to give, or the kind of tests to measure the students level of learning, I should not be teaching the course. Students may think they know the best kind of test to measure their level of achievement, but they may not be right. Students really do not know if I am better at designing multiple choice questions than grading discussion questions. If I am a poor grader of discussion questions they are much better off with the multiple choice and only the teacher knows which he can do best.
- Q. Is it true that the majority of students thrive on success or knowing the level of accomplishment?
- A. Scannell - Yes, this is a motivating force. What do grades represent though? I think grades should represent the level of student achievement in the course. Personally, I do not care whether the student comes to class or not, or whether he started all the way down at the bottom and came up or not. All I want to know is what he has achieved in the course without regard to where he was when he came in. I am only attempting to measure how much he has achieved in my course. My grade is defined in terms of performance over the subject matter covered.
- Q. Do you base your score only on the final exam?
- A. Scannell - No, I base the grade on the best composite score I can possibly arrive at in the course. I may give several tests and all will have the same relative weight in determining the grade unless I deliberately weigh one more than the other, which is uncommon. If I find an exam is unreliable in terms of standard



Informal small group discussions between sessions played an important part in the exchange of ideas and information. Shown in the top photo L. to R. are Warren Prawl, Gerald Wilson and Howard Mitchell. In the bottom photo L. to R. are Gale Fuller, Larry Aleamoni, David Mitchell and Don Hoyt.

## Student Panel Statements (cont.)

deviation, I may give more weight to the more reliable exams. I may also use other kinds of information in arriving at the final grade.

Q. Does the instructor have a responsibility to prepare students for the exam to enable them to perform their best?

A. Scannell - Who is going to determine whether or not the student performed at his best? He may say that he blacked out - how do you know that? I know some students are very persuasive and it is hard to tell when there is actually a real case of not being able to perform at the best level. However, I do know this sometimes happens and should be avoided if possible. We will discuss test anxiety tomorrow and give results of some research on the subject. I still raise the question, how do you know when students are under-performing on a test?

## WHY TEST?

Dr. Dale P. Scannell  
University of Iowa

Without doubt the most common reason why college instructors give tests is to collect information on which to base the marks required by the central administration of their schools. This reason for testing is so exclusive that many people find it difficult to think of testing as anything separate from the process of judging student achievement. In fact, for many people the terms measurement and evaluation are assumed to be synonymous with testing. While it is true that testing is one way to measure, and that measurement is a process necessary for evaluation, these three processes are truly separate and distinct from each other. The title of this session, "why test," implies that there is not just the one reason, but rather that there are a variety of contributions that classroom testing can make to the teaching-learning process. The reasons for testing necessarily derive from the nature of the substance to be measured and from the type of evaluation to be made with the data obtained. It seems to me that there is an inverse relationship between educational level and exploitation of test data (the number of different reasons for testing). Elementary school teachers derive many benefits from classroom test data, while instructors of graduate and professional courses give a minimum of tests--just enough to derive grades.

There are certainly many exceptions to the general situation noted above, and undoubtedly many professors derive some of the many potential values from test data, in informal ways. Thus, my purpose in this session is not to confront you with something that is entirely new to you, but rather to survey with you some of the more important potential values of test data. In addition, I will make a few comments about the unique characteristics of tests designed for the purposes I will mention. I have decided arbitrarily to discuss the ways in which tests can contribute to the teaching-learning process under six separate headings. These are: 1. To assign grades 2. To diagnose student learning difficulties 3. To determine whether minimum mastery has been achieved 4. To motivate students 5. To supplement and enhance the learning process 6. To measure teaching effectiveness.

Each of these six topics could provide the basis for a talk, and therefore I must limit my comments to the high points and hope that during the discussion we can explore further those that are most relevant or most troublesome in the College of Agriculture. These six topics obviously are not independent of one another, but this overlap reflects the actual complexity of the teaching-learning process and our state of knowledge about cause and effect relationships within an educational setting. Seldom, if ever does a teacher do something that has but a single effect on the students. Usually an action designed for one purpose will influence other aspects of classroom activities and student achievements. Classroom testing is obviously



The complete group during a presentation by one of the guest speakers.

in this category, and although the topics overlap, it will serve our purpose best to consider them separately.

### I. TO ASSIGN GRADES

In almost all educational systems instructors are required to assign marks to indicate the level of accomplishment reached by their students. Since this use of test scores is required of instructors, and other uses are not, it is easy to understand why this is the most common reason for giving tests. Regardless of the type marking system used by an institution, or the philosophy associated with the implementation of that system, there are two characteristics that a test must possess if it is to be useful in this process. The first is that the test must rank students accurately, and the second is that success on the test must be directly related to the amount of achievement made by the students in this course. As obvious as these two points are, it is surprising how many courses there are in which the examinations do not meet these twin criteria. In some courses performance on a test is more closely related to student ability to write clear prose, to the student's general level of intelligence, or to the richness of his background in areas related to the course. (Sometimes I think that many of us in college teaching are assigning grades more of the basis of what students bring with them to the course than on what they derive from our own teaching.) In addition, very few classroom tests are designed to be maximally effective. It is fairly easy to demonstrate that to be most effective for the process of ranking a group of people, each item on a test should be answered correctly by approximately 50 percent of the group for which the item is appropriate. And yet most of the tests used in universities are much easier than this. Many instructors still believe that there is something magical about the number, 100, and that anything less than 60 percent on a test must be regarded as unsatisfactory performance.

I think that testing practices would be improved immensely if instructors merely realized that every test represents only a sample of the many kinds of accomplishments that could be included on the tests. The meaning of 100 percent on a test depends on what the test in total represents. We cannot say that students who answered 100 percent of the questions on the test have made 100 percent of the possible accomplishments in that course. That is, we cannot assume that there has been 100 percent mastery because a student made no errors on a given test. Most instructors, however, still hold to the optimistic belief that they can design a test on which 60 percent of the possible score represents a minimum satisfactory level of achievement. As a result, many of the items on classroom tests are less effective than they could be. If everyone taking a test gets an item correct or if everyone misses the item, then that item contributes nothing to our knowledge of differences in level of student achievement. The item makes no contribution to helping us rank students accurately.

Some instructors claim that testing and grading are the most

unpleasant aspects of teaching. To support this point of view, they indicate that teachers cannot effectively assume both the roles of tutor and judge. To guide students as well as possible during the instructional process and then to sit in judgment seems to these people to be incompatible responsibilities.

We cannot deny the attractiveness of this claim, at least on the surface. In some institutions the problem is partially avoided by having all major examinations developed by external committees or by an individual other than the instructor who has taught the course. In some of the United States and in some foreign countries the problem has been reduced partially (and replaced by other problems) by the use of external examinations prepared by national boards. Obviously, this approach would not be immediately applicable in our system of education.

In spite of the superficial attractiveness of this argument, it should be noted that a teacher sits in judgment at all times in the instructional process. Each time a teacher makes an assignment or recommends an outside reading or modifies the instruction for an individual student, this teacher has been sitting in judgment. In fact, we could develop a good case to support the position that one cannot be an effective teacher without being, in many ways, a good judge.

It seems to me that instructors are uncomfortable in the position of judge primarily because they know that their judgments are based on insufficient and questionably relevant evidence. Thus, this problem is rationalized by blaming the system and the requirement of marking, when in fact the happiest solution might well be to find ways to obtain better evidence of student achievement. The problems may never be overcome completely, but they can be reduced if instructors know how to develop better tests, understand more completely the nature of educational measurement, and know the limitations of the data they must interpret. I assume that your recognition of these facts contributed to your decision to conduct this year's symposium on testing, and I applaud you for this.

## II. TO DIAGNOSE LEARNING DIFFICULTIES

There is a school of thought today in psychology that anyone can learn anything to any level, if given sufficient time and resources. Aptitude for learning is thus defined in terms of the time required to complete a learning task. If this position is correct, the effective instructor should be able to guide all students to the desired level of achievement if he has sufficient time and adequate information about the learning capabilities of the students in a class. Of course, the instructor's job would be simplified if the institution had an accurate way of selecting for admission to the university only those students who had reached the minimum level of mastery on prerequisite skills.

Even though we may be reluctant to accept this school of thought completely, and I am, there is a message here for all instructors. We

can all do a better job of guiding individual student effort than we are now doing. If better information were available about the learning problems that the students in our classes have, we could individualize instruction better and help to increase the achievement of all of our students. To do this, of course, we need to diagnose the learning difficulties of our students and this in turn requires data from a certain kind of test.

Tests that are used primarily for grading purposes are designed to discriminate as much as possible among students who have achieved differentially. In addition, these tests must be structured carefully to reflect all aspects of achievement that have some permanent value. Primary emphasis is on the students' abilities to use learned information in solving new problems that have not been experienced in the learning sequence.

In contrast to the achievement (discriminatory) test, diagnostic instruments concentrate only on the steps of the learning process which are known to present difficulties to student progress. Balance among content and skills, and ability to discriminate, are of no interest in diagnostic testing.

In most disciplines diagnostic testing is used to support or verify the instructor's observations. If a student seems to be having difficulty with a certain process or skill, the instructor administers a test over just that narrow area. Of course, the key to diagnostic testing is the instructor's ability to analyze his teaching into its component parts.

### III. TO DETERMINE WHETHER MINIMUM MASTERY HAS BEEN ACCOMPLISHED

Very few college instructors recognize some of the inherent characteristics of educational measurement. For example, classroom tests are concerned with student ability to deal with concepts, but concepts exist only because we have conceived of and defined them. Thus, operational definitions are essential for the concepts we are attempting to measure. Frequently tests provide the only or the best operational definition for a concept that is available. For example, achievement in general biochemistry is a concept that has no inherent physical referent. What we mean by "achievement in general biochemistry" becomes apparent only after that term has been operationally defined. Does it mean ability to associate the names all chemical elements with the appropriate chemical symbols? Does it mean more than this, for example, ability to balance the redox equations for the reactions that occur within an animal body? Does it include the ability to predict what biological change will occur in a certain solution with a given concentration? What biological changes, what solutions?

Most of the classroom measurement problems that you are associated with involve a broad spectrum of content, student skills, and possible applications. For most of your purposes you have never found it necessary to analyze each aspect of the course to the extent required for the

development of mastery tests. However, there are some situations which require evidence of student mastery of basic concepts. Two examples are apparent. In some courses the material proceeds in a necessary sequence from simple to complex. The understanding of material in the fourth week of the semester requires a minimum understanding of the material covered in the first three weeks. Similarly, the courses which comprise some disciplines must be organized within a given sequence because each subsequent course builds on the preceding material.

In sequential areas tests of minimum mastery are required. These tests must, in effect, define the skills that are essential if students are to make reasonable progress in the next step of the sequence. The tests may have an adequate ceiling so that differences in achievement above the minimum essentials can be detected for purposes of assigning grades, or the two types of tests can be administered as separate units. As in so many other areas of measurement, the crucial issue here is the ability of the instructor to identify those skills that are truly essential, and to devise the test questions that will reflect accomplishment of the minimum skills.

#### IV. TO MOTIVATE STUDENT EFFORT

The motivational uses of tests are frowned upon by many people. They claim that it is more important for a student to be motivated by a desire to learn than by the threat of an examination. While I think we all would agree that it would be desirable for students to want to learn for the sake of learning, we would also recognize that in many courses this is an unrealistic expectation.

Many instructors fail to realize the motivational advantages of tests; some instructors abuse tests by going to the other extreme. Somewhere between there is a level that teachers should seek to find. If students, in studying for a test, are performing desirable learning activities and developing toward the goals of instruction, then the motivational aspects of testing are more desirable than undesirable. Of course, undesirable practices can be employed. If tests are not accurate representations of what is important, if test items are unrealistically difficult, if tests are used as punishment techniques, if tests are used in a threatening way, or if teachers put primary emphasis on the scores and not on the learning the scores reflect, then of course we would all agree that tests should not be used for motivational purposes.

#### V. TO ENHANCE THE LEARNING PROCESS

There are several ways that tests can be used to supplement other teaching techniques. As I have implied earlier, tests provide an operational definition for what an instructor holds as the important contribution of his course to the curriculum of which it is a part. Perceptive students realize that tests provide the true definition of what it is important for them to accomplish, regardless of the instructor's propaganda. You may tell students that memorization is not

sufficient, that rote learning may be part of but not the terminal aspect of achievement, that ability to apply generalizations in solving novel problems is the end you seek, but if your tests emphasize facts, memorization, and trivia, students will study to learn facts and if necessary trivia.

Tests serve a guidance function with greater impact than assignments, discussions, or statements of goals. Students study for what they expect to be held responsible for on tests. Thus, a wise instructor will draw on tests to help guide students into desirable learning activities.

There is another way in which tests can interact with the learning process. Knowledge of success and reinforcement of learning are important aspects of effective learning and retention. Psychological research has indicated that students will make more rapid and more permanent advancement through a learning problem when they have some indication of their progress up to a given date. Tests, in the form of quizzes and written exercises, provide feedback to the student about the progress that he is making. Desirable advancement is thus reinforced, and students are given some indication of their weak areas so that they can use their study time in the most appropriate way.

#### VI. TO MEASURE TEACHING EFFECTIVENESS

The measure of a teacher is the achievement of his students, and diligent instructors are constantly alert for evidence of their successes and failures. Effective tests can provide feedback to the instructor concerning the areas in which his teaching has been less than completely satisfactory. Tests can suggest which activities are making a positive contribution to student development, and which should be replaced or revised. As the technology and knowledge of an area changes, an instructor must change his materials and perhaps his overall strategy. Tests cannot provide definitive evidence but they certainly can provide suggestions for the person who desires to teach a maximally effective course.

For a test to be useful in the process of evaluating teaching effectiveness, it must be relevant and specific to the desired achievements in the course. Great care must be exercised by an instructor so that he does not draw comfort from tests that reflect the students' native endowment, prior education, verbal facility, or some combination of factors not related to the effectiveness of his own efforts. Properly handled, however, tests can be an effective tool for curriculum reform and instructional updating.

#### Summary

Six reasons for administering classroom tests have been reviewed,

and the unique features required of tests that serve those specific purposes have been discussed briefly. Better testing is part of better teaching--the improvement of testing must necessarily improve the overall quality of a course.

## TYPES OF EXAMINATIONS

Dr. Lawrence M. Aleamoni  
University of Illinois

If one were to ask: "How many different types of examinations are there?" the response might be a bit surprising for there are more varieties of examinations than most people realize or care to know about.

In the Sixth Mental Measurements Yearbook, edited by Oscar K. Buros (1965) I found 15 general categories of tests ranging from Achievement Test Batteries to Vocational Tests. In addition, the total number of more specifically defined categories of tests, under the 15 general, was 86 (e.g. under Vocational we find Clerical and Manual Dexterity tests). Finally, the total number of unique tests reviewed over all these categories was 3,469 (for years 1938-1964).

Examinations of a different kind are constructed, administered and evaluated by you and me every day in many common places. For example how many times have you looked at a pretty girl and asked yourself "I wonder if she...?" I will let you fill in the rest of the stem and speculate on the number of alternatives. Or how often have you applied a test of your own making before buying a car, house, etc.? Undoubtedly you would be able to add many more examples to this list of examinations.

To begin with, let us define a psychological test as being essentially an objective and standardized measure of a sample of behavior that can provide inferences about the people being tested. We will now look at several pairs of contrasted types of tests such as: paper and pencil versus performance, individual versus group, achievement versus aptitude, speed versus power, tests with "face validity" versus tests with real validity, long versus short, and free-response versus limited-response (sometimes referred to as essay versus objective).

These tests can actually be further classified on the basis of form and content where achievement and aptitude tests would qualify under the latter classification.

The familiar paper and pencil type of test provides each person with a test booklet containing all the test items. Each person answers the item by either writing in the test booklet or on a separate answer sheet (or sometimes both). In some paper-and-pencil tests, the item information is presented on phonograph records, tape recordings, video-tapes, etc. Examples would include tests of musical aptitude and tests of foreign language aptitude or achievement.

In performance tests, an individual may be required to manipulate objects, pictures, blocks, or cash registers or perform more complex

activities in a typical everyday life situation (such as having a pretty girl walk past a line of men). Performance tests have usually been restricted to individual testing due to the difficulty and expense of providing duplicate sets of materials. Another reason is that in most performance tests each person could easily see what the others are doing.

Group tests differ from individual tests in that the former permit instructions to be given to many people at the same time, each of them writing his own answers. Group tests can be given to a single individual if desired. Many individual tests require careful oral questioning or observation of behavior and, in effect, usually become interviews in which the tester asks questions and writes down the testee's answers.

Achievement tests are designed to measure an individual's proficiency in a particular area and can be used to predict performance in that area. Driving tests, typing tests and tests in specific subject matter fields, such as biology, chemistry, agriculture, and the like are also considered to be primarily achievement tests.

Aptitude tests, on the other hand, are designed to predict an inexperienced individual's potentialities, his ability to learn the job or his general overall learning ability. The former could be illustrated by certain clerical aptitude tests and the latter by the "general intelligence" tests.

A differentiation will now be made between speed and power tests. A pure speed test is one in which each person's score depends entirely on how many questions he is able to answer in the time allowed. Such a test is usually constructed from items of uniformly low difficulty, all of which are well within the ability of the people being tested. The time limit is then made so short that no one can finish all the items. A pure power test, on the other hand, has a time limit long enough to allow each person to attempt all items. The distribution of the difficulty of the items is highly positively skewed, and the test includes some items too difficult for anyone to solve, so that nobody can attain a perfect score. It should be noted, therefore, that both speed and power tests are designed to prevent the achievement of perfect scores as perfect scores are indeterminate and do not discriminate.

Most tests actually depend upon both power and speed in varying proportions.

The item "face validity" is used to refer to a test whose content "looks" as if it measures what is meant to be measured, whether or not it actually does. Test users should be very cautious in selecting tests based on inferred rather than statistical relationships to measure or predict some desired criterion. The reason for this note of caution is that a test can appear highly unlikely to measure what the tester may be interested in until a correlation is calculated between the test score and the criterion and shows that a high relationship actually exists. However, good face validity can help to keep

the motivation of the test taker high.

Test length is integrally tied to sampling since the inclusion of all items pertinent to a given area of knowledge would make a test inefficiently long. Therefore, we must use enough items in a test to show the degree to which each person possesses that knowledge in comparison with others. Since we very seldom test one field of knowledge in an examination we must obtain several adequate samples of items representing these different fields. The cost of constructing, administering, and scoring a test along with the item sampling determine test length. In practice, a test should include as many items as are necessary for adequate assessment of the fields of knowledge but few enough so that the test is practical to administer and within the physical capacities of the test taker.

In free-response tests individuals are asked questions in which they have relatively unrestricted freedom in expressing their best answers. On the other hand, in limited-response test items, the individuals may be limited to choices among or manipulations of several suggested answers or elements.

The free-response type of test includes two common forms, namely, the essay and the oral. The limited-response type of test includes matching, arrangement (or rearrangement), true-false, multiple choice, and completion items.

If one wants to (a) expend little time in preparing an examination, (b) provide the opportunity for the individual to select, organize, and integrate learned material, (c) evaluate effective writing, and (d) provide the possibility for bluffing or writing around the topic, then an essay test should be used. However, these "advantages" may be offset by weaknesses in the essay test's reliability of scoring, adequacy of sampling individual achievement, labor required in scoring, and freedom from distortion of grading by skill in expression and quality of handwriting.

The oral examination is a good way to discover what thought processes an individual uses in solving a problem.

In the preschool, kindergarten, and lower primary grades the oral examination may be the only way to evaluate achievement because the pupils have not yet learned to read. In higher grades and at the college level they are often desirable when the instructor wishes to evaluate a student's ability to discuss broad types of problems in which integration of knowledge in several areas is important.

On the negative side, oral examinations have the same limitations as essay examinations--poor sampling of content, great consumption of time, and low reliability.

In general, large-scale testing of students is confined almost entirely to tests of the limited-response type. This form has the advantage that the scoring can readily be made much more objective and reliable than the free-response, and provide the same relative score for a student on equivalent forms of a test.

Before discussing the types of limited-response items, it would be profitable to consider some of the principles that apply to all of them. Some of these principles are merely common sense notions, but their discussion is justified by the frequency with which teacher-made tests violate them. Others reflect the research and experience of test experts and would not be apparent to novices.

1. Avoid obvious, trivial, meaningless, and ambiguous items.
2. Observe the rules of rhetoric, grammar, and punctuation.
3. Avoid items that have no answer upon which all experts will agree.
4. Avoid "trick" or "catch" items, such as, items phrased so that the correct answer depends on obscure key words to which even good students are unlikely to give sufficient attention.
5. Avoid items which contain "irrelevant cues." These are items phrased so that the correct answer may be determined merely by using intelligence without real knowledge of the achievement at which the item is aimed. Illustrative of an irrelevant cue is the following:

Man is an (1) plant, (2) reptile, (3) animal, (4) bird.

The irrelevant cue here is the article "an," which indicates that the correct answer must begin with a vowel.

6. Avoid items that furnish the answers to other items, because this will render one of the items useless for evaluation purposes.

We will now describe the various types of items characteristic of the limited-response test.

Matching items employ the same set of responses for a cluster of several similar stimulus words. There are a variety of ways to match responses to stimulus words. Some examples are events and dates, events and places, events and results, inventions and inventors, processes and products, and causes and effects. Similarly, the stimuli and responses may be presented in many varying forms--diagrams, maps, pictures, chronologically or logically arranged stimuli with numbered gaps between them.

To prevent guessing, there are usually more items in one column than in the other. A short example of a matching item is:

Directions: After each animal write the letter corresponding to its gestation and incubation period.

<u>Gestation and Incubation Period (in days)</u>	<u>Animals</u>
A. 330-340	1. Swine
B. 280-283	2. Chickens
C. 143-150	3. Mares
D. 112-114	4. Sheep
E. 67-70	5. Cattle
F. 20-21	

The arrangement items require the student to put into some specified order a series of randomly presented material. Any kind of specified order may be called for, such as chronology, difficulty, importance, length, weight, or logic.

An example of a chronological order, arrangement item is:

Directions: Given below are groups of three events or men whose letters you are to write in chronological order to the right of the items.

1. (A) Wilson, (B) Lincoln, (C) Washington were Presidents of the United States . . . . . \_\_\_\_\_
2. (A) Catholics, (B) Quakers, (C) Puritans were religious groups who settled in America . . . . . \_\_\_\_\_

A major difficulty with this type of item is in scoring each possible arrangement according to the extent of its deviation from the correct arrangement.

The true-false item is the simplest to prepare and is also quite widely adaptable. It tends to be less discriminating, item for item, than the multiple choice type, and somewhat more subject to ambiguity and misinterpretation. The high proportion of items that can be answered correctly by chance, and the random error that chance responses introduce in the scores, is a major limiting factor. However, if a person is careful in constructing a true-false item and avoids most of the common pitfalls, such as lifting textbook sentences and changing one or two words, he can produce a very discriminating examination.

Multiple choice test items are currently the most highly regarded and widely used form of objective test item. They are adaptable to the measurement of most important educational outcomes--knowledge, understanding, and judgment; ability to solve problems, to recommend appropriate action, to make predictions. Almost any understanding or ability that can be tested by means of any other item form--completion, matching, arrangement, true-false, oral or essay--can also be tested by means of multiple choice test items.

Students often find multiple choice questions less ambiguous than completion or true-false items. Instructors find it easier to defend the correct answers to them.

Although chance errors resulting from guessing are still evident here they are much smaller due to the larger number of possible alternatives. The good multiple choice test item is difficult to construct. Devising plausible incorrect alternatives places a heavy burden on the ingenuity and psychological insight of the instructor. Constructing a simple multiple choice item with four alternatives may require as much work as constructing four true-false or simple question items.

Completion items are those in which a statement is presented with a critical key word or phrase missing. The student is then asked to recall the missing words and insert them in the appropriate blanks.

Good completion items usually cannot be made merely by leaving out one or more words from a sentence or passage from the textbook, because such words often make sense only as a result of their context. Without the context the completion item becomes ambiguous. Instructors should write completion items in their own words, make them self contained, and make sure that they limit the possible correct responses to exactly the kind of achievement intended.

The organization, description, definitions, and selection of types of examinations presented in this paper are by no means exhaustive. If you are interested in further enlightenment then I recommend looking at the references listed at the bottom of this page.

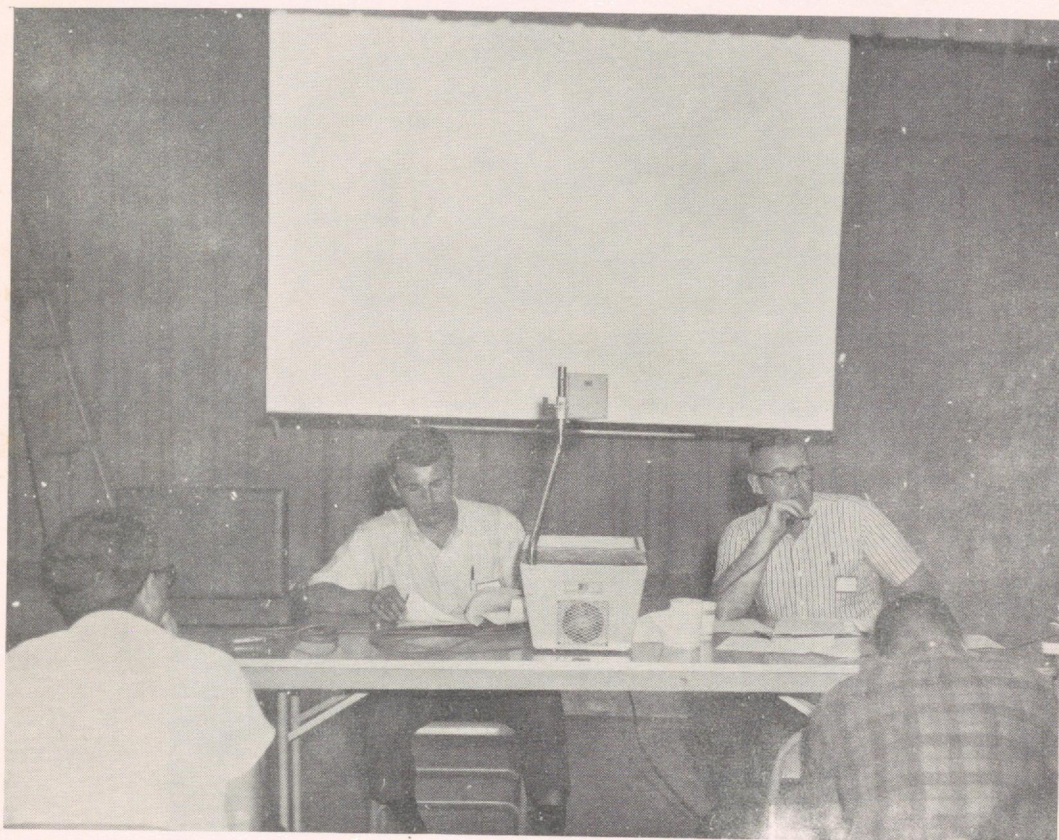
Many people liken a speech to a pair of deer horns: a point here and a point there and a lot of bull in between. I hope that you were able to concentrate on the points.

#### References

- Adkins, Dorothy C. Construction and Analysis of Achievement Tests. Washington, D.C.: U.S. Government Printing Office, 1947.
- Anastasi, Anne. Psychological Testing. (2nd ed.) New York: Macmillan, 1961.
- Cronbach, L. J. Essentials of Psychological Testing. (2nd ed.) New York: Harper & Row, 1960.
- Ebel, R. L. Measuring Educational Achievement. New Jersey: Prentice-Hall, Inc., 1965.
- Lyman, H. B. Test Scores and What They Mean. New Jersey: Prentice-Hall, Inc., 1963.
- Remmers, H. H.; Gage, N. L. & Rummel, J. F. A Practical Introduction to Measurement and Evaluation. (2nd ed.) New York: Harper & Row, 1965.
- Tyler, Leona E. Tests and Measurements. New Jersey: Prentice-Hall, 1963.

Questions and Discussion

- Q. What do you think of open book exams?
- A. They are among the highest level exams if properly designed. They should measure the use of information when it is required rather than recall of specific information.
- Q. Knowing that some authorities on testing and evaluation use true-false questions altogether while others use multiple choice, is it not possible to mix these types of questions on the same test?
- A. There is no apparent reason why you should not mix types of questions. The type of question should depend on the type of test item which will most effectively provide the information you want. Completion, true-false and matching items do a fairly good job of measuring a student's verbal level learning such as structural group of organic compounds, etc.; for testing knowledge of names, etc., matching is the most efficient type. For measuring students' ability to apply concepts or data you can't use matching. Some people think that a complete free response type question such as discussion or essay is the only way to test this kind of knowledge. I feel that well designed multiple choice items are very effective for this purpose. True-false items are more difficult to use for this purpose, but it is possible to design them in such a way that they will give reliable information. There are very few cases where we have a true dicotomy and to design a true-false question where this is not the case is very difficult. There are many difficulties in designing good true-false items. They have many more inherent limitations than the multiple choice type.
- Q. In grading true-false questions do you grade off double, or more for incorrect answer than are allowed credit for the correct answer?
- A. No, I do not count off for guessing.
- Q. Of what value are pre-tests in a teaching program?
- A. For instructional purposes they are fine but have no place in grading. They can be used to determine the level of instruction at which a course should be started, etc., but the pre-test score should not enter into the grade. You cannot or should not attempt to measure growth in a course for grading purposes. Grades should reflect achievement over materials covered in course.
- Q. Do all instructors have an obligation to grade grammar on answers on test questions?



Drs. Aleamoni (left) and Scannell (right) answered questions following the formal presentation of papers.

- A. Yes, I would say so, but not for credit on the exam unless it is a course in grammar. Spelling, punctuation, etc., should be corrected, but it should not be included in the student's grade.
- Q. The point has been made that it requires considerable time to make up good multiple choice questions. Is it safe to re-use these questions, or should new ones be made up each time?
- A. Aleamoni - The main stem, or question can be re-worded or slightly changed to such an extent that it is possible to safely use the same questions again and not easily be recognized as being the same.

Scannell - It is possible to maintain strict item or test security to prevent questions from getting in student files and to re-use them. However, to do this one must have a fairly large pool of items accumulated over a period of time.

The group response was such that they did not believe this would be possible.



Following the first days crowded program schedule the Ag and Engineering Faculty (Engineers were conducting a conference at Rock Springs too) engaged in a vigorous volleyball game before the evening steak fry (top). Two things are obvious: First, Engineering Dean Ralph Nevins (right) was not anxious that the volleyball scores on the card be read, and second that Dean Carroll Hess (left) represents a well fed aggie team (bottom).

## LABORATORY IN TEST CONSTRUCTION AND EVALUATION

Dr. Dale Scannell and Dr. Larry Aleamoni

Each of the resource specialists gave a brief description of the operation and functions of the offices for which they work. Both have similar functions in assisting the individual instructors in preparation and grading of examinations as well as analyzing the results. All services are free of charge to instructors and departments since their services are budgeted in the overall university budget.

Following this presentation Dr. Aleamoni presented ideas and suggestions for preparing and administering exams. A publication from the Educational Testing Service, Princeton, New Jersey was used as a guide. Some of the suggestions are summarized below.

### Essay Tests

"Making Your Own Tests" focuses on objective tests of school achievement. This is not to suggest that free-writing exercises, which are integral tools of instruction, are not also useful in evaluation. They have great value when you want to find out how well students can supply a rather thorough analysis of a situation or how well they can express themselves in writing. However, within the necessary limitations on classroom testing, essay tests present at least three problems:

1. The number of topics you can cover within a class period is limited.
2. The reliability of essay grading leaves much to be desired.
3. The burden on the teacher of grading the essay papers of a whole class and returning them promptly is great.

It is suggested that essay questions be reserved for special papers assigned for work outside of class and to a few students at a time.

### Planning Your Objective Test

- Step 1: List the major topics covered in your particular teaching unit. This list should not exceed five.
- Step 2: Indicate the number of items you want to devote to each topic. (For convenience, "Making Your Own Tests" focuses on 50-item tests, if your items are unusually complex or you have a short testing period, you may want to use fewer items.)

Step 3: List under each topic the things you want students to know about, understand, or be able to do.

Step 4: Collect materials on which to base items. (Some of these materials may actually be presented on the test. Others will serve only as the basis of test items.)

Consider pictorial materials as well as textual materials, materials presented orally as well as materials presented visually.

Step 5: Begin writing the four choice items for your test.

Note that items may be phrased as questions or may be presented in completion form, that they may require students to select the best answer or the exception, that they may have specific stems or not, and that they may be presented in sets based on common material. Items may be written to measure remembering, understanding, and thinking.

### Mechanics of Item Writing

#### Item cards:

Professional item writers write each item on a separate 5 x 8 card. You may very well want to follow the same flexible procedure. At the bottom of the card, write the topic and skill categories for the item. Write the intended key (A, B, C, or D) on the back of the card. If a set of items is based on a single piece of material, that material may be written out (or described) on a separate card and clipped to the item cards that go with it.

#### Item stems:

If an item has a stem, the stem should meet these standards:

1. It should set the task. Generally the student should know before he reads the choices just what problem he must solve. Here are examples of good and poor item stems:

GOOD What is the inverse operation used to check division?

POOR Which of the following is true?

POOR Multiplication is . . .

2. The stem must allow for the number of choices you have decided upon. If you are committed to four choice items, the following stem might trick you into the absurd fourth choice shown.

When you add two even numbers, the result is:

- A always an even number      C sometimes an even number and  
sometimes an odd number
- B always an odd number      D neither an even number nor an  
odd number

3. The stem should be worded as briefly and clearly as possible.
4. The stem should be grammatical both within itself and in its relationship to the choices. If the ordinary rules of grammar aren't observed, this kind of thing can happen.

The roots of the equation  $x^2 - x - 20$  are

A - 4, + 5

B + 4, - 5

C 4

D 5

5. The stem should not include material which automatically determines the correct choice or rules out one or more incorrect choices. In the above example, "roots" and "are" in the stem tell the student who may not know anything about the solution of quadratic equations that the right answer cannot be C or D.

Item choices:

1. Like item stems, item choices should be brief and clear.
2. They should also be grammatical.
3. Item stems should not "give away" information. "Give aways" may include physical characteristics, as well as wording. For example, in the following item the right answer may be chosen for quite the wrong reason because it is shorter than any of the other choices.

According to the Stage Manager, the First Act of "Our Town" is called Daily Life; the Second Act, Love and Marriage. Which of these is the best name for the Third Act?

- A. Growing Up, Love, Marriage, and Children
- B. None Can Escape the Judgment Day
- C. Life in a Republican, Protestant Town
- D. Death

4. Item choices should be parallel in terms of grammar, as well as physical properties. The following item is faulty in almost every respect, including the lack of grammatical parallelism of choices:

Franklin Roosevelt was:

- A. the 14th President of the United States
- B. the son of Theodore Roosevelt
- C. successful in revising the complement of the Supreme Court
- D. elected to his first term by a majority of the popular vote.

Note the following example of extreme lack of parallelism of choices, in terms of content.

The first President of the United States was:

- A. Christopher Columbus
- B. George Washington
- C. John F. Kennedy
- D. Kangaroo

Incorrect choices:

For every correct choice you write, you will need to construct incorrect choices which have some plausibility to students with varying degrees of information or misinformation. There are three general approaches to writing such choices.

1. Administer open-end questions to students and have them supply their own answers. Then use the most frequent wrong answers as incorrect choices when you reshape the questions into multiple choice form for administration to another group of students. For example this question might be given:

On the arithmetic test, Bill got 32 as an answer to one problem. In working this problem, Bill's only mistake was multiplying by 4 in the last step when he should have divided by 4. What is the correct answer to the problem?

(This procedure is time-consuming and may not be possible in many school situations.)

2. Develop incorrect choices on the basis of your knowledge of common misconceptions students hold and errors which they make. For example, in the arithmetic question shown previously, a student might divide 32 by 4 to get 8, or subtract 4 from 32 to get 28, or multiply 32 by 4 to get 128. These numbers: 8, 28 and 128, are all good candidates for incorrect choices.
3. Think about reasonable confusions in terms of the stem or the correct answer. For example, in constructing a vocabulary test including the word "ingenuous", you might use the incorrect choice "clever."

Don't forget the possibility of including such choices as the following, but when you use them, make sure that they represent the correct answer in some cases.

It cannot be determined from the information given.

None of these

No error

All of the above

By all means, avoid trickery in constructing your items. For example, if you're testing student knowledge about the statement involved in some historical event, don't include and count as wrong a slightly misspelled version of a prominent participant's name.

#### Multiple Choice Item Types

Question.

If  $n$  is an odd number, what is the next larger odd number?

- A.  $n - 2$
- B.  $n - 1$
- C.  $n + 1$
- \*D.  $n + 2$

Completion.

In the Hindu civilization, the highest social class or caste is the:

- \*A. priests
- B. farmers
- C. soldiers
- D. merchants

Best answer.

The agency of the United Nations which is most like the United States Senate is the:

- \*A. General Assembly
- B. Security Council
- C. Secretariat
- D. International Court of Justice

Exception, a typical characteristic.

Which art reproduction would you be LEAST likely to find in Grover's Comers ("Our Town")?

- A. Whistler's "Mother"
- \*B. El Greco's "View of Toledo"
- C. "Washington Crossing the Delaware"
- D. "The Last Supper"

No specific stem.

In each item, choose the name that does not belong with the other three.

1.
  - A. Thomas Hardy
  - B. John Galsworthy
  - C. Thomas Wolfe
  - \*D. Arnold Barnett
2.
  - A. Aeschylus
  - B. Euripides
  - C. Aristophanes
  - \*D. Demosthenes

In each item, mark the line in which any error of punctuation or usage occurs; mark D if there is no error.

1. A. It's hard to believe that  
B. Both the paper and the book  
\*C. Was stolen from my desk  
D. No error
2. A. Each of the other boys entered  
B. His father's occupation, but  
C. I pursued a course of my own  
\*D. No error

Item Set.

Here is an excerpt from a student's summary of an article he had read:

(1) There is in England today just three classes, upper, middle, and lower. (2) The upper class have gone down a lot from what it was once. (3) You can only tell them now by their language. (4) This wasn't this way at all in Victoria and Edward's day. (5) Then the upper class was, in addition to language, richer, well-dressed, and better educated.

1. Which of the following would make the greatest improvement in sentence 1?
  - A. Starting with "In England today"
  - \*B. Substituting "are" for "is"
  - C. Substituting "only" for "just"
  - D. Adding "classes" after "lower"
2. In which of the following sentences could a colon be used to advantage?
  - \*A. Sentence 1
  - B. Sentence 2
  - C. Sentence 4
  - D. Sentence 5
3. Which sentence violates the principle of parallel structure?
  - A. Sentence 1
  - B. Sentence 3
  - C. Sentence 4
  - \*D. Sentence 5

Following discussion of ideas presented in the ETS\* publication Drs. Scannell and Aleamoni summarized the material with the following suggestions:

1. Our attempts should be to stretch out the scores as much as possible with differences being due to differences in achievement in the course. (The effectiveness of exams in accomplishing this can be measured by an item analysis, details for which are given in a publication Number 5 by ETS "Short Cut Statistics for Teacher Made Tests.")
2. The most effective test for discriminating purposes will be composed of items that are if possible about mid-difficulty. That is, about one half of the group should answer the items correctly. Items which are answered, or missed, by almost everyone contribute very little to test discrimination.
3. Organize items by content or concepts on the exam.
4. Items requiring the most time to answer should be kept toward the end of the test.
5. In writing test items (multiple choice) it is easiest to write the stem first, the correct answer next, and the distractors. This alternative should then be assigned at random by use of random table, etc.
6. Directions for the test should be as complete as possible such as how much time is available, number of points on each question, how to mark the answers, how it will be scored and anything that will influence the examinee's performance. Always use a separate answer sheet regardless of the kind of questions used.
7. Tests should be administered with as little distraction as possible.

#### Questions and Discussion

- Q. Do you use alternate forms or questions when giving exams to large classes where students cannot be seated in alternate seats?
- A. Yes, we even prepare four form tests where the same questions are arranged in four different orders.
- Q. What is the advantage of a four response multiple choice test over the free response (completion) for testing the same time, or subject, but require one word answers?

\* Educational Testing Service, Princeton, New Jersey.

- A. (1) Ambiguous questions are easier to answer with multiple choice type. (2) Students interpret discussion questions differently and in a multiple choice item you can provide reasonable alternatives from which the student only selects one answer and (3) the question can be more specific.
- Q. Is it possible to have more than one correct answer to a multiple choice question?
- A. No, when the approach is necessary it is best to use two true false questions. This is true if you have two correct responses in mind. It is too difficult to grade this kind of multiple choice question.
- Q. How can tests be used for motivation of students?
- A. Occasional tests are good for this purpose. It is better to maintain a mild constant level of motivation rather than to go along at a low level then jumping to a high. Learning is better under the first case. This means fairly evenly spaced tests, some being small quizzes serves this purpose.
- Q. Should short quizzes be announced prior to the time they are given?
- A. You will get a better indication of student learning from announced quizzes. Unannounced quizzes can be effective in maintaining student effort, but they don't give very good information for measurement purposes.

Concluding Remarks Concerning Multiple Choice Questions  
Dr. Dale Scannell

One of the virtues of a multiple choice test is that you can modify quite easily the difficulty of the items by modifying the responses you give. In general the more similar the responses are the more difficult the item is going to be and the more dissimilar the responses, the easier the item will be. This means you can adjust the appropriate level of difficulty to the students in class. As an example, what year did Texas gain statehood? (a) 1835, (b) 1840, (c) 1845, (d) 1850. With this small a range, only five years difference, the person really had to know American history. If the span had been: (a) 1730, (b) 1775, (c) 1845, (d) 1910, the last question would be much easier. This shows how you can modify questions to accomplish your objective.

Multiple choice tests are excellent for definitions of terms. The best approach is to present the term to be defined in the stem then four or five reasonable answers or responses.

They can be used to measure a student's knowledge of facts, generalizations, ability to apply these in given situations or solving problems, ability to formulate hypothesis, ability to recognize when data are sufficient, etc. You might be happily surprised that multiple choice items can do a good job for you in your courses.

## DISCUSSION GROUPS\*

Group 1      How might students be involved in formulating policies on testing and grading?

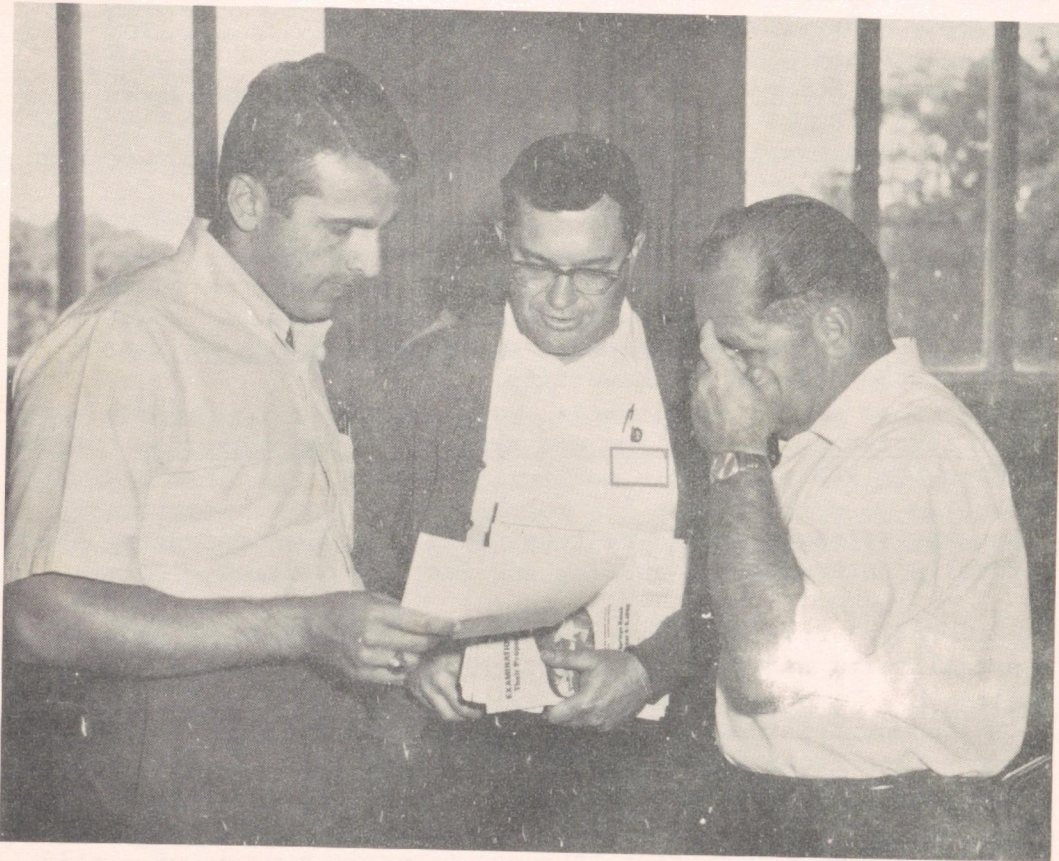
F. R. Anstaedt - Grain Science & Industry  
J. J. Albract - Ag Education  
F. R. Carpenter - Administration  
J. K. Greig, Jr. - Horticulture & Forestry (chairman)  
J. A. Hobbs - Agronomy  
Jim Hoover - Animal Science & Industry  
M. L. Manuel - Ag Economics  
R. B. Mills - Entomology (recorder)  
D. J. Mugler - Dairy & Poultry Science  
W. S. Ruliffson - Biochemistry  
R. R. Schalles - Animal Science & Industry

Decided to discuss the question in two parts: practices in the classroom and policies of the college or university.

Student involvement in testing and grading in the classroom. Following are ideas from the group:

1. Student opinions should be considered, but the instructor should finally determine the grading and testing practices.
2. Get student opinions after the course is finished to be used in establishing and revising practices. Use opinions of former students.
3. If student opinions are to be used in current semester, get them early then establish practices to be followed and make certain students are informed.
4. Discussion of tests after they are taken is a means of getting student suggestions.
5. Students could make up questions as an outside assignment and the instructor use these as aids in test construction.
6. Give students "extra points" for "extra activities," e.g., attending judging schools, getting a story in Collegian, etc. By this means a student can influence his grade.
7. Let students indicate at the end of the semester the grade he thinks he deserves. Use this not for alteration of the instructor's grade but only as another student opinion.
8. One departmental club has discussed course and curriculum matters. Grades and testing could be discussed by other such clubs.

\*Resource persons and Dean Hess visited all discussion groups for brief periods.



Wilber Ringler (center), Chairman of the Group Discussions, consults with Larry Aleamoni (left) and Dean Carroll Hess (right) concerning their participation in these groups.

The consensus of opinion on the above was that each instructor should be sensitive to student opinions and suggestions and use these as he sees fit to improve his grading and testing.

Policies of College or University which involve students.

1. The university now has policies on grading and testing, some of which were originated and encouraged by students, e.g., pass-fail, final exam requirements, grading scale, listing of instructors for sections of classes, etc.
2. It was mentioned that lawsuits had been filed against institutions where students felt that established written university policy had not been followed.
3. Scannell stated that there should be a written policy on these items and that students and instructors should be made aware of them. Also a policy should be established under which a student may appeal if he feels he has been treated unfairly. Students should be represented when these policies are established.

Other comments

1. Colleges and universities will need to have established policies to counteract the students which want complete control of the universities. Again students should be represented in establishing these.
2. The group unanimously encourages the use of student evaluations of instruction. These are to be used by the instructor for self-improvement.

Conclusion: Student's opinions and suggestions should be used by the instructor, as he wishes, to aid him in improving his testing and grading. Students should be represented in the establishing of policies by the university on grading and testing.

Group 2      What should be the role of examinations?

A. W. Adams - Dairy & Poultry Science  
 H. R. Bradley - Ag. Education  
 E. P. Farrell - Grain Science & Industry (recorder)  
 C. V. Hall - Horticulture & Forestry  
 C. V. Hess - Administration  
 Jack Jordon - NDEA Fellow  
 G. H. Kiracofe - Animal Science & Industry  
 E. L. Mader - Agronomy  
 Ross Mickelsen - Dairy Science  
 C. L. Norton - Dairy & Poultry Science (chairman)  
 E. F. Smith - Animal Science & Industry  
 H. E. Thompson - Entomology

It was agreed that the six reasons for examinations given by Dr. Dale P. Scannall are acceptable. They are:

- 1) To assign grades to the students.
- 2) To diagnose learning difficulties.
- 3) To determine if a minimum mastery of the subject has been achieved.
- 4) To motivate students.
- 5) To enhance the learning process.
- 6) To measure effectiveness of teaching.

All agreed that grades are necessary and that exams are an important aid in arriving at a true grade.

All agreed that, if by some chance the university no longer required or used grades, they would continue to use examinations for other purposes. Exams would be used to motivate students and to measure effectiveness of teaching even if grades were eliminated.

Exams early in the course could be an aid to the instructor in diagnosing learning difficulties of the student.

Exams over course work with no corrections or grades would be quite ineffective.

Some thought that the prompt review of errors on exams will be an aid in teaching the student.

Pre-testing: Some thought that one or more exams given at the start of a course could guide the instructor in how or what he offers in the early part of the course. This would supplement prior fulfillment of prerequisites.

The assignment of grades to two or more students with equal numerical scores close to the division between an "A" and a "B" or a "B" and a "C"

was considered a problem. One suggested that the instructor could offer another exam or assignment to all who are tied at the dividing line. The grade on this extra exam or assignment could possibly differentiate between those tied. Students might be persuaded that exams are important by reminding them that all through life they are being tested for performance by their employers, business and social contacts.

Group 3      What criteria and standards should be followed for determining course grades?

Richard Bassette - Dairy & Poultry Science  
 Ken Jorns - Student Panel  
 K. A. Huston - Dairy Science  
 Paul L. Kelley - Ag Economics  
 B. A. Koch - Animal Science & Industry  
 R. I. Lipper - Ag Engineering  
 C. E. Long - Horticulture & Forestry  
 G. D. Miller - Grain Science & Industry  
 H. Roberts - Dairy & Poultry Science (recorder)  
 H. Tuma - Animal Science & Industry  
 D. A. Wilbur - Entomology  
 V. Withee - Agronomy (chairman)

The discussion was begun by asking the group if we should report grades. The group agreed that the university requires a report of grades earned by each student for the course under study.

To evaluate the performance of a student, some form of grading must be used. If a 5 point system (A,B,C,D, and F) is used, then what is an "A" student or a "B" student? It was felt that this system is based upon an average. The question was then asked - What is an average student? It was agreed that there is no set system for the university and that the instructor must use his own judgment in determining what constitutes an average student. Some panel members felt that a grade given to a student may be based upon previous classes and not upon the class presently under consideration. Some discussion on the size of class in relation to determining grades for each student was discussed. Smaller classes create greater problems in assigning grades while large classes (20 or more) present a more definite division between grades. A better curve should be realized with a larger class.

It was pointed out that other criteria should be considered in evaluating the student. This should be left up to the individual instructor in determining these points.

A student member of the group expressed the view that students should be informed of the grading system early in the course. He also recommended

that students should be informed of their grades during the semester so that improvements may be made.

The pass or fail (2 point) system was felt inadequate for the proper evaluation of the students. If such a system were employed, the fail level should be "C" and not "D".

In summary, all members agreed that some form of grading system should be retained and that the present 5 point grading system is adequate. It was felt that no more than this number should be used and that fewer than 5 may be advantageous. Opinions were that the 2 point system does not give a true evaluation of the student.

The five point system is based upon an average (C) as determined by the instructor on the basis of past experiences. This average tends to move upward in advance courses from the freshman year to the senior year.

The group felt that students in their major field were protected by the instructor because of a closer association with the student in and out of class. The judgment of the instructor was felt to be necessary in determining the grade of the student. All members suggested that the performance of the student determine the grade given him.

Group 4      Should the College of Agriculture adopt an honor system of taking examinations?

J. G. Berry - Poultry Science  
 D. Bradley - Economics  
 B. E. Brent - Animal Science & Industry (chairman)  
 Michael Burns - Student Panel  
 J. V. Craig - Dairy & Poultry Science (recorder)  
 D. H. Kropf - Animal Science & Industry  
 E. B. Macy - Ag Journalism  
 J. H. McCoy - Ag Economics  
 N. W. Miles - Horticulture & Forestry  
 A. B. Ward - Grain Science & Industry  
 J. Wheat - Animal Science & Industry  
 G. E. Wilde - Entomology

A preliminary discussion of faculty experience with and desirability of honor examinations indicated no majority preference. The University of Minnesota system was presented by a former Minnesota student and discussed. Opinions were expressed that cheating is a minor problem at Kansas State University. It was the majority opinion that cheating would probably be inhibited about as effectively by the honor system as by proctoring of examinations. Success of the system was believed to hinge on favorable student attitudes. It was then concluded that the honor system orientation material from the University of Minnesota should be made available for examination by members of the Agricultural Student Council. The discussion group believed that the honor system of examinations should be given objective consideration by the Agricultural Faculty if the students want it.

Group 5 Some universities adjust final course grades for differences in grading standards among faculty, departments, and colleges. Discuss the merit of such a system for KSU.

O. W. Bidwell - Agronomy (recorder)  
 H. D. Blocker - Entomology  
 R. K. Burkhard - Biochemistry  
 Robert Dobson - Student Panel  
 C. L. Drake - Animal Science & Industry  
 A. H. Kahrs - Dairy & Poultry Science  
 Miles McKee - Animal Science & Industry (chairman)  
 S. T. McCready - NDEA Fellow  
 George Montgomery - Ag Economics  
 R. E. Odom - Horticulture & Forestry  
 W. L. Prawl - Ext. Of. Programs & Training  
 R. Sucher - Grain Science & Industry

I. Merits of a grade standardization system at KSU:

1. A grade-standardization system (GSS) is designed to improve the standard of measurement and is valuable in interpreting student grades. The GSS makes it theoretically possible to compare levels of proficiency among students . . .
  - . . . in different sections of the same course;
  - . . . in different courses of the same department;
  - . . . in courses of different departments of the same college;
  - . . . in courses of different colleges; and
  - . . . in courses of different universities.
2. The GSS may be valuable in comparing grades of students of similar departments of different universities.
3. The GSS may help eliminate the problems associated with differences of grading vigor among instructors, departments, and possibly colleges.
4. The committee believes that the GSS may have its greatest value in large classes and those of the Freshman and Sophomore year. The committee believes that it would be more difficult to administer the GSS in classes of small numbers (less than 50) and those of upper-classmen and graduate students.
5. The GSS may provide a most useful standard for evaluating students for graduate school.
6. The GSS may provide an appropriate means by which the university may upgrade its instructional effort.

II. Problems related to the establishment of a grade stabilization system.

1. The mechanical conversion of grades by a disinterested party or agency of the university might reduce the good teacher-student relationship that has existed in the College of Agriculture.
2. The GSS might affect that aspect of the student's personality related to later success in life . . . a quality not highly correlated with college grades.
3. The design and execution of a grade stabilization system would be very difficult, particularly across colleges, universities, and in some cases across departments within a college.

Recommendation:

The committee recommends that the College of Agriculture consider cooperating with Don Hoyt of the University Office of Educational Research in attempting an experimental grade stabilization program as a means of improving the instructional effort.

Group 6      What factors should be included in a "composite evaluation" of a college graduate?

Roscoe Ellis, Jr. - Agronomy (chairman)  
 R. J. Elzinga - Entomology  
 Gale Fuller - Student Panel  
 R. W. Funsch - Horticulture & Forestry  
 C. Melton - Animal Science & Industry  
 H. L. Mitchell - Biochemistry  
 J. D. Mitchell - Dairy & Poultry Science  
 Frank Orazem - Ag. Economics (recorder)  
 Curtis Trent - Ext. Of. Programs & Trg.  
 G. T. Wilson - Horticulture & Forestry  
 Walter Smith - Animal Science

Discussion

The members of the group realize that the grade point average achieved by a student may be a poor indicator for measuring or predicting student's accomplishments or performance in the future. It should not be used as a sole criterion for evaluating student's accomplishments of the past. There are other factors which may have an important bearing on the student's performance in professional employment. The relative significance of these factors may vary according to differences in jobs to be performed in various industries or governmental agencies. The employers may set their own

criteria and select the factors they deem important for an employee to meet in order to discharge the job's responsibilities satisfactorily.

Some of the factors, other than the grade point average, which may be used for a "composite evaluation" of a college graduate are these:

- Understanding of problem solving processes
- Ability of getting along with other people
- Initiative
- Professional attitude
- Responsibility conscientiousness
- Participation in campus activities (departmental clubs, student government, etc.)
- Memberships in professional and honorary organizations, and leadership positions held
- Background - rural, urban
- Employment as a student
- Motivation and adaptability
- Character
- Courses (it was emphasized that all problem courses taken should reflect the nature of the problem on the student's manuscript)
- Hobbies and, if permissible, the high school rank

The difficulty of an accurate evaluation of an over-all ability and potential for any one student was recognized but at the same time it was pointed out that further studies may help select factors or variables which together with the grade point average may make student's evaluation more meaningful.

The group recommends that the College of Agriculture consider developing a form which would be used for compilation of information related to an evaluation of every student in the College of Agriculture. This should be done with professional help, such as, from College of Education and the University's Counselling Center. The responsibility of gathering the needed information should rest with individual departments.

Group 7 Some universities have an office of instructional resources to help faculty in evaluating test construction, test procedures, help design instructional aids, and provide an overall consulting service for improving teaching effectiveness. Discuss the merit of such an office at KSU.

- W. H. Borst - 4-H
- K. C. Feltner - Agronomy (recorder)
- Ted Fisher - NDEA Fellow
- W. A. Geyer - Horticulture & Forestry
- L. Harbers - Animal Science & Industry
- T. L. Hopkins - Entomology
- Brad Kerbs - Student Panel
- C. S. Menzies - Animal Science & Industry
- J. L. Morrill - Dairy & Poultry Science
- D. B. Parrish - Biochemistry (chairman)
- D. L. Stuteville - Plant Pathology
- I. E. Werner - Ag Economics

Group 7 was unanimous in its belief that a real need exists for many services that might be provided by an "Office of Instructional Services." Moreover, most expressed the belief that services made available by such an office would be made use of by departments represented by the group.

Some thought was given to the nature of services that ought to initially be offered at KSU by such an office. Some suggestions:

1. Orientation for new classroom instructors to include information relating a) kinds of teaching and evaluating services available and b) principles underlying sound course blueprinting.
2. Consultation relating to effective instruction (i.e., classroom techniques, examination construction, self-evaluation, visual and audio aids) for new and existing staff.
3. Could serve as an aware but confidential agency for collection, analysis and interpretation of instruments used by students to evaluate courses and instructors.
4. Computer programs for grading examinations and for providing minimal statistics to provide an evaluation of testing tools.

The group recognized a concomitant need for better availability of instructional aids (slides, transparencies, tapes, programmed instruction), and decided that these services should be in close physical proximity to the "Office of Instructional Services" if not an integral part of it.

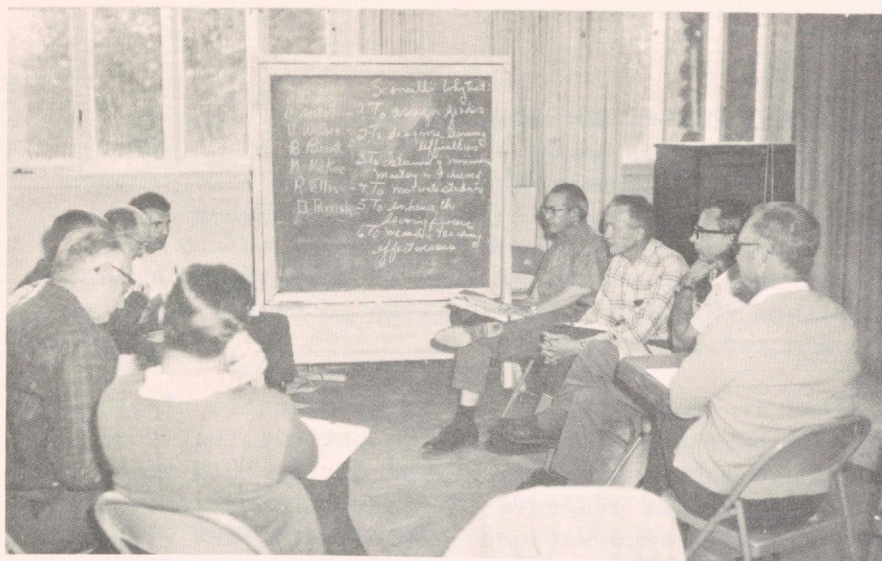
Disadvantages posed by various members of the group included:

1. Test items that lend themselves to computer analysis are not preferred by many instructors, particularly in small classes.
2. Lines of communication necessary to effectively use such a service would require more, not less, faculty teaching time.

The group formulated a list of precautions that they recommend subsequent planning groups take into consideration. They are:

1. Structure the operational procedure to protect against classroom teachers becoming preoccupied with educational statistics at the expense of good teaching.
2. Recognize that such services will not substitute for teaching time but rather will increase it.
3. Recognize that such services will not alleviate serious shortcomings that now exist in the amount and nature of physical teaching facilities.
4. Ongoing programs could suffer budgetarily if such an office came about through "skimming from existing moneys." Therefore, new line-budget funding should be sought.

Finally, the group concurred in the belief that a) a sincere interest in teaching is growing among university teachers across the nation, and b) this faculty interest is the prime requisite for the success of an instructional service unit.



All participants were divided into one of seven discussion groups each with an assigned topic. The discussion group chairman is meeting with individual group chairmen and recorder (top), one group is shown in session (center) and group chairmen answering questions from the audience following oral reports (bottom).

## PSYCHOLOGICAL IMPLICATIONS

Dr. Dale Scannell

There has been a lot of discussion relating to anxiety and tests. However, before we get into this topic I would like to relate results of a study that was conducted in California.

Some people selected at random two groups of students in classes. They told the instructors that there was reason to believe that some students in the class would make remarkable growth during this year. They gave the instructors names of students in class which were selected at random. A control group was also established, but without telling the instructors. Sure enough the group which they had informed the teacher about did make more growth during the year than the control group. I don't know what this says to us, but apparently we have some subtle kinds of communication with other individuals that we are unaware of.

Back to the topic of anxiety which is produced by tests. Our purpose is to cover a variety of ideas related to the psychological implications of examinations. Psychologists have known for a long time that performance is facilitated by a certain degree of excitation in subjects. People who are mildly anxious tend to out-perform those with comparable ability who are not motivated at all. We talk about anxiety as though it has a detrimental effect on student performance on exams, when in fact the research suggests to us that a certain amount of anxiety may enhance the performance of an individual on an examination. However, the studies that have been conducted on the general topic of test anxiety have not produced conclusive results. But, from the results available there are a few generalizations we might pose with some reluctance. The results seem to support several observations.

1. The amount of anxiety that the student has while taking a test is negatively related with his ability. That is, low ability students tend to be more anxious than high ability students.
2. In addition there tends to be a positive correlation between level of anxiety during a test and level of aspiration. The higher one wants to perform the more anxious he will be.
3. Anxiety among students taking a test tends to decrease rapidly with experience that students have with a particular type of test. The more often they are exposed to the examination situation, the less anxious they become in that particular kind of situation.
4. A kind of general conclusion might be that a certain amount of mild anxiety will facilitate performance, but when the anxiety reaches a higher point it begins to interfere with performance on a test. Many people recommend that in teaching you try to maintain the motivation of your students at a mild level and distribute it throughout the semester rather than have it peak at a certain time just prior to taking an examination.

I don't know how concerned we should be as instructors in attempting to remove anxiety among students. I do think we should attempt to eliminate unnecessary frustrations that create anxiety as much as possible. Some things that can be done are: (1) prepare good clean exam copy, express ideas clearly and include complete instructions, (2) have one copy per student, (3) announce exams in advance and tell students what subjects the exam will cover and (4) give fair questions.

### Questions and Discussion

- Q. What are the psychological effects of an exam which is too long for students to complete in the allotted time, even though they are told they will actually only be graded on the basis of those completed?
- A. Results show that the highest reliability, or best performance comes when 85 to 90 percent of the students complete the exam in the allotted time. Almost all standardized tests are designed with this in mind. I would suggest that we be more generous than that with classroom tests so that almost all students will have an opportunity to complete the exam. Give ample time even though there will almost always be a few students who will take all of the time available and more.
- Q. How valid is a student's complaint when he has two one hour exams on the same day when he has had ample time to study for both?
- A. Aleamoni - This depends somewhat on when the exams are given. If the student has prepared himself well enough he should be able to handle both examinations. As a rule students know that most mid-term and final exams come during a short period of time and we know that what is learned over a long period of time is retained longer than that which is crammed during a short period of time. Therefore, students should keep up on studying throughout the semester rather than cram the night before a test. This should reduce anxiety in the case of two exams due on the same day.
- Q. Should you tell students what kind of questions will be included on an examination?
- A. Scannell - Yes, even give them sample type questions.
- Q. Should you use a correction for guessing formula on objective type tests?
- A. Discussion of pros and cons by both resource persons, but the conclusion was in most cases, it is not necessary.

- Q. Comment on pop quiz effects or influences on students.
- A. Aleamoni - Unannounced pop quizzes raise the anxiety levels of students. Of course you can announce at the beginning of the semester that you will give weekly quizzes and avoid this reaction to a certain extent. My main objection is that you may be attempting to measure something the student has not learned. Generally, I am not in favor of unannounced pop quizzes. They do help for motivational purposes, but not measurement.
- Q. How long does it take to write a good multiple choice question?
- A. Discussion by Scannell, Aleamoni and Hoyt - The conclusion was that if you could write three or four good questions per hour that you would be doing well. In many cases one per hour would be normal. However, an instructor's skill and speed will develop with practice. The main point is --- don't put off test construction until the zero hour. Write a few items every day.
- Q. Should a different grading standard be used for graduate and undergraduate students when in the same class?
- A. All three resource specialists discussed the question and generally agreed that there should be no justification for this if the course was properly organized.
- Q. When do you write your test questions?
- A. Aleamoni - As I go along, preferably after each lecture. Then the questions are edited later in order to improve wording, etc.
- Scannell - I tell people that too.
- Q. What about other forms of anxiety such as financial or personal problems?
- A. Scannell - We know that there are personal factors which would interfere with performance on a test. On any given day there are factors unrelated to the examination which may cause a student to perform differently than on another day, or test. This is why grades from a number of exams present a more reliable estimate of the student's ability than do scores from a single test given on a single day.
- Q. What kinds of responses result from anxiety and how much of this is produced from examinations?

- A. Scannell - There are many responses to anxiety and one of the common ones on examinations is that some students resort to cheating. They try to avoid the problem in this way. The best way to avoid anxiety is by attacking the problem itself, take the exam and get it out of the way.
- Q. How do you feel about the honor system as it relates to cheating and relief from anxiety among students?
- A. Scannell - Personally I feel that the anxiety level among students who would be required to proctor other students would be greater. The fewer distractions a student can have when taking an exam, the lower the level of anxiety. Therefore, I don't think it is right for one student to be required or concerned about what another student is doing while he is taking the exam. Instructors can and should proctor the exam because it is almost impossible to write an exam that all students will understand completely. Therefore, the instructor should be there to answer questions and his walking around the room should not bother the students, unless someone has something to hide. If this is the case he should be bothered.

## CONCLUDING REMARKS

Dean Carroll V. Hess

My concluding remarks will be very brief. As I listened to the discussions these past two days, I listed several desirable impacts which I think this symposium can have upon the instructional program in our College. They are:

1. That you now have a greater awareness of the broadened scope, purposes and complexity of designing good examinations. You must certainly recognize that their proper design and utilization is an integral part of good teaching.
2. That you will take away from this symposium some better guidelines to: (a) improved tests, (b) construction and administration of exams, (c) desirability of test analysis and proper procedures to be followed, (d) assignment of grades, and (e) possible psychological impacts of exams on students.
3. That you leave this symposium with an improved attitude toward the adoption of a procedure for evaluating your course and instruction in the interest of instructor self-improvement.

I believe the chief function of the Deans office to be one of stimulating and facilitating a variety of efforts designed to improve the quality of the educational experience of our students. I welcome your bombardment of my office with ideas toward this end.

Too frequently, we remind the student of his responsibilities in the attainment of an education--whereas we also need to remind ourselves, as teachers and administrators, of our responsibilities to the student on this educational process.

I sleep well knowing that you, the faculty in agriculture, are very cognizant of your responsibilities to the students. More importantly, you reflect an enthusiasm for attaining even higher levels of teaching achievement. Your participation in this symposium and past instructional seminars on campus attest to this.

Let us return to campus charging ourselves and each other with a greater determination to keep the educational needs and aspirations of our students uppermost in our minds and be guided accordingly.

Thank you again for your deep interest and attention in this symposium.



Families of the participants and invited guests were treated to a chicken barbecue with all the trimmings on the final afternoon of the session.

Some of the invited guests shown left to right in bottom photo are Drs. John Noonan, Associate Dean of Graduate School, Robert Kruh, Dean of Graduate School, John L. Brown, Vice President for Academic Affairs, with their families. In right foreground is Mrs. Carroll Hess.

Participants in Symposium and Guests

ADMINISTRATION

G. H. Beck  
F. R. Carpenter  
C. V. Hess  
V. Larson

AGRICULTURAL ECONOMICS & ECONOMICS

P. L. Kelley  
J. H. McCoy  
George Montgomery  
Frank Orazem  
I. E. Werner  
M. Manuel  
D. Bradley

AGRICULTURAL EDUCATION

J. J. Albracht  
H. R. Bradley

AGRICULTURAL ENGINEERING

R. I. Lipper  
George Larson

AGRICULTURAL JOURNALISM

E. B. Macy

AGRONOMY

O. W. Bidwell  
Roscoe Ellis, Jr.  
K. C. Feltner  
J. A. Hobbs  
E. L. Mader  
V. Withee  
J. Vesecky

ANIMAL SCIENCE & INDUSTRY

B. E. Brent  
C. L. Drake  
D. L. Good  
L. Harbers  
R. H. Hines  
Jim Hoover  
G. H. Kiracofe  
B. A. Koch  
D. H. Kropf

ANIMAL SCIENCE & INDUSTRY (CONT.)

Miles McKee  
C. Melton  
C. S. Menzies  
R. R. Schalles  
E. F. Smith  
W. Smith  
H. Tuma  
J. Wheat

BIOCHEMISTRY

R. K. Burkhard  
H. L. Mitchell  
D. B. Parrish  
W. S. Ruliffson

DAIRY & POULTRY SCIENCE

R. Mickelsen  
A. W. Adams  
Richard Bassette  
J. W. Craig  
A. H. Kahrs  
J. D. Mitchell  
J. L. Morrill  
D. J. Mugler  
C. L. Norton  
H. Roberts  
K. A. Huston  
J. Berry

ENTOMOLOGY

H. D. Blocker  
R. J. Elzinga  
T. L. Hopkins  
R. B. Mills  
H. E. Thompson  
D. A. Wilbur  
G. E. Wilde

EXTENSION OFFICE OF PROGRAMS &  
TRAINING

W. L. Prawl  
W. E. Ringler  
Curtis Trent

## 4-H

W. H. Borst

## GRAIN SCIENCE &amp; INDUSTRY

F. R. Anstaett

E. P. Farrell

G. D. Miller

A. B. Ward

R. Sucher

## HORTICULTURE &amp; FORESTRY

R. W. Funsch

W. A. Geyer

J. K. Greig, Jr.

C. V. Hall

C. E. Long

N. W. Miles

R. E. Odom

G. T. Wilson

## PLANT PATHOLOGY

D. L. Stuteville

## NDEA FELLOWS

Ken Sorensen

Jack Jordan

Jan Cipra

Ted Fisher

Stuart McCreedy

## STUDENT PANEL

Ken Jorns

Michael Burns

Robert Dobson

Gale Fuller

Brad Kerbs

## RESOURCE PERSONNEL

Dale P. Scannell

Lawrence M. Aleamoni

Donald Hoyt

## GUESTS

Roy Freeland

Robert Kruh

James McComas

Chet Peters

John L. Brown

John Noonan

Floyd Smith

Dick Owens

Harold Jones

G. Larson

L. Erpelding

E. Baugher

P. Stevenson

R. W. Campbell

A. Moore

H. Knutson

## Suggested References

University of Minnesota, Bureau of Institutional Research, Bulletin on Classroom Testing, Number 1. "Why Do We Test?" by John E. Stecklein.

Bulletin on Classroom Testing, Number 2. "Essay Tests: Why and How?" by John E. Stecklein.

Bulletin on Classroom Testing, Number 3. "What is a Good Test?" by John E. Stecklein.

Bulletin on Classroom Testing, Number 4. "How to Write Multiple-Choice Test Items." by John E. Stecklein.

Bulletin on Classroom Testing, Number 5. "How to Write True-False Test Items." by John E. Stecklein.

Bulletin on Classroom Testing, Number 6. "How to Write Matching Test Items." by John E. Stecklein.

Bulletin on Classroom Testing, Number 7. "How to Measure More Than Facts with Multiple Choice Items." by John E. Stecklein.

Bulletin on Classroom Testing, Number 8. "How to Make An Item Analysis of An Objective Test." by John E. Stecklein.

Bulletin on Classroom Testing, Number 9. "How to Make a 'Content-Objectives' Test Analysis." by Charles Schumacher.

Bulletin on Classroom Testing, Number 10. "How to Make An Item Analysis of An Essay or Problem Test." by Robert L. Lathrop.

Examinations Service Bulletin, Number 11. "A Point of View About Grading." by Robert L. Lathrop.

Examinations Service Bulletin, Number 12. "Grading Operations--A Simplified Approach." by Robert L. Lathrop.

Testing Bulletin No. 1. "Essay Tests - General Considerations." Prepared by The Board of Examiners, Michigan State University, February, 1956.

Testing Bulletin No. 5. "The Preparation and Use of Objective Tests." Prepared by the Office of Evaluation Services, The Basic College, Michigan State University, May, 1957.

Testing Bulletin No. 6. "The Preparation and Use of Objective Tests." Prepared by the Office of Evaluation Services, The Basic College, Michigan State University, November, 1957.

Testing Bulletin No. 7. "Oral Examinations." Prepared by the Office of Evaluation Services, The Basic College, Michigan State University, June, 1958.

"Constructing and Using Examinations for Evaluation of Teaching." by John F. Thompson, Department of Agricultural and Extension Education, University of Wisconsin, Madison, May, 1968.

"Making Your Own Tests." Developed by Cooperative Test Division, Educational Testing Service, Princeton, New Jersey.

"ETS Builds a Test." Educational Testing Service, Princeton, New Jersey.

"Multiple-Choice Questions: A Close Look." Educational Testing Service, Princeton, New Jersey.

"Selecting An Achievement Test: Principles and Procedures." Number 3, Evaluation and Advisory Service Series, Educational Testing Service, Princeton, New Jersey.

"Making the Classroom Test: A Guide for Teachers." Number 4, Evaluation and Advisory Service Series, Educational Testing Service, Princeton, New Jersey.

"Short-cut Statistics for Teacher-Made Tests." Number 5, Evaluation and Advisory Service Series, Educational Testing Service, Princeton, New Jersey.

