DOCUMENT RESUME

ED 273 168 HE 019 505

AUTHOR MacCuish, Donald A.
TITLE A Quality Assurance Model for Higher Education: A Pilot Study.
PUB DATE 12 Feb 86
NOTE 17p.; Paper prepared for the Testing and Quality Assurance in Higher Education Conference (Miami, FL, February 12, 1986). For related document, see HE 019 506.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accountability; Classification; *College Faculty; Course Descriptions; *Course Objectives; *Educational Quality; Faculty Evaluation; Higher Education; Student Evaluation; *Teacher Made Tests; Test Construction; *Test Reliability

ABSTRACT

A model for assuring quality in the development of course objectives and classroom and exit examinations is presented. The model was based on a pilot study with 131 faculty at the University of Central Florida. It was found that 91% of teaching faculty create 100% of the tests they use to evaluate student performance. The faculty seemed to use course descriptions fairly regularly to develop course objectives. Faculty do not use a taxonomy in developing teacher-made test. However, few faculty obtain data on the reliability of their testing devices. According to the model, faculty members would have a good course description from which they would develop realistic and attainable course objectives. The course objectives would then be used to develop classroom and exit examinations, as well as course content. A taxonomy of testing, such as Bloom's, should then be used as a guide to constructing tests. Reliability data derived from test evaluation should be used to improve instruction and the measuring devise itself. Since tests determine whether the student has mastered educational objectives, use of this model would be a part of a quality assurance program. The model would also provide one objective and measurable input to the complex process of faculty evaluation. (SW)

A Quality Assurance Model for Higher Education: A Pilot Study

by

Donald A. MacCuish
Manager
Southern Operations
ONLINE Computer Systems
20251 Century Blvd.
Germantown, MD 20874

A Paper Prepared for the
Testing and Quality Assurance
in Higher Education
Conference

February 12, 1986

-Abstract-

Two of the growing issues in higher education are faculty accountability and faculty evaluation. These two issues have important implications for testing, particularly classroom and exit examinations, and quality assurance. This paper illustrates how these seemingly divergent issues are interrelated. You cannot have faculty accountability or meaningful faculty evaluations without well-developed classroom and exit examinations or a good quality assurance program.

Jerome Bruner, in the Process of Education, discusses the concept of "spiral curriculum" meaning that a thought, idea, or concept trancends courses and leads to more advanced thoughts, ideas, or concepts. If we look at any college curricula, this "spiral curriculum" concept is vividly illustrated in the course sequencing within educational programs. The purpose of the classroom examination is to determine whether or not the student has mastered those thoughts, ideas, and concepts well enough to move to the next stage. This is an interim quality controls check. At the end of the educational experience, some institutions require final quality assurance checks; qualification examinations are an example.

The question, then, is how do we insure that the quality assurance process works? To simply require classroom or exit examinations is inappropriate. Quality assurance is a process. Industry has understood this for years. Industry may not have implemented the concept correctly, but it has been understood. Education must also implement and maintain a scientifically based testing and quality assurance program.

This paper discusses how this can be accomplished. In the Spring of 1985, a pilot study was conducted at the University of Central Florida (UCF). The

original purpose of this study was to determine whether or not college faculty really used the statistical concepts they stressed in the classroom. The literature suggesting that teachers in the public schools are remiss if they do not evaluate their testing devices, is readily available in the professional journals. Suggestions on how to do statistical analysis is also widely reported in the literature. The interesting fact is that the college faculty that write about statistics in evaluation do not follow their own techniques.

According to this study, 91% of the faculty at UCF develop 100% of the tests they use in evaluation of student performance in their classes. The average number of tests created per course is 4.122, while the average number given per course is 4.069. The study also reveals that faculty members seldom obtain reliability data on their teacher made tests.

How do you have quality assurance if the method of evaluation is not evaluated? Further analysis of the data indicates some very interesting factors which bring this entire process together. The trends discovered in this pilot study suggest that we can design, develop, and implement a well-planned and organized testing and quality assurance program in higher education by using the course development model described in this paper. This program, if implemented, brings into focus faculty accountability and can be one input to the faculty evaluation process.

Introduction

Lyman A. Glenny and Frank A. Schmidtlein have written that as the student population declines and States scrutinize their budgets more closely, State legislatures and departments of higher education will, out of necessity, enter the State college system demanding strict accountability. Glenny and Schmidtlein beleive one of the most likely areas of State invasion will be

dictated criteria for evaluating new and existing academic programs to include faculty effectiveness. They also believe that State legislatures will oversee the administration of evaluation procedures for pay increases, lay-offs, promotions, etc. The final area of intrusion will be student preparedness to complete tasks in their fields of specialization.

Holley and others, support the views expressed by Glenny and Schmidtlein. Holley and his associates, however, say that the intrusion into the hallowed halls of academia has already begun. They cite as examples, Title VII of the Civil Rights Act of 1964 which became effective at the college level in 1979 as a result of the U.S. Supreme Court case Board of Trustees vs Sweeney (1979). They also cite the Jepsen vs Florida Board of Regents (1980) 5th Circuit Court of Appeals decision pertaining to the Equal Employment Opportunity Commission guidelines.

If state legislatures actively start the process of entering into evaluating faculty or programs, can the Boards of Trustees of Private Colleges be far behind? If the State legislatures or Boards of Trustees, or both enter the field of faculty or program evaluations, what form will these evaluations take? If these are to be like other programs which law makers left to their own devices create, we educators are in trouble! I support the conclusions of Glenny, Schmidtlein, and Holley. I think that it is obvious that the intrusion, if not already begun, is inevitable. We have a choice. We can ignore the situation and let it develop by itself. We can fight it. Or, we can develop one or several models that will help not only the authorities evaluate us and our programs, but also assist us in our own quality assurance programs. I think that the first two options are inappropriate. We should, I believe, develop our own models. My research, albeit only a pilot study, allows me to suggest such a model.

Quality Assurance

Before beginning the discussion and description of the pilot study, a word has to be said about quality. The concept of quality has been around for a long time. It is also a misunderstood term because it means different things to different people. Frequently we mistake goodness for quality. They are different. Philip Crosby, founder of the Quality College in Orlando, states that quality has four absolutes. The first of these is that by definition, quality means conformance to requirements. A Volkswagon beetle is a quality product if it conforms to the specifications. A Lincoln Town Car is a quality product if it conforms to the requirements. In education, whether it be the course examination or our performance as educators, if the requirements are not specified we will never have a chance of having 'quality' programs. The second absolute is prevention. By adopting a course development model, we can prevent nonconformance to course requirements. Had the course development model been instituted at UCF in the Spring of 1985, they would not have had the problems they did in one of their chemistry courses. Error-free performance is the third absolute. This means that we do it right the first time. By documenting our mistakes we can ensure that we do not make them a second time. If we evaluate our tests we can also evaluate our instruction. Measurement is the last absolute. By evaluating our courses and instruction, we can then take corrective action and eliminate the nonconformances.

Discussion

This pilot study began as a project to investigate the use of statistics by college faculty. In many of our colleges of education, we are told that we must evaluate our tests and other measurement devices used to grade student performance. This is supposedly one of the reasons we take statistics courses in

graduate programs. The other reason, of course, is our theses and dissertations require a statistics package. I initially wanted to see whether the faculty at UCF followed what they preached. As I analyzed the data, however, I came to realize that what I was really studying was not use of statistics by college faculty. I was, in reality, studying the theory of course development.

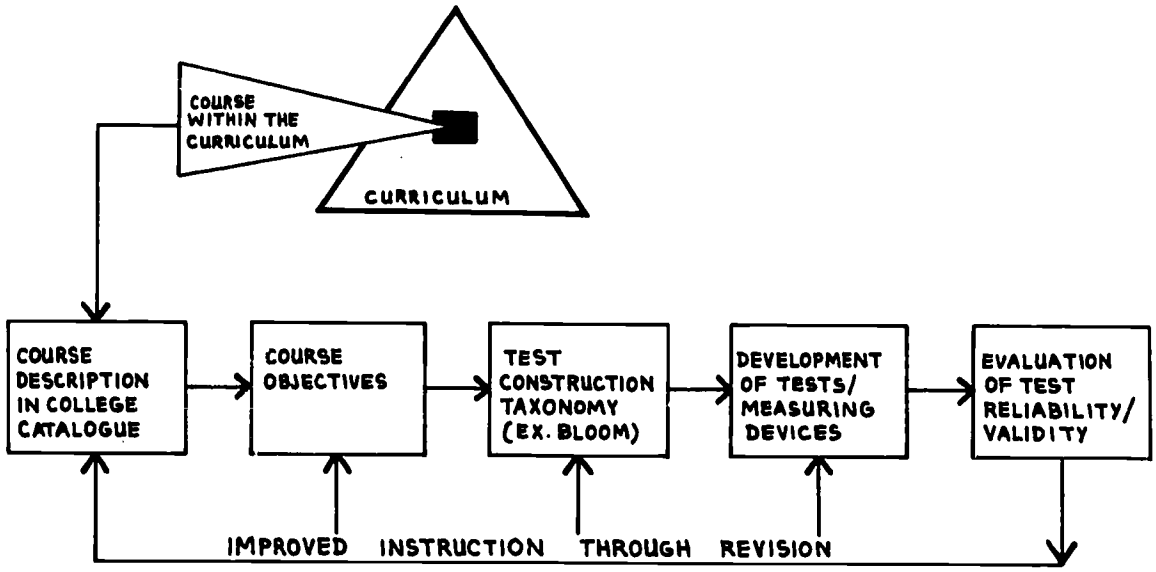
Jerome Bruner, in the Process of Education, discusses the "spiral curriculum" meaning that a curriculum must be designed so that a thought, idea, or concept transcends into more advanced thoughts, ideas, or concepts. If we look at college curricula, this "spiral curriculum" concept is vividly illustrated in the course sequencing within educational programs. Any curricula a student enters and then exits several years later includes basic skills and knowledge courses, prerequisite courses, and advanced courses. Each of the latter is dependent upon the former. We see this in the UCF College of Education's doctoral program, specifically the statistics requirement. We take four courses. The first is an advanced basic course, the second is a course on multi-variate analysis, the third includes log linear analysis and questionnaire design, and the fourth is a research design course. The outcome of the last course is a dissertation proposal and discussion of the statistical analysis to be used in one's dissertation. It would be very, very difficult to meet the objectives of the fourth class, much less do an adequate job on a dissertation, without the three previous courses. The "spiral curriculum" gives us a flow diagram.

Bruner says that all subject matter has a structure. He stresses the importance of student knowledge and awareness of this structure at the appropriate time. Barry J. Wadsworth, in Piaget's Theory of Cognitive Development, describes the concept of "heirarchy of prerequisite learning." This concept is similar to Bruner's concept of structure. Gagne, a behaviorist, and Piaget, a developmental cognivist, both agree that good learning is dependent upon

presenting the learner with materials based on this hierarchial order, or more simply, following a logical sequence of instruction within the structure of a subject.

This course development theory is patterned after the Instructional Systems Design model of analysis, design, development, implementation, and evaluation. The development of any program is outlined in the curriculum which identifies the courses perculiar to it. Course descriptions, as contained in university catalogues, are supposed to be used to derive course objectives. From course objectives, the professor develops his or her tests as well as the instruction. Tests, according to Bloom, should be written in both a logical and scientific fashion. His taxonomy was developed to assist educators in this process. Evaluation is an important element in the theory of course development. First, the tests must be valid and reliable. Second, they must be scientifically improved. Finally, the feedback from student performance should lead to revised and refined instruction.

COURSE DEVELOPMENT MODEL
(THE MODEL)



Do you remember when you were in college? Did you not look at the course descriptions to find out what the course was about, and what was being taught? What was your reaction when you learned that in some of those courses the content did not agree with the descriptions? Did any of your professors give examinations on materials not covered in class? How did they handle your questions about these inconsistancies? Did they say, "I expect college level students to read a lot and absorb all the materials pertaining to this very important subject?" Did you feel that you got the wrong end of the deal? Why should a student have to guess what is expected? Why should a student have to wonder about what the important and salient points of a subject are? Quality assurance can be facili tated by letting people know what is required and how we intend to equip them with the skills necessary to perform to expectation.

The design of this pilot study required that a survey be distributed to all four hundred and five teaching faculty at the University of Central Florida. One hundred and thirty-one completed surveys were returned. This represents a 32 percent return rate. This sample, although biased, represents one third of the total population, so is assumed to be fairly representative. The breakdown of returned surveys by colleges within the university is as follows: Arts and Sciences - 44%, Business Administration - 18%, College of Education - 21%, College of Engineering - 12%, and Health Sciences - 5%.

A frequency distribution was obtained and from this distribution the Mean, Standard Deviation, Standard Error, Skewness, Kurtosis, and other statistical data evaluated. This data is located in Table 1 for your review. Between group significance is at the .05 level in those cases in which it is significant.

One factor very important to this theory is the generation and use of teacher made tests as opposed to department made or standardized tests in the

	MEAN: ARTS & SCIENCES	MEAN: BUSINESS ADMINISTRATION	MEAN: EDUCATION	MEAN: ENGINEERING	MEAN: HEALTH	GROUP MEAN	STANDARD ERROR / DEVIATION	SKEWNESS	KURTOSIS	BETWEEN GROUP SIGNIFICANCE
TEACHER MADE TESTS	98.42	95.00	98.14	100.0	93.57	97.67	0.828 9.473	-5.034	27.58	NO
NUMBER OF TESTS CREATED/COURSE	4.228	4.250	3.407	3.875	6.142	4.122	0.187 2.141	1.901	6.090	NO
NUMBER OF TESTS GIVEN/COURSE	4.105	4.041	3.666	3.750	6.142	4.068	0.172 1.974	1.976	7.420	NO
HOW FREQUENTLY DO YOU CURVE	2.631	2.750	3.030	2.250	2.714	2.694	0.093 1.066	-0.328	-1.113	NO
COURSE DESCRIPTION TO COURSE OBJ	2.122	1.833	1.481	2.000	1.857	1.908	0.099 1.133	0.827	-0.842	NO
USE COURSE OBJ TO DEVELOP TESTS	1.754	2.041	1.370	2.062	1.285	1.740	0.084 0.957	1.131	0.219	YES
USE A TAXONOMY FOR TEST DEVELOP	3.789	3.625	2.074	3.750	2.428	3.328	0.089 1.019	-1.186	-0.055	YES
OBTAIN RELIABILITY FREQUENTLY	3.122	3.291	2.851	3.250	2.142	3.061	0.093 1.065	-0.782	-0.704	NO
USE RELIABILITY DATA	3.140	3.250	2.814	3.375	2.428	3.084	0.091 1.045	-0.744	-0.767	NO
AVERAGE RELIABILITY	3.140	3.541	2.814	3.687	2.714	3.190	0.107 1.229	-0.928	-0.867	NO
VALIDITY ONE	3.526	3.875	3.222	3.937	2.571	3.526	0.108 1.236	-0.745	-0.388	YES
VALIDITY TWO	0.754	0.416	1.074	0.500	0.857	0.732	0.124 1.419	1.617	1.006	NO
VALIDITY THREE	0.263	0.375	0.407	0.000	0.428	0.290	0.083 0.973	3.157	8.377	NO
CALCULATE STAT VALIDITY	3.543	3.458	2.888	3.437	2.285	3.313	0.079 0.904	-0.981	-0.276	YES
AVERAGE VALIDITY	4.596	4.625	3.481	4.375	3.285	4.274	0.121 1.387	-1.472	0.357	YES
DO ITEM ANALYSIS	3.403	3.041	2.333	3.312	2.285	3.045	0.095 1.087	-0.639	-1.047	YES
REVISE TEST BASED ON ITEM DISCRIM PWR	3.386	3.250	2.592	3.625	2.571	3.183	0.081 1.036	-0.755	-0.957	YES
OBTAIN STANDARD ERROR	3.596	3.583	3.222	3.312	2.285	3.412	0.081 0.927	-1.503	1.173	YES
OBTAIN STANDARD DEVIATION	3.280	3.166	2.925	3.000	2.142	3.091	0.101 1.160	-0.872	-0.814	NO

TABLE 1

student evaluation process. Ninety-one percent of the teaching faculty create 100 percent of the tests they use to evaluate student performance. The mean for this figure was 97.672. With a standard error of 0.828, I can project that the mean of the population will be between 96.05 and 99.294. Both the skewness (-5.034) and the kurtosis (27.583), show that the distribution is not normally distributed. Thus, I concluded that a significant portion of the UCF faculty do in fact create their own tests to evaluate student performance. I suspect that this conclusion would hold true at other institutions as well.

I also found that the average number of tests created per course to be 4.122. The average number given per course is 4.069. The median and mode for both is 4.00. The correlation between the two seems to be significant. A Pearson Correlation indicated the incidence of correlation to be strong (.72).

The mean for using course descriptions to develop course objectives is 1.908, or the faculty seems to use course descriptions fairly regularly to develop course objectives. There was no significance between colleges. Fifty-three percent of the respondents indicated that they always use course descriptions to develop course objectives, while twenty-three percent frequently use the descriptions to develop course objectives. This explains the standard deviation of 1.133. The standard error of 0.099 indicates that the population mean would be quite close to the sample (1.908+or-.19).

Members of the faculty do not use a taxonomy to help them develop their teacher made tests. Sixty-five percent of the faculty never use a taxonomy. The level of significance between colleges is less than .01, which means that there is virtually no chance that the results are due to sampling error.

The results of the analysis of the data also revealed that the faculty seldom, if ever, obtain the reliability ratings for their tests. This is

significant when one realizes that the majority of tests and test questions given by faculty lend themselves to statistical analysis. This is even more important because the university has a computer center on campus which will do a data analysis of examination results for the faculty members. One question raised is, do faculty members know that this service is available?

The result of the analysis was sufficient to warrant a more indepth look at the data. As a result, several cross tabulations were obtained. These first, were cross tabulations of using course discriptions to develop objectives, using objectives to develop tests, obtaining reliability information, and using reliability information by colleges within the university. Although there were the differences in the patterns between the colleges, the variations were not significant.

Two, three variable log linear analyses were performed. The two independent variables in each case were using course descriptions and developing objectives. The dependent variables were obtaining reliability on tests and using reliability to evaluate tests. Table 2 shows the results of the log linear analysis with the obtaining of reliability as the dependent variable. Model 5 is the model which shows that there is a strong interaction between using course descriptions to develop objectives and using course objectives to develop tests. This means that there is also an interaction between using course descriptions and obtaining reliability. Table 3 shows the results of the log linear analysis while using reliability as the dependent variable. The results between the two log linear programs are similar. Unfortunately, faculty members who use course descriptions to develop course objectives and use course objectives to develop their tests do not obtain data on the reliability of their testing devices.

TABLE TWO

Chi Square	df		Component Analysis	df
134.57344	62			
103.95084	60	Desc.	30.62260	2
65.45541	57	Obj.	38.49543	3
49.37898	54	OReli.	16.07643	3
15.86690	45	D on Obj.	33.51208	9
8.16124	36	D on OR	7.70566	9
7.50723	23	Obj. on OR	0.065401	9
0.0	22	D on Ob on OR	7.50723	5

TABLE THREE

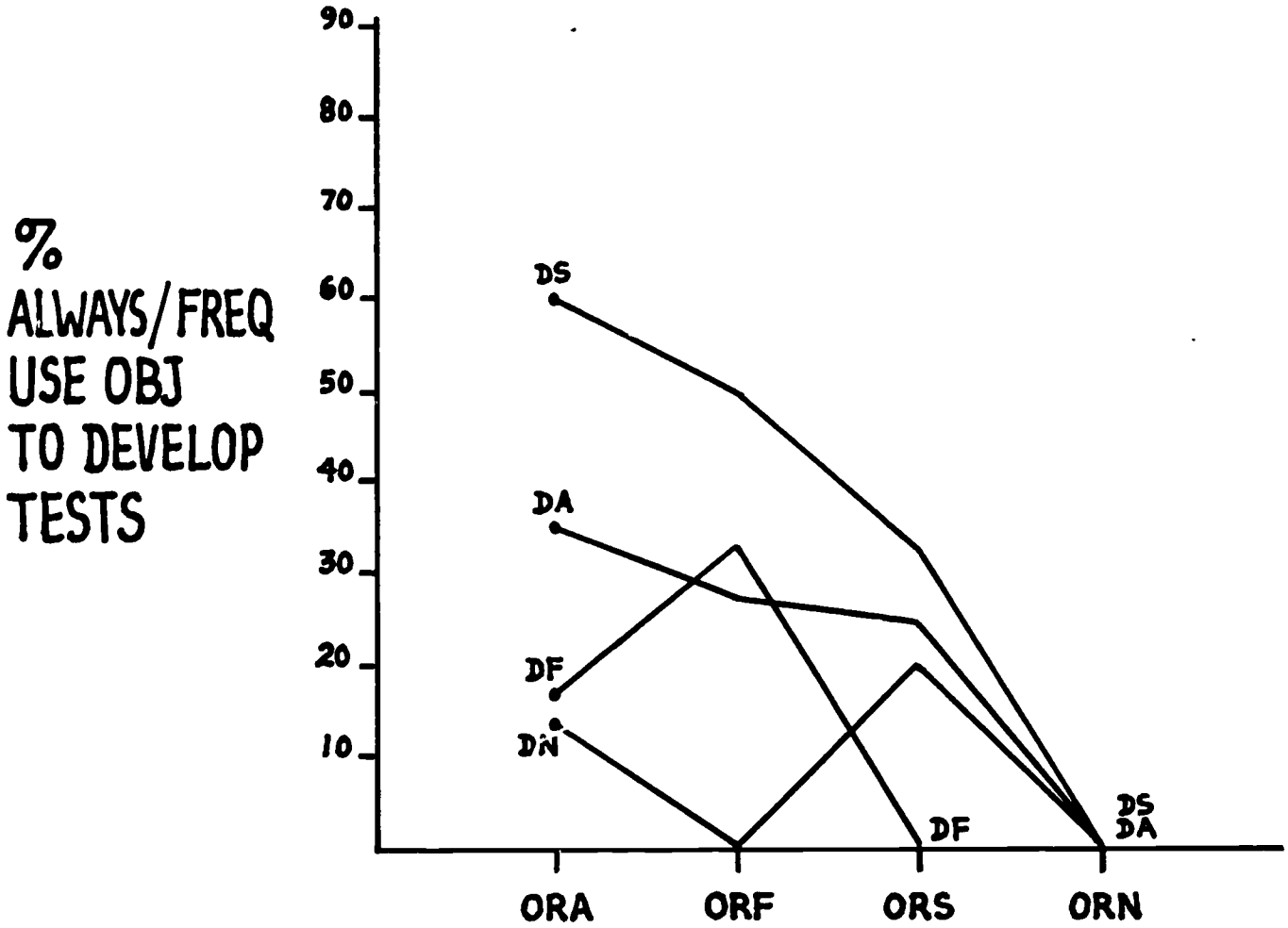
Chi Square	df		Component Analysis	df
126.38297	62			
92.30995	60	Desc.	34.07302	2
61.30515	57	Obj.	31.0048	3
45.04537	54	UReli.	16.25978	3
14.44798	45	D on Obj.	30.59739	9
8.30819	36	D on UR	6.13979	9
6.93657	27	Obj. on UR	1.37162	9
0.0	23	D on Ob on UR	6.93657	4

Figure 1 graphically depicts this fact. The impact of the two independent variables appears to be negative as opposed to being positive.

The desired relationship we want to see is that the use of course descriptions to develop course objectives is positive and strong, which is the case. We also want to see that this relationship is equally positive and strong when it comes to obtaining and using reliability data to improve teacher made tests. Unfortunately, this has not been the case. The relationship is strong, but negative.

What does this all mean to us as educators? At the University of Central Florida, the faculty already has the makings of a quality assurance program in place. It is one which is practiced and used by a large element of the faulty. We only have to strenghten it. I believe that the faculty at other institutions of higher learning are similar to those I have found at UCF. We can influence the adoption of a quality assurance program which is already acceptable. We only have to implement the rest of our model.

We can now educate our faculty on how to close the quality assurance loop. If we can instruct them on how to obtain and use reliability data, they can then use this information to evaluate their tests and other measuring devices. Once this is accomplished, they will be in a position to identify the areas of their instruction which need to be improved, or perhaps to change the course descriptions, or to even refine their course objectives. Their position can be further strengthened by using a taxonomy, such as Bloom's, to assist them in writing their tests and examinations.



LEGEND

DA - Always uses Course Descriptions
DF - Frequently uses Course Descriptions
DS - Seldom uses Course Descriptions
DN - Never uses Course Descriptions
ORA - Always Obtains Reliability
ORF - Frequently Obtains Reliability
ORS - Seldom Obtains Reliability
ORN - Nover Obtains Reliability

FIGURE 1

Conclusion

In this paper, we presented a course development model which holds that a faculty member, in order to develop instruction which harmonizes with the college curriculum, should have a good course description and from that course description should develop realistic and attainable course objectives. The course objectives should then be used to develop classroom and exit examinations, as well as course content. A taxonomy of testing, such as Bloom's, should then be used as a guide to construct the tests and examinations. Reliability data derived from test evaluation should be used to improve both instruction and the measuring device itself.

If this model, which is currently being used to some extent by the faculty at the UCF, were to be adopted and used it would fit all the requisites of a quality assurance program. The course and accountability requirements would be specified. There would be a preventative system or mechanism in place to ensure quality. We would be able to document all the steps in developing our courses of instruction and improving our educational programs. Finally we would have a means of accurately measuring our deviations or nonconformances.

What would the impact be with regards to faculty accountability and evaluation? We would have a way of showing legislatures, boards of trustees, students, and anyone else how we logically met the objectives of the "spiral curriculum". Second, use of this model would provide one objective and measurable input to the complex process of faculty evaluation. Lawrence Poole says that researchers agree that no one method of faculty evaluation is sufficient. Use of this model would provide one objective and measurable input to this complex process. This model represents a quantifiable behavioral input which the professor can control. It thereby reduces the impact of subjective evaluations in the evaluation program.

REFERENCES

Andrews, Dee H. and Ludwika A. Goodson. "A Comparative Analysis of Models of Instructional Design," Journal of Instructional Development, 3, No. 4, (1980), pp. 2-16.

Braskamp, Larry A. and James D. Rath. "On Evaluating Professionals in Higher Education," Evaluation News, 3, No. 2, (1982), pp. 85-90.

Dawson, Judith A. and Darrel N. Caulley. "The Group Interview as an Evaluation Technique in Higher Education," Educational Evaluation and Policy Analysis, 3, No. 4, (1981), pp. 61-66.

Dierderich, Paul B. Short-Cut Statistics for Teacher-Made Tests. (Princeton: Educational Testing Service, 1973).

Driscoll, Marcy P. "Alternative Paradigms for Research in Instructional Systems," Journal of Instructional Development, 7, No. 4, (1984), pp. 2-5.

Everett, Michael D. "The Impact of Student Evaluations of Teaching on the Quality of Education," Liberal Education, Winter (1981), pp. 327-335.

Gilley, Jerry W. and Robert L. Dean. "Instructional Design Utilizing Content and Experience: A Model," Performance and Instructional Journal, 24, No. 10, (1985), pp. 20-24.

Glenny, Lyman A. and Frank A. Schmidtlein. "The Role of the State in the Governance of Higher Education," Educational Evaluation and Policy Analysis, 5, No. 2, (1983), pp. 113-153.

Holley, William H. Jr., Richard Cottrelt, and R. James Kessler. "Assessment of Faculty Performance Appraisal," The Journal of the College and University Personnel Association, 33, No. 1, (1982), pp. 14-18.

Poole, Lawrence H. and Donald A. Dellow. "Evaluation of Full-time Faculty," New Directions for Community Colleges, No. 41, (1983), pp. 19-32.

Rubeck, Robert F. and Donald Gill. "Incident Evaluation a New Technique for Student Evaluations of Their Instructors," Educational Technology, 23, No. 9, (1983), pp. 43-45.

Schrock, Timothy J. "A Way to Empirically Analyze Kompetence: A Model for Faculty Evaluation," Evaluation News, 4, No. 2, (1983), pp. 82-85.

Stahl, William K. "Academic Program Review in a Community College," Evaluation News, 4, No. 2, (1983), pp. 61-64.

TRADOC Pamplet 350-30, Interservice Procedures for Instructional Systems Development, Washington: Department of the Army, (1975).