DOCUMENT RESUME

ED 088 944 TM 003 506

TITLE	Proceedings of the Annual State Conference on Educational Research (24th Burlingame, California, November, 1972)--"Evaluation--Try It, You'll Like It!"
INSTITUTION	California Teachers Association, Burlingame. Dept. of Research.
SPONS AGENCY	California Advisory Council on Educational Research, Burlingame.
REPORT NO	RP-48
PUB DATE	Nov 72
NOTE	113p.
AVAILABLE FROM	California Teachers Association, 1705 Murchison Drive, Burlingame, California 94010 ($2.00)
EDRS PRICE	MF-$0.75 HC-$5.40
DESCRIPTORS	Administrator Evaluation; *Conference Reports; *Cost Effectiveness; Educational Accountability; Educational Assessment; *Evaluation; Instructional Improvement; *Public Schools; Student Evaluation; Teacher Evaluation

ABSTRACT

The meeting was organized around 10 symposia concerning the following areas: teacher evaluation through measurement of student achievement, statewide evaluation, district evaluation procedures, individualized evaluation standards, learning improvement, competency based curriculum, administrative evaluation, student evaluation of the system, and cost effectiveness. Papers presented at each symposium are included in this publication. (RC)

CALIFORNIA ADVISORY COUNCIL ON EDUCATIONAL RESEARCH

RESEARCH RÉSUMÉ

NUMBER 48
NOVEMBER 1973

Proceedings of the 24th Annual State Conference on Educational Research

"EVALUATION – TRY IT, YOU'LL LIKE IT!"

CALIFORNIA TEACHERS ASSOCIATION 1705 Murchison Drive Burlingame

$ 2.00

CALIFORNIA TEACHERS ASSOCIATION RESEARCH STAFF

GARFORD G. GORDON
Research Executive

HAL R. WEATHERBE
Assistant Research Executive

DONALD P. GLASER
Assistant Research Executive

LOUISE HINDMAN
Research Assistant

DORMILEE TASSOS
Research Assistant

SHIRLEY SCHMITT
Research Assistant

MARY EVELYN PYBURN
Research Assistant

PAMELA M. SMITH
Research Assistant

JOSEPHINE BROCK
Secretary

WILMA NICHOLS
Secretary

ANNE PROTOPOPOFF
Librarian

LOIS WING
Assistant Librarian

MARY FULLER
Computer Consultant

PROCEEDINGS OF THE 24th ANNUAL STATE CONFERENCE ON EDUCATIONAL RESEARCH

"EVALUATION--TRY IT, YOU'LL LIKE IT!"

HYATT HOUSE HOTEL

SAN JOSE

Conducted by

CALIFORNIA TEACHERS ASSOCIATION

for

CALIFORNIA ADVISORY COUNCIL

ON

EDUCATIONAL RESEARCH

NOVEMBER 9-10, 1972

INTRODUCTION

CONFERENCE PROCEEDINGS HAVE BEEN AN ANNUAL FEATURE SINCE 1958. PRIOR TO THAT DATE, AN ANNUAL REPORT WAS GIVEN ABOUT THE CONFERENCE AFTER ITS CONCLUSION.

EACH PARTICIPANT THIS YEAR SUBMITTED HIS REMARKS ON SPECIALLY DESIGNED PAPER. EACH PAGE WAS PHOTOGRAPHED IN ORDER TO PREPARE A PLATE SO THAT THE PROCEEDINGS COULD BE PRINTED AS SUBMITTED BY THE AUTHORS, WITHOUT EDITING BY THE CTA STAFF.

CTA ASSISTANT RESEARCH EXECUTIVE, DONALD P. GLASER, IS CONFERENCE PROGRAM MANAGER AND JOSEPHINE BROCK IS CONFERENCE PROGRAM SECRETARY.

TABLE OF CONTENTS

FIRST GENERAL SESSION

SPEAKER: LEE J. CRONBACH
VIDA JACKS PROFESSOR OF EDUCATION
STANFORD UNIVERSITY

TOPIC: "WHAT CAN GO WRONG IN EVALUATION?"

WHAT CAN GO WRONG IN EVALUATION?

Lee J. Cronbach

(Lee J. Cronbach is Vida Jacks Professor of Education at Stanford University. He received his B.A. degree from Fresno State College, his M.A. from the University of California at Berkeley and his Ph.D. from the University of Chicago.)

The term evaluation is now much used, and evaluations are in great demand. The term is ambiguous, because there are many kinds of evaluation. Too often, there are misunderstandings because legislators, educators, and the public expect one kind of inquiry to serve all purposes. Professionals misapply principles of evaluation because they do not realize that the validity of the principle is limited to certain kinds of investigation.

Evaluation is the collection and analysis of data that will help a decision maker. Too often it is considered to be an entirely scientific inquiry, but it is a social process, and a means of educating the decision makers. Evaluation is not passive, neutral, or value-free. The evaluator has responsibility to improve the program he evaluates, not merely to turn in a report card.

The data from which evaluations are made are often too limited. This is especially the case when evaluations are dominated by objective tests and by narrowly conceived "behavioral objectives". The essence of quality in an educational program is its ability to improve performance on transfer tests, i.e., on matters not studied. Evaluations that range beyond the prescribed curricular content can threaten teachers and students, but they are indispensible. Attention needs to be given to those outcomes that arise out of the whole educational process, not being the domain of any one course or subject field.

The analysis of evaluative data is often too simple. There are limits to the depth of an analysis a target audience can profit from, but one should run the risk of analyzing more deeply rather than less deeply.

Good evaluation is costly. The person planning an evaluation must make hard choices to decide what information will have the greatest effect on program quality. He must prepare a report that will be meaningful to the decision maker, and educationally valid. No standard rules can tell him how to accomplish this in any one setting.

SECOND GENERAL SESSION

SPEAKER: HARRY F. SILBERMAN
TASK FORCE LEADER
NATIONAL INSTITUTE OF EDUCATION

TOPIC: "PROGRAM PLANNING FOR NIE"

Dr. Harry F. Silberman

PROGRAM PLANNING FOR NIE

The NIE and the Office of Education are now co-equal branches of the new Education Division in HEW created by the Education Amendments of 1972.

The NIE legislation specified that the Institute shall seek to improve education, including career education, in the United States. To achieve its purpose, NIE must be able to identify the goals of American education, but in a pluralistic society it is difficult to specify a set of goals with which everyone will agree. Nevertheless, some attempts have been made in our studies to describe a set of goals on which there is some agreement and we considered the problems most frequently cited as obstacles to achieving those goals. The problems most frequently cited were: (1) Access; (2) Participation; (3) Productivity; (4) Substance.

The program suggestions that are most often proposed to overcome these obstacles require that certain factors be changed; (1) Laws; (2) Incentives; (3) Information; (4) Agents.

NIE R&D programs call for substantial experimentation with these four variables. Program plans reflect both new initiatives and NIE's responsibility to assume management of certain on-going research and development activities previously supported by the Office of Education.

Some of the on-going programs that were transferred from the Office of Education include: The National Center for Educational Communications, the Experimental Schools Programs, the Career Education Program, the Regional Laboratory and R&D Centers Program, the Researcher Training Program, and the Unsolicited Research Program. These programs comprise approximately 80% of the National Institute of Education's budget.

The program areas in which planning for new initiatives is proceeding are as follows: 1. Home-based Education; 2. School-Work Transition; 3. Community Participation; 4. Post Secondary Alternatives; 5. School Finance; 6. New Measures for Education; 7. Self Directed Education; 8. Productivity.

An important share of NIE's attention should be devoted to broad policy issues related to education. Much of this work may be done intramurally as part of a "think tank" operation in which senior staff on temporary leave from their home institutions spend their time at NIE in an attempt to synthesize what is known about crucial issues and make recommendations to governmental officials on matters pertaining to education.

There are several important outcomes from such systematization and analysis of R&D findings. Without an explicit effort to organize what we have learned from past R&D effort, we shall continue to travel the

same cul-de-sacs. An institutionalized memory of past failures will conserve NIE resources so that new program plans will be more likely to have a cumulative effect. Another possible outcome of a systematic NIE attempt to synthesize R&D is more realistic advice for legislators and executives at Federal, State, and local levels, who must make decisions on educational policy. Where solutions to problems are not available, where knowledge gaps prevent analysis and existing R&D activities are not already underway, alternative program plans will be prepared for new NIE research initiatives. Thus, NIE might support directed studies on such topics as school finance, organization reform, or accreditation and certification practices. Longitudinal studies of entire communities might also be undertaken to assess the relative influence of major determinants of educational achievement, such as family, mass media, peer culture, employer practices, schooling, etc.

THIRD GENERAL SESSION

SPEAKER: ASSEMBLYMAN JOHN STULL
80th ASSEMBLY DISTRICT
STATE OF CALIFORNIA

TOPIC: "TEACHER COMPETENCY ACT, AB 293"

TEACHER COMPETENCY ACT, AB 293

ASSEMBLYMAN JOHN STULL
80th ASSEMBLY DISTRICT
STATE OF CALIFORNIA

LADIES AND GENTLEMEN:

I APPRECIATE HAVING THIS OPPORTUNITY TO SHARE SOME THOUGHTS WITH YOU ON MY BILL ON CALIFORNIA TEACHER TENURE--A MATTER WHICH SHOULD BE OF VITAL CONCERN TO ALL OF US, AS SCHOOL ADMINISTRATORS, AS EDUCATORS, AS PARENTS, AS TAXPAYERS, AND AS CITIZENS OF THIS STATE.

OUR WHOLE SITUATION REMINDS ME OF THE OLD STORY ABOUT THE DOCTOR, THE ENGINEER, AND THE POLITICIAN WHO WERE ARGUING ABOUT WHICH OF THEIR PROFESSIONS WAS THE OLDEST. THE DOCTOR SAID THAT, OF COURSE, MEDICINE WAS THE OLDEST, THAT MANKIND HAD ALWAYS HAD PHYSICIANS, AND THAT THEY ARE EVEN MENTIONED IN THE BIBLE.

"THAT'S NOTHING," SAID THE ENGINEER, "THE BIBLE TELLS US HOW THE WORLD WAS CREATED OUT OF CHAOS, AND HOW COULD THERE BE ANY ORDER BROUGHT OUT OF CHAOS WITHOUT AN ENGINEER TO HELP?"

WHEREUPON THE POLITICIAN BROKE IN. "WAIT A MINUTE," HE SAID, "WHO DO YOU THINK CREATED THE CHAOS?"

WELL, I'LL ADMIT TO BEING A POLITICIAN MYSELF, BUT I DON'T LIKE THE CURRENT CHAOS ANY BETTER THAN THE REST OF YOU. THAT'S ONE OF THE REASONS THAT PASSAGE OF AB 293 PROVIDES MUCH SATISFACTION--IT'S A BRIGHT NOTE.

I HAVE SERVED IN THE LEGISLATURE NOW FOR OVER SIX YEARS, AND I HAVE BEEN REQUESTED TIME AFTER TIME TO PROVIDE MORE FUNDS FOR SCHOOLS. I HAVE BEEN ASKED TO ALTER THE CREDENTIALING LAWS, LOWER CLASS SIZE, INCREASE THE NUMBER OF NURSES, PSYCHOLOGISTS,

AND COUNSELORS, RAISE TEACHER SALARIES AND EVEN TO ELIMINATE THE PITHING OF LIVE ANIMALS. NOW FOR EACH OF THESE REQUESTS THERE HAS BEEN A REASONED ARGUMENT DESIGNED TO DEMONSTRATE THE ESSENTIALITY OF THE CAUSE. BUT I WANT YOU TO KNOW THAT IN NO CASE WAS EVIDENCE PRESENTED WHICH WOULD CLEARLY DEMONSTRATE THAT STUDENTS WOULD LEARN MORE WITH ANY OF THESE CHANGES.

I HAVE FROM TIME TO TIME SUPPORTED THE REQUESTS FOR INCREASES IN EACH OF THE ABOVE SERVICES. AND I HAVE FROM TIME TO TIME OPPOSED SUCH REQUESTS. BUT WITH EACH PASSING YEAR I BEGAN MORE AND MORE TO QUESTION THE BASIS FOR EACH REQUEST.

WHAT I ALWAYS WANTED TO KNOW WAS WHAT WILL BE THE BENEFIT TO CHILDREN FROM SUCH ALLOCATIONS OF DOLLARS. WHAT I CARE ABOUT IS NOT SO MUCH HOW SERVICES ARE DELIVERED BUT RATHER HOW WELL ARE THEY RECEIVED. I WANT THE GOVERNMENTAL PROCESS TO CONCENTRATE ON THE OUTPUT RATHER THAN ON THE INPUT.

WHAT AB 293 ATTEMPTS TO DO IS TO FOCUS THE ATTENTION OF THE SCHOOL COMMUNITY ON THE PROGRESS OF CHILDREN. I DO NOT BELIEVE THAT IT SHOULD BE THE FUNCTION OF THE LEGISLATURE TO TELL PROFESSIONALS HOW TO PROVIDE SERVICES, BUT PROFESSIONAL FLEXIBILITY--OFTEN CALLED ACADEMIC FREEDOM--REQUIRES AS ITS COUNTERWEIGHT PROFESSIONAL ACCOUNTABILITY BASED UPON THE PRODUCT OF THE PROFESSIONAL SERVICE. SO AB 293 ATTEMPTS TO CONCENTRATE ATTENTION UPON THE PRODUCT OF THE EDUCATIONAL SYSTEM, THROUGH EVALUATION.

TO ME AB 293 WAS A FIRST ATTEMPT TO CREATE AN EVALUATION SYSTEM COUPLED WITH A REVISED DUE-PROCESS SYSTEM WHICH WILL AFFIRM OR DENY THE EFFICACY OF THE EVALUATION SYSTEM. THAT IS

REALLY A COMPLEX SENTENCE, BUT IN TRUTH THIS IS THE ESSENCE OF AB 293.

TO SPEAK FAVORABLY OF TEACHER EVALUATION IS TO VENTURE INTO HAZARDOUS TERRITORY. EVALUATION OF SCHOOL PERSONNEL HAS NEVER BEEN POPULAR. THE EDUCATION PROFESSION HAS RESISTED, AVOIDED, RATIONALIZED, OR WATERED DOWN NEARLY ALL EFFORTS TO CONFRONT THE ISSUE OF EVALUATION ON OPEN ANALYTICAL GROUNDS.

THE WISDOM WITH WHICH ANY EVALUATION SYSTEM IS IMPLEMENTED WILL HAVE MUCH TO DO WITH NOT ONLY THE PROMISE OF HIGHER-QUALITY EDUCATION FOR OUR CHILDREN, BUT, INDEED, THE VIEWS OF SOCIETY TOWARD EDUCATION IN GENERAL. THE OVERALL PURPOSE OF EVALUATION IN AB 293 IS INSTRUCTIONAL IMPROVEMENT. IT IS TO HELP EACH TEACHER AND EACH ADMINISTRATOR SUCCEED IN HIS OR HER MISSION TO OPEN OUR WORLD TO CHILDREN.

ADMITTEDLY, THE IMPLEMENTATION OF THE EVALUATION CALLED FOR IN AB 293 WILL TAKE TIME, RESOURCES, AND TALENT--VERY LIKELY MORE THAN IS AVAILABLE TODAY--BUT IT IS IMPERATIVE THAT WE MOVE NOW TO BEGIN.

NOW, I AM AWARE THAT THE EVALUATION PROVISIONS OF AB 293 HAVE CAUSED A GREAT DEAL OF CONSTERNATION WITHIN THE EDUCATIONAL PROFESSION. I HAVE BEEN TOLD THAT AB 293 IS RIGID, BUREAUCRATIC, AND EXTREMELY UNREALISTIC. I HAVE BEEN TOLD THAT AB 293 REQUIRES THE USE OF STANDARDIZED TESTS TO MEASURE PUPILS' PROGRESS. I HAVE BEEN CASTIGATED BY THOSE WHO BELIEVE THAT AB 293 IS SUBTERFUGE FOR THE IMPLEMENTATION OF PPBS, AND I HAVE BEEN ACCUSED OF BEING PUNITIVE TOWARDS TEACHERS, IGNORANT OF EDUCATIONAL PROCESSES, AND ANTI PUBLIC SCHOOLS.

WELL, NOBODY IS PERFECT.

WHAT I BELIEVE AB 293 WAS DESIGNED TO DO IS TO SET ABOUT A PROCESS WHICH WILL REQUIRE A CLOSER EXAMINATION OF THE PROGRESS OF STUDENTS AND THE FUNCTIONS OF THOSE PAID TO PROMOTE SUCH PROGRESS IN THE PUBLIC SCHOOLS.

INITIAL LEGISLATIVE STUDIES ON THE SUBJECT OF TENURE BEGAN IN 1968 WITH THE SENATE EDUCATION COMMITTEE REVIEWING CALIFORNIA'S TENURE STRUCTURE DURING INTERIM STUDIES. IN 1969, THE ASSEMBLY SUBCOMMITTEE ON EDUCATIONAL ENVIRONMENT, WHICH I CHAIRED, REVIEWED TENURE IN ELEMENTARY, SECONDARY, AND HIGHER EDUCATION. IN 1970 AND 1971 LEGISLATION RESULTING IN AB 293 WAS REVIEWED.

IN BOTH THE ASSEMBLY AND SENATE STUDIES OF EXISTING TENURE LAWS, THE EVALUATION SYSTEM WAS OF KEY CONCERN. EVEN IN DISCUSSING DISMISSAL PROCEDURES, ONE CONCEPTUAL QUESTION WHICH ALWAYS AROSE WAS, "HOW DO YOU ASSESS COMPETENCY?" IN AB 293 I FEEL THE LEGISLATURE HAS PROVIDED YOU, THE EDUCATIONAL COMMUNITY, WITH THE GUIDELINES NECESSARY FOR A REASONABLE ASSESSMENT OF COMPETENCY. I WOULD, THEREFORE, NOW LIKE TO REVIEW WITH YOU THE EVALUATION SECTION OF THE NEW LAW.

AMONG OTHER THINGS, THE LAW REQUIRES THAT EACH SCHOOL BOARD ADOPT A UNIFORM SET OF EVALUATION AND ASSESSMENT GUIDELINES FOR USE IN EVALUATING THE PROFESSIONAL COMPETENCY OF ITS CERTIFICATED PERSONNEL, AND FURTHER REQUIRES THAT THESE GUIDELINES BE PUT IN WRITING. IN OTHER WORDS, SCHOOLS BOARDS ARE, UNDER THE NEW LAW, REQUIRED TO PUT DOWN ON PAPER AND MAKE AVAILABLE TO THEIR CERTIFICATED PERSONNEL GUIDELINES OF EXPECTED JOB PERFORMANCE.

GUIDELINES DESIGNED TO ASSIST THE CERTIFICATED PERSON IN KNOWING WHAT IS EXPECTED OF HIM AND HOW THE SCHOOL BOARD WILL GO ABOUT REVIEWING HIS PERFORMANCE.

IN THE DEVELOPMENT OF THESE EVALUATION AND ASSESSMENT GUIDELINES AND PROCEDURES, THE NEW LAW REQUIRES THAT SCHOOL BOARDS AVAIL THEMSELVES OF THE ADVICE OF CERTIFICATED INSTRUCTIONAL PERSONNEL. IT WAS MY INTENTION FROM THE OUTSET THAT CERTIFICATED PERSONNEL BE REVIEWED BY THEIR PEERS. TO ENSURE THAT THE CRITERIA ESTABLISHED ARE APPROPRIATE TO THE FIELD IN WHICH THE EMPLOYEE IS CERTIFICATED, I FELT THAT INPUT FROM THE CERTIFICATED PERSONNEL THEMSELVES IN THE FORMULATIVE STAGES WAS ESSENTIAL. IN WRITING THIS PROVISION INTO THE LAW, I ENVISIONED THIS INVOLVEMENT BEING ACCOMPLISHED EITHER THROUGH A TEACHERS' COMMITTEE, APPOINTED BY THE SUPERINTENDENT, OR WITH THE ASSISTANCE OF THE CERTIFICATED EMPLOYEE COUNCIL. NO MATTER WHICH METHOD IS IN FACT UTILIZED, CERTIFICATED PERSONNEL INPUT AT THE FORMULATIVE LEVEL IS ESSENTIAL TO ENSURE A QUALITY PRODUCT, AND I AM HOPEFUL THAT YOU WILL KEEP THIS PORTION OF THE NEW LAW IN MIND AS YOU PROCEED WITH ITS IMPLEMENTATION.

IN FORMULATING EVALUATION AND ASSESSMENT GUIDELINES, THE NEW LAW CLEARLY ESTABLISHES FOUR AREAS THAT MUST BE CONSIDERED:

1. THE ESTABLISHMENT OF STANDARDS OF EXPECTED STUDENT PROGRESS IN EACH AREA OF STUDY AND TECHNIQUES FOR THE ASSESSMENT OF THAT PROGRESS;

2. ASSESSMENT OF CERTIFICATED PERSONNEL COMPETENCE AS IT RELATES TO THE ESTABLISHED STANDARDS;

3. ASSESSMENT OF OTHER DUTIES NORMALLY REQUIRED TO BE PERFORMED BY CERTIFICATED EMPLOYEES AS AN ADJUNCT TO THEIR REGULAR DUTIES AND ASSIGNMENTS; AND,

4. THE ESTABLISHMENT OF PROCEDURES AND TECHNIQUES FOR ASCERTAINING THAT THE CERTIFICATED EMPLOYEE IS MAINTAINING PROPER CONTROL AND IS PRESERVING A SUITABLE LEARNING ENVIRONMENT.

NOW IN LEGAL TERMS, THESE FOUR POINTS ARE, I GUESS, EASILY UNDERSTOOD. BUT IN LAYMAN TERMS, SUCH AS YOU AND I CAN UNDERSTAND, WHAT IS THIS PORTION OF THE NEW LAW SAYING?

BASICALLY, THE FIRST TWO POINTS REQUIRE YOU TO ESTABLISH AN EXPECTED LEVEL OF STUDENT PROGRESS OVER A GIVEN PERIOD OF TIME AND THEN ADOPT A SYSTEM OF EVALUATION TO SEE IF PROGRESS HAS BEEN ACHIEVED. YOU ARE THEN ASKED TO EVALUATE THE COMPETENCY OF THE CERTIFICATED EMPLOYEE ON THE BASIS OF HOW WELL HIS STUDENTS PROGRESSED. IN OTHER WORDS, WHAT DO YOU EXPECT THE STUDENT TO LEARN, AND HAVE EMPLOYEES DONE THEIR BEST TO BRING THE STUDENT ALONG SATISFACTORILY TO MEET THAT EXPECTATION?

THE REMAINING TWO POINTS INVOLVE JUDGING A CERTIFICATED EMPLOYEE'S PERFORMANCE IN PROFESSIONAL WORK RELATED TO HIS PRIMARY NORMAL ASSIGNMENTS, AND HIS ABILITY TO MAINTAIN DISCIPLINE AND GOOD ORDER AMONG HIS STUDENTS.

IT HAS COME TO MY ATTENTION THAT MANY ADMINISTRATORS HAVE INTERPRETED THIS FOURTH POINT--PERTAINING TO THE PRESERVATION OF PROPER CONTROL AND A SUITABLE LEARNING ENVIRONMENT--IN SUCH A WAY THAT IT APPLIES SOLELY TO INSTRUCTORS. THIS IS CLEARLY NOT THE INTENT OF THIS POINT, NOR IS IT THE INTENT OF THE LAW.

THIS LAW APPLIES TO ALL CERTIFICATED EMPLOYEES, FROM SUPERINTENDENT ON DOWN. EVERY CERTIFICATED EMPLOYEE OF A SCHOOL DISTRICT IS RESPONSIBLE IN HIS OWN AREA TO ENSURE THAT EVERY STUDENT IS AFFORDED THE BEST POSSIBLE EDUCATION BY THE BEST QUALIFIED CERTIFICATED PERSONNEL. THIS WAS MY INTENT, AND I BELIEVE IT IS CLEARLY THE INTENT OF THE LAW.

ALTHOUGH THESE FOUR POINTS CONSTITUTE IN THE MINDS OF MANY EDUCATORS THE HEART OF THE LAW, OF EQUAL IMPORTANCE TO ME IS THE REQUIREMENT THAT ANY GUIDELINES WHICH ARE ESTABLISHED MUST INCLUDE ADEQUATE PROVISIONS FOR FOLLOW-UP COUNSELING. MORE SPECIFICALLY, THIS NEW LAW REQUIRES YOU TO TAKE FOUR SPECIFIC STEPS BEFORE ANY DISMISSAL ACTION CAN BE TAKEN AGAINST A CERTIFICATED EMPLOYEE CONSIDERED DEFICIENT IN HIS EVALUATION RATINGS:

1. THE CERTIFICATED EMPLOYEE MUST BE NOTIFIED IN WRITING OF THE UNSATISFACTORY PERFORMANCE;

2. THE WRITTEN NOTICE OF UNSATISFACTORY PERFORMANCE MUST BE CLEARLY AND COMPREHENSIVELY STATED;

3. THE EVALUATOR MUST MEET AND DISCUSS WITH THE EMPLOYEE AND MAKE SPECIFIC RECOMMENDATIONS ON IMPROVEMENT OF HIS PERFORMANCE; AND,

4. POSITIVE STEPS MUST BE TAKEN TO ASSIST THE EMPLOYEE IN UPGRADING HIS PERFORMANCE.

ALL TOO OFTEN PROFESSIONAL PERSONNEL ARE NOT GIVEN THE OPPORTUNITY TO HAVE THEIR WORK REGULARLY EVALUATED AND, IN TURN, GIVEN THE OPPORTUNITY TO IMPROVE IN AREAS IN WHICH THEY MAY BE DEFICIENT. THIS SECTION OF THE NEW LAW NOT ONLY PROVIDES THIS OPPORTUNITY--IT MANDATES IT.

IN THIS CONTEXT, THEN, I DO NOT FIND THE EVALUATION PROVISIONS OF AB 293 ANYTHING MORE THAN A CODIFICATION OF WHAT SOUND EDUCATIONAL PRACTICE HAS BEEN FOR MANY YEARS.

I AM FULLY AWARE THAT THE IMPLEMENTATION OF THESE POLICIES HAS REQUIRED AND WILL CONTINUE TO REQUIRE A GREAT AMOUNT OF DILIGENT WORK.

IT MAY BE THAT CARRYING OUT THESE EVALUATION PROCEDURES WILL REQUIRE CHANGING THE ROLE OF MANY EXISTING EDUCATIONAL PROFESSIONALS.

IT MAY BE THAT THE ESTABLISHMENT OF EDUCATIONAL STANDARDS WILL CAUSE SOME SERIOUS RETHINKING OF EDUCATIONAL OBJECTIVES.

IT MAY BE THAT CARRYING OUT THE EVALUATION PROCESS WILL REQUIRE REARRANGEMENT OF RESOURCE ALLOCATIONS.

IF SUCH IS THE CASE, I WILL BE PLEASED; FOR THESE CONSIDERATIONS ALONE WILL STRONGLY INDICATE THAT WE ARE REFOCUSING OUR ATTENTION ON THE OUTCOMES OF OUR ACTIONS.

I HAVE BEEN TOLD SO OFTEN BY EDUCATORS THAT IT IS IMPOSSIBLE TO EFFECTIVELY EVALUATE PROFESSIONAL PERSONNEL. YET, EVERY TIME I GO TO A SCHOOL I CAN RECEIVE, PRIVATELY, A LIST OF THE FIVE BEST AND FIVE WORST PROFESSIONALS IN THE SCHOOL. AND, I MIGHT ADD, THE LIST IS PRETTY MUCH THE SAME WHETHER IT COMES FROM TEACHERS OR ADMINISTRATORS. THE EVALUATION SECTION OF AB 293 ASKS THAT THOSE JUDGMENTS BE ACTED UPON EITHER TO MAKE THE WORST BETTER OR TO HELP THEM FIND OTHER EMPLOYMENT.

THE DISMISSAL OR DUE-PROCESS PROCEDURES OF AB 293 ARE AN ATTEMPT TO ESTABLISH A MORE PROFESSIONAL REVIEW OF CASES WHERE ATTEMPTS AT UPGRADING HAVE ALLEGEDLY NOT BEEN SUCCESSFUL.

PREVIOUS LAW REQUIRED THAT IF A LOCAL DISTRICT WISHED TO DISMISS A CERTIFICATED TENURED EMPLOYEE IT WOULD BE REQUIRED TO FILE A CASE IN SUPERIOR COURT. TENURED CERTIFICATED EMPLOYEES WERE THE ONLY CLASS OF PROFESSIONALS FOR WHICH SUCH ACTION WAS REQUIRED. IT WAS OUR FEELING THAT SUCH A PROCEDURE DID NOT CONTAIN ENOUGH PROFESSIONAL INVOLVEMENT.

THE DUE-PROCESS PROCEDURES IN AB 293 PROVIDE THAT IF A CASE INVOLVES QUESTIONS OF FITNESS FOR SERVICES (INCOMPETENCY, DISHONESTY, WILFUL REFUSAL TO OBEY SCHOOL LAWS AND REGULATIONS, ETC.), IT WILL BE DECIDED BY A THREE-MEMBER PANEL.

WE REFER TO THIS BODY AS THE COMMISSION ON PROFESSIONAL COMPETENCY. THE COMMISSION ON PROFESSIONAL COMPETENCY SHALL SIT WHEN CHARGES INVOLVE THE QUALIFICATIONS OF THE CERTIFICATED EMPLOYEE TO CONTINUE IN HIS DESIGNATED CAPACITY, BASED UPON THE EVALUATION AND ASSESSMENT GUIDELINES ESTABLISHED BY THE LOCAL DISTRICT. THE COMMISSION WILL CONSIST OF THREE MEMBERS--THE HEARING OFFICER AND TWO EXPERIENCED PERSONNEL WHO SHALL HAVE AT LEAST FIVE YEAR'S EXPERIENCE IN THE SPECIFIC EDUCATIONAL FUNCTION OF THE ACCUSED, AS SET FORTH IN SECTION 13055 OF THE EDUCATION CODE. ONE OF THE TWO SHALL BE CHOSEN BY THE EMPLOYING AUTHORITY AND THE OTHER BY THE EMPLOYEE IN QUESTION. THE HEARING OFFICER FROM THE OFFICE OF ADMINISTRATIVE PROCEDURE SHALL SERVE AS CHAIRMAN AND A VOTING MEMBER OF THE THREE-MAN PANEL.

IN THOSE CASES REQUIRING A COMMISSION ON PROFESSIONAL COMPETENCE, THE PANEL SHALL, BY A MAJORITY VOTE, PREPARE A WRITTEN DISPOSITION CONTAINING THE FINDINGS OF FACT, DETERMINATIONS OF ISSUES, AND A DECISION EITHER TO RETAIN OR TO DISMISS THE

EMPLOYEE. ITS DECISION SHALL BE DEEMED TO BE THE FINAL DECISION OF THE SCHOOL DISTRICT'S GOVERNING BOARD. IN CASES HEARD ONLY BEFORE THE HEARING OFFICER, HIS DECISION SHALL ALSO BE BINDING ON THE SCHOOL BOARD.

THE HEARING OFFICER ALONE WILL CONSIDER ALL CHARGES DEALING WITH ACTS OF CRIMINAL SYNDICALISM, PHYSICAL OR MENTAL CONDITIONS WHICH AFFECT ABILITY TO PERFORM, CONVICTION OF A FELONY OR CRIME OF MORAL TURPITUDE, VIOLATIONS OF SPECIFIC CODES, OR KNOWING MEMBERSHIP IN THE COMMUNIST PARTY.

EITHER PARTY MAY, OF COURSE, APPEAL THE DECISION TO THE SUPERIOR COURT. IN INCLUDING THIS APPEAL PROCEDURE, HOWEVER, IT WAS NOT THEN, NOR IS IT NOW, MY INTENTION TO ALLOW THE COURTS TO HAVE TRIAL DE NOVO IN HEARING THE CASE. ALTHOUGH THE LAW DOES READ THAT THE COURTS MAY EXERCISE THEIR "INDEPENDENT JUDGMENT ON THE EVIDENCE," IT ALSO STATES THAT THE COURTS SHALL REVIEW ANY APPEALS IN THE SAME MANNER AS A DECISION MADE BY A HEARING OFFICER UNDER CHAPTER 5 OF THE GOVERNMENT CODE.

SO, AB 293 REALLY IS JUST A MANDATE OF AN EVALUATION SYSTEM COMBINED WITH CIRCUMSPECT CHANGES IN A DUE-PROCESS SYSTEM WHICH WOULD BE UTILIZED ONLY IN CASES WHERE ALL OTHER AVENUES FOR CHANGE HAD BEEN INEFFECTIVE.

AS MANY OF YOU ARE ACUTELY AWARE, LOCAL DISTRICTS HAVE BEEN WRESTLING WITH THE IMPLEMENTATION OF AB 293 FOR NEARLY A YEAR NOW AND SOME DISTRICTS HAVE, I UNDERSTAND, PUT FORTH AMAZINGLY COMPETENT FIRST EFFORTS. I AM OFTEN ASKED WHICH PROGRAMS WHICH HAVE BEEN ESTABLISHED MOST NEARLY MEET MY INTENTIONS AS THE AUTHOR OF AB 293. TO MY OWN AMAZEMENT I AM QUITE OFTEN ASKED

THIS QUESTION BY THE SAME PEOPLE WHO HAVE QUESTIONED MY COMPETENCE AS A LEGISLATOR FOR EVEN INTRODUCING THE BILL IN THE FIRST PLACE.

MY RESPONSE IS QUITE SIMPLE. ANY PLAN WHICH MEETS THE GUIDELINES OF THE BILL AND HAS GENERAL ACCEPTANCE AMONG TEACHERS, ADMINISTRATORS, PARENTS, AND SCHOOL BOARD MEMBERS OF A COMMUNITY IS A RESPONSIBLE PLAN.

IT WAS NEVER MY INTENTION TO ESTABLISH ANY STATEWIDE EVALUATION STANDARDS. I, OF COURSE, REALIZED THAT DISTRICTS WOULD, WHERE POSSIBLE, SHARE APPROACHES TO THE PROBLEM AND UTILIZE THE BEST INFORMATION AVAILABLE TO ADAPT EVALUATION PROCEDURES TO MEET THE NEEDS OF ALL CHILDREN IN THEIR DISTRICTS.

BUT AB 293 IS AN ATTEMPT TO ACCOMPLISH THE ESTABLISHMENT OF A FLEXIBLE MECHANISM FOR SETTING STANDARDS AT THE LOCAL LEVEL. I MUST EMPHASIZE THAT STANDARDS REQUIRED BY THIS LAW DO NOT REQUIRE A SINGLE CRITERIA. THE LAW DOES NOT REQUIRE THAT ALL CHILDREN LEARN AT THE SAME RATE. THE LAW DOES NOT REQUIRE THAT ALL TEACHERS NOR ADMINISTRATORS THINK ALIKE OR ACT ALIKE.

AB 293 DOES NOT TELL LOCAL DISTRICTS HOW TO EVALUATE--IT SIMPLY TELLS DISTRICTS THAT THEY MUST. FOR EXAMPLE, DESPITE PERSISTENT RUMORS TO THE CONTRARY, THE TERMS "TESTING" OR "MEASUREMENT" WILL NOT BE FOUND ANYWHERE IN AB 293.

I PERSONALLY BELIEVE THAT TESTING DEVICES WILL AND SHOULD BE USED AS A PART OF THE EVALUATION PROCESS, BUT THE EXTENT OF THE USAGE IS UP TO THE LOCAL SCHOOL BOARD IN CONSULTATION WITH AFFECTED CLASSES OF CERTIFICATED EMPLOYEES. I HAVE SUPPORTED ONLY ONE AMENDMENT TO AB 293 SINCE ITS ENACTMENT, AND THAT WAS TO SPECIFICALLY STATE THAT NORMS ESTABLISHED FOR

STANDARDIZED TESTS WOULD NOT BE UTILIZED FOR TEACHER EVALUATION. I BELIEVE THAT WE MUST LOOK AT INDIVIDUAL STUDENT PROGRESS RATHER THAN SIMPLY THE NORM OF ANY PARTICULAR CLASS.

I MUST NOTE THAT IMPLEMENTING AN EVALUATION SYSTEM IS NOT EASY; NO TASK OF CHANGE IS EVER EASY. AND I AM NOT SO NAIVE TO THINK THAT SYSTEMS DEVELOPED TODAY WILL NOT BE REVISED FOR THE BETTER TOMORROW. BUT THE TASK OF EVALUATING PROFESSIONAL QUALITY IN AND OF ITSELF WILL HAVE A POSITIVE EFFECT ON THE CHILDREN OF THIS STATE. THE PROCESS OF EVALUATION IS ONGOING AND MUST ADAPT TO THE NEEDS OF A RAPIDLY CHANGING SOCIETY. I WOULD BE DISAPPOINTED IF ANY SCHOOL DISTRICT ADOPTED AN EVALUATION SYSTEM FOR 1972-73 AND DID NOT CHANGE IT IN 1973-74.

BUT GIVEN ALL THIS, WE STILL MUST GET ON WITH THE TASK. I HAVE BEEN HEARTENED TO RECEIVE COPIES OF MANY PLANS ADOPTED BY LOCAL DISTRICTS WHICH SEEM TO MEET THE POSITIVE APPROACH OF THE LAW. ALAMEDA, SOUTH WHITTIER, MT. DIABLO UNIFIED, LAWNDALE, AND OTHERS TAKEN MADE POSITIVE STEPS ALREADY. I AM SURE THAT ALL DISTRICTS IN TIME WILL MEET THIS PROBLEM IN A POSITIVE MANNER.

NOW, ALL OF THIS MAY GIVE YOU SOME IDEA OF JUST WHAT WAS INTENDED WITH THE PASSAGE OF AB 293, BUT I WOULD NOT BE CANDID IF I DID NOT DWELL ON A MORE SERIOUS PROBLEM WHICH MOTIVATED ME TO PERSIST IN THIS FIELD AND WHICH NOW, MORE THAN EVER, APPEARS TO BE WIDESPREAD IN THE FIELD OF EDUCATION.

IN A RECENT ARTICLE IN THE LOS ANGELES TIMES, SIDNEY SLOMICH TELLS A STORY: "ONE NIGHT A MAN DREAMT THAT A MONSTER WAS ON HIS CHEST CHOKING HIM, TRYING TO KILL HIM. THE MAN WOKE IN TERROR AND SAW THE MONSTER ABOVE HIM. 'WHAT IS GOING TO HAPPEN

TO ME?' THE MAN CRIED. "DON'T ASK ME," REPLIED THE MONSTER. "IT'S YOUR DREAM."

I THINK THIS STORY APPLIES TO MUCH OF WHAT IS WRONG WITH PUBLIC EDUCATION TODAY. PROFESSIONAL EDUCATORS ARE QUICK TO SEE A THREAT TO THEIR EXISTENCE IN ANY DISCUSSION OF EDUCATIONAL CHANGE.

I HAVE RECEIVED MASSIVE NUMBERS OF LETTERS ON AB 293 SINCE ITS ENACTMENT. THE TONE OF A MAJORITY OF THESE LETTERS IS HOSTILE TO ANY FORM OF EVALUATION. THE UNDERLYING CURRENT OF SUCH LETTERS IMPLIED THAT PUBLIC SCRUTINY OF EDUCATION WAS AKIN TO SACRILEGE.

IN MANY AREAS OF EDUCATION IT IS AS THOUGH THERE HAS BEEN A MORATORIUM DECLARED ON COMMON SENSE.

FOR MANY YEARS PARENTS HAVE RETREATED BEFORE THE EXPERTS TO THE DEGREE OF ABSURDITY. AND THE EXPERTS HAVE BEEN INCREASINGLY UNREADY TO OPENLY BRING THEIR PLANS BEFORE THE PUBLIC FOR SCRUTINY.

NOWHERE IS THIS MORE OBVIOUS THAN IN THE FIELD OF TENURE. AB 293 DID NOT ELIMINATE TENURE--IT SIMPLY PUT THE LIGHT OF LAW ON THE ACTIVITIES OF PUBLICLY EMPLOYED SCHOOL PERSONNEL. FANATICAL DEFENSE OF LACKADAISICAL PERFORMANCE BY A FEW CERTIFICATED EMPLOYEES HAS NOT HELPED THE TRUE PROFESSIONALS BUT HAS MERELY SERVED TO LOWER THE STANDARDS AND SULLY THE REPUTATION OF ALL EDUCATORS.

BUT THE PUBLIC HAS FINALLY REALIZED THAT THE MONSTER OF SIDNEY SLOMICH WAS IN ITS OWN DREAM AND THAT IN THE FINAL ANALYSIS IT COULD CONTROL THE OUTCOME. CONTRARY TO THE ATTITUDE

OF SOME IN OUR SOCIETY, THE PUBLIC SECTOR--THOSE WHO ARE ULTIMATELY RESPONSIBLE FOR THE SCHOOLS--CAN AND WILL WITHDRAW SUPPORT IF IT IS NOT REASSURED THAT PROFESSIONAL QUALITY IS THE ORDER OF THE DAY.

THERE IS A CONFIDENCE GAP. AND IF CONFIDENCE IN OUR SCHOOLS IS NOT INCREASED, IT WILL BE THE CHILDREN WHO CONTINUE TO SUFFER.

AND ONE FINAL WORD ABOUT EDUCATION. I AM AWARE OF THE FACT THAT NEW DUTIES AND HIGHER QUALITY CANNOT BE ACHIEVED WITHOUT FINANCIAL SUPPORT FOR THE SCHOOLS--WITHOUT, I MIGHT ADD, THE CREATION OF A SANE SYSTEM OF SCHOOL FINANCE IN THIS STATE. BUT THE PUBLIC WILL NOT STRONGLY SUPPORT INCREASINGLY HIGHER LEVELS OF SCHOOL SUPPORT WITHOUT BEING CONVINCED THAT INCREASED QUALITY WILL RESULT FROM THAT FINANCIAL COMMITMENT.

I BELIEVE THAT THE MANNER IN WHICH EDUCATORS AND THE PUBLIC RESPOND TO THE CHALLENGE OF IMPLEMENTING AB 293 WILL HAVE MUCH TO DO WITH INCREASED FINANCIAL SUPPORT FOR THE SCHOOLS.

THE GOOD FAITH AND RESPONSIBILITY WITH WHICH THE TEACHING PROFESSION MEETS THIS NEW LAW WILL HAVE MUCH TO DO WITH PUBLIC ATTITUDE. TENURE REFORM IS AN IDEA WHOSE TIME HAS COME, AND COOPERATION BY ALL CERTIFICATED EMPLOYEES AND SCHOOL DISTRICTS IN IMPLEMENTING THE NEW LAW IS THE ONLY INSURANCE THEY HAVE THAT TENURE WILL BE MAINTAINED.

CICERO TOLD THE UNVARNISHED TRUTH IN SAYING THAT THOSE WHO HAVE NO KNOWLEDGE OF WHAT HAS GONE BEFORE MUST FOREVER REMAIN CHILDREN.

WHAT HAS GONE BEFORE IN TEACHER TENURE HAS NOT BEEN SUCCESSFUL--IT MUST NOT BE REPEATED. IT HAS JEOPARDIZED THE PRODUCT. A QUALITY PRODUCT--A YOUNG PERSON PREPARED FOR LIVING--SHOULD

BE OUR COMMON GOAL IN THE SCHOOLS. THIS CAN ONLY BE ACHIEVED WITH DILIGENCE, DEDICATION, AND COMMON SENSE ON THE PART OF ALL CONCERNED.

IT'S UP TO ALL OF US.

LET'S GET ON WITH IT!

# # #

FOURTH GENERAL SESSION

SPEAKER: ARTHUR P. COLADARCI
DEAN, SCHOOL OF EDUCATION
STANFORD UNIVERSITY

TOPIC: "EVALUATION: SACRED AND PROFANE"

EVALUATION: SACRED AND PROFANE
(Abstract)

Arthur P. Coladarci

(Arthur Coladarci is Dean of the School of Education at Stanford University and Professor of Education and of Psychology. He received his B.A. degree from Western Connecticut College and both his M.A. and Ph.D. from Yale University.)

The level of our technical and conceptual understanding of evaluational processes has risen emphatically in the past quarter century. The application of this knowledge in our educational institutions, however, continues to be relatively primative, uncritical and ineffective.

In these comments, I wish to suggest some distinctions that may be heuristic in this connection and a danger that is avoidable.

Three Faces of Evaluation

It is useful, I believe, to realize that "evaluation" is not a process, similar in form, dynamics and logic wherever it is applied. Evaluation may differ considerably across differing evaluative purposes and functions. Three rather distinct purposes are involved and they differ in process and consequences (I "piggy-back" here on a distinction offered by Lee Cronbach almost a decade ago):

(1) Evaluation for decisions about administrative, regulatory and procedures. The purpose here is that of judging the adequacy and efficiency of the administrative organization, operations and procedured rules of the institution, with emphasis on structure rather than functions of the institution. The possible concerns are infinite in number: the degree to which the institution's organization permits change, the fidelity of budget administration, the availability of a place to which a teacher may retire to contemplate his

divine essence or (less divinely) repair a faculty zipper.

(2) Evaluation for decisions about students. The purpose here is to make and remake decisions on selection, "needs," status, placement, career, assignment, etc. In this function individual differences are central and relevant, in contrast to (1) above.

(3) Evaluation for decisions about the institutional program, method, content and organization of learning. The reference here is clearly to the primary raison d'être of an educational institution -- and, accordingly, one would expect this evaluational function to be dominant in attention and competence. However, I am constrained to the conclusion that, confounding this function with the previous two, we perform it badly in the everyday work of the school. The controlling concern is (or, rather, should be) assessing the effectiveness of instructional techniques, organization of instruction and materials of instruction. (Compare the modal use of tests to determine whether a pupil should be consigned to hell, purgatory or heaven.) The conception of "teaching as hypothesis-making," which I introduced many years ago, continues to be helpful (unlike most of my earlier ideas!) at this juncture. Consider, however and unhappily, how little of the enormous "testing and evaluating" business in our schools is explicitly or even accidentally addressed to the assessment and improvement of their raison-di'êntre function -- teaching. At the system level, each year, batteries of tests are administered, scored, summarized and stored with only rare reference to their implications for system-improvement. Also, at the system-level, and assuming (awkwardly) that the tests are relevant to instructional purposes, consider the continuing inefficiency in time and costs in administering all tests to all pupils -- when the data needed for system decision can be obtained by sampling both pupils and test subsets. At the teacher-specific level (to me the most critical and productive), consider how infrequently data are produced that are relevant in kind and immediacy-of-availability to assessing and modifying instructional decisions. (I suspect that the savings in time and funds resulting from sampling pupils and tests in system-wide testing would go a long way toward making it possible for teachers to learn how to develop and use measurement-evaluation procedures immediately appropriate to the continuing teacher-specific task.)

The third evaluation function, on which I have just commented on has both summative and formative components, to use the language of Scriven in a now classic presentation. It appears to me that while the summative mode will and should be applied in making decisions about whether an institutional procedure should be

"exported" and in periodic reporting to the schools' clients, we must direct priority attention to the application of formative evaluation procedures in the daily professional life of the teacher and administrator. It is probable, however, that a massive in-service education program is needed. The practicing professional educator must learn that formative evaluation is neither "untidy" or unscientific; if done conscionably and well; it can be disciplined and rigorous, albiet not as unambiguous as summative evaluation techniques. While summative evaluation may be conceptualized in many ways, I like the following formulation (with apologies to Stake):

(a) The continual definition of the purposes and rationale for instruction. This is needed also to locate all of the reference groups that will be affected by the instruction and must be involved.

(b) The continual identification of "antecedent data" -- all conditions existing prior to instruction; "entry behaviors," in parallel recent language.

(c) The continuous identification of "transactional data." The instructional process is thought of as comprising countless transactions among pupils, materials, teachers, etc.

(d) The continuous identification of "outcome data." Here we refer to all of the data bearing on the intentions of instruction.

The formative task, inherently complex but not impossible, is constantly to look for congruencies between transactions and outcomes, remedying as we go. Done well, instruction is never the same; i.e., it is always forming and reforming.

The Danger of Oversimplification

I take the liberty of offering a long quotation from a statement by Stake earlier this year; it cannot be improved upon, in my view:

> The whole cloth is a grand accumulation of intense transactions and outcomes. The teachers intend to deliver on many promises and to take advantage of many targets of opportunity in the usual school setting. Each child brings his own complex convictions, misunderstandings, and propensities, and he takes away some of these and others as well. Each classroom is a community with rules and stresses, competition, compassion. Yesterday's subgroups are not tomorrow's. Things are learned, unlearned, relearned, much as shoelaces are knotted, untied, broken, retied. An educational program

has countless objectives, many of them dormant until a crisis arises. The priorities vary over time, from person to person, and no statement of program objectives ever devised has come close to representing the real world intents of people involved in an educational program. This is the whole cloth of an educational program.

I shouldn't imply that we can't get reasonable consensus as to what the priority objectives are, but that's consensus. The unspoken objectives -- such as safety in the classroom, sharing of work responsibilities, developing a sense of humor, a respect for rules, a tolerance of ambiguity, and so on -- are left to take care of themselves, at least until a crisis arises. Then these objectives may preempt all others. One can get a simple consensus for a list of top objectives as long as no one takes the consensus too seriously. Children can still get more than a primitive education, the primitive education that the consensus statement describes. A century ago, the Swiss historian, Jacob Burckhardt, stated that the essence of tyranny was the denial of complexity, and he forsaw that ours would be the age of the great simplifiers. To quote from Patrick Moynihan on Burkhardt, 'He was right, this is the single greatest temptation of the times, it's the great corrupter, and it must be resisted with purpose and energy.' Consensus is a great simplifier, and so is theory: statistical processes are simplifiers, test scores are simplifiers, simple representations of the complex. These simplifiers help us by reducing the complex phenomena to something within our power of comprehension, but they also mislead us, by saying that education is much less than it really is.

We work day by day with simplifications, with statements of objectives, with central tenuencies, with criteria tests, and we become transfixed by them, losing our awareness of the fundamental activities of teaching and learning. We do it to ourselves and we do it to our audiences, that is, the people for whom we are articulating the program evaluation. Evaluators should be helping people keep in touch with the reality of instruction, but our scrapbooks are full of enlargements of enlargements.

There is abundant evidence that the danger of oversimplification in evaluation is real. It is illustrated unhappily in the naive and distorting reductionism that occurs in jumping

from definitions of purpose to operations of measurement: a current emphatic example is found in school system acrobatics in attempting to comply with the requirements of the Stull legislation.

SYMPOSIUM I

TEACHER EVALUATION THROUGH MEASUREMENT OF STUDENT ACHIEVEMENT

Chairman: Dr. W. James Popham
Professor of Education
University of California, Los Angeles

Participants: John D. McNeil
Professor of Education
University of California, Los Angeles

TEACHER EVALUATION THROUGH STUDENT ACHIEVEMENT

John D. McNeil

(John D. McNeil is Professor of Education at the University of California, Los Angeles. He received his B.A. and M.A. degree from San Diego State College and Doctorate degree from Teachers College, Columbia University.)

"By the work one knows the workman." The success of the teacher is best shown in the progress of his pupils. Those evaluating teachers have, however, seldom followed this maxim. Judgment of the teacher's classroom processes, general background, characteristics, personal conduct in the school and community has dominated the realm of teacher assessment. Now, a number of circumstances are making it necessary for those who would judge teachers to collect evidence regarding the teacher's ability to effect pre-specified changes in pupils. There is, for instance, civil rights legislation demanding that decisions, such as teacher selection, rest upon data directly related to work with pupils, not upon the educational and cultural accomplishments of the teacher himself. The desire of legislators, administrators, and leaders of teachers' organizations to restore public confidence in schools is another factor in motivating measurement of outcomes. The public wants assurance that children are protected from incompetent teachers. Evidence collected systematically regarding comparative standing of teachers in terms of student achievement is believed to be more reassuring than tenure (longevity) and licensing (courses completed). This is not to say that the only reason for evaluating teachers by results is so the unfit can be weeded out. On the contrary, the number of teachers who may be fired for failing to advance the learning of pupils is insignificant in comparison with the number of those being evaluated so that they can improve.

Reservation in accepting student achievement as the principal criterion of effectiveness centers chiefly on concerns about the adequacy of measures for assessing a wide range of outcomes and failure to account for instructional variables that the teacher does not control. Are these preoccupations valid? Is educational technology answering them? Let's see:

1. Criticism that measures of pupil progress are insensitive. Standardized tests have been the traditional instruments for measuring the effects of instruction. These tests are frequently vulnerable to the charges that they are not likely to contain both the breadth and depth of content coverage necessary to make a detailed assessment. Further, the results of such tests jar the American's conscience about equality by the fact that the tests are designed to give a set of scores which will produce a "normal" distribution dooming half of a class or half a school to below average.

Response. There has been an increase in availability of criterion-referenced tests which focus on many specific objectives, such as critical skills in

learning to read, which nearly all pupils are expected to master. Whereas a standardized test may measure, say, five objectives for a course of study, a set of criterion-referenced tests permits the measurement of five times that number of objectives, some of which might be prerequisite to more encompassing terminal objectives. With the standardized tests, the teacher gets credit only when pupils succeed on the five objectives regardless of the number of prerequisite skills effectively taught. With criterion-referenced tests, teachers have a better chance to receive recognition for whatever they have taught successfully.

2. Criticism that teachers will teach too trivial, easily attained objectives.

Response. The practice of holding preinstructional conference helps to ensure more warranted intents. At the conference, proposed objectives are reviewed by supervisors, colleagues, parents, and pupils. The quality of the conference is enhanced by discussing the probability that the learner needs the competency specified in the objective, that the objective is teachable, and that the objective will be acquired only through efforts of the teacher.

3. Criticism that collecting evidence of pupil progress is too costly.

Response. One hundred years ago inspectors in British schools engaged in pupil sampling, examining every fifth pupil. This sampling procedure is economical, especially when one is measuring attainment of objectives that call for observation of pupils' processes and oral production. Currently, sampling procedures also include the sampling of items and tests. With item sampling, each student gets a sample test or sample item drawn from a large population of tasks that are taught in the course. Different students take different tests, thereby, reducing time required for testing.

4. Assessment of pupil progress does not reveal why the teacher is succeeding or failing.

Response. More than seventy observation schedules are available for use in recording or analyzing a teacher's practices. These schedules can be applied to observation of single lessons and units of instruction by supervisors, colleagues, pupils, and even the teachers (provided a record of classroom interactions has been made).

A most recent tool The Teacher Improvement Kit has been designed to help the teacher diagnose teaching strengths and weaknesses and develop improved teaching strategies. Kits treating different subject matter and kinds of teaching tasks offer a range of minilessons including background information about the task, pupil posttests, normative data, and specific procedures for getting better results should a need for improvement be indicated.

5. Comparison of teachers on the basis of pupil progress is unfair inasmuch as the conditions under which instruction occur are not standardized.

Response. Researchers at Instructional Appraisal Services are now developing controlled performance tests that afford all participating teachers equal

opportunity to display their competency to effect changes in learners (both achievement and affect). Their performance tests control for teaching tasks, instructional material, time for preparation and instruction, and ability and predisposition of pupils.

* Examples of sources for newer instructional technology include:

1. Criterion-referenced tests
   a. Wisconsin Research and Development Center--The Wisconsin Design for Reading--University of Wisconsin, Madison, 53706

   b. Instructional Objectives Exchange--P.O. Box 24095, Los Angeles, California 90024

2. Observation schedules--Mirrors for Behavior. Research for Better Schools, Philadelphia, Pennsylvania

3. Teaching Improvement Kits--Instructional Appraisal Services, P.O. Box 24821, Los Angeles, California 90024

AA5

SYMPOSIUM II

STATEWIDE EVALUATION AND TESTING

Chairman: Dr. Mabel C. Purl
Director of Research and Evaluation
Riverside Unified Schools

Participants: Alex Law, Chief
Program Evaluation Office
State Department of Education

Assemblyman Leroy F. Greene
3rd Assembly District
State of California

SYMPOSIUM III

EVALUATING DISTRICT EVALUATION PROCEDURES

Chairman: Dr. Lester W. Ristow
Assistant Director
Division of Research and Pupil Personnel Services
Los Angeles County Schools

Participants: J. Alden Vanderpool
Teacher Education Executive
California Teachers Association

Harold W. Stombs
Consultant to California Teachers Association
On Employee Evaluation Systems

Garford G. Gordon
Research Executive
California Teachers Association

Theodore Bass
Professional Services Executive
California Teachers Association

CERTIFICATED EMPLOYEE EVALUATION PLANS

J. Alden Vanderpool
Harold W. Stombs

(J. Alden Vanderpool is Teacher Education Executive for the California Teachers Association, Burlingame. He received his B.A. degree from Central Washington State College, his M.Ed. degree from the University of Washington, and his Ed.D. degree from Stanford University. Harold W. Stombs is Consultant to the CTA Evaluation Review Service. He received his B.A. and M.A. degrees from San Jose State College.)

In the 1971 session, the California State Legislature passed the Stull-Rodda Professional Competency Act. While retaining employee tenure, the Act outlines new procedures for dismissal of certificated employees. One of the causes for dismissal is incompetence, a charge which has rarely been successfully used in dismissal proceedings in California.

Section 13407 requires the notice of incompetency to include the evaluation made pursuant to Article 5.5. This article stipulates four elements which must be included in the evaluation: expected student progress, preserving a suitable learning environment, maintaining proper control, and other duties normally required as an adjunct to regular assignments. Common practice has based evaluation on presage and process criteria. Evaluation now must also be based upon the product criterion, expected student progress.

The Stull-Rodda Act requires that these criteria apply not only to teachers, but to all certificated employees. The employing authority also has a responsibility to provide assistance to any employee whose performance is in need of improvement.

Most of the evaluation plans we have examined start from the premise that evaluation should be a positive procedure that should result in improvement in the instructional program. This is commendable and may tend to reduce the threat that is implicit in evaluation. The consequences of evaluation should be positive steps to correct deficiencies rather than punishment or dismissal. Whether or not the process of evaluation will fulfill expectations for an improved educational product depends on the criteria used to describe a good instructional program.

Several problems result from attempts to implement the Stull-Rodda Act. One of the most common problems among those we have encountered is that some districts are producing evaluation plans that suggest that they do not understand the intent of the law. They have written a few goals and objectives and plan to continue evaluating as they have always done. While their intentions are undoubtedly good, years of practice, lack of understanding and expertise, and the monumental size of the task have combined so that, in our opinion, few districts have fully or successfully implemented the Stull-Rodda Act. With no proven models to rely upon, this conclusion is not alarming because districts cannot be expected to build new evaluation systems in one year.

A more profound problem, and one that should be of concern to researchers, is that we do not really know what specific teacher behaviors will result in better learning by the students. This may be due in part to the lack of a theory or theories of instruction, with the result that there is little systematic application of learning theories to the practical teaching situation.

Another serious problem is how the evaluative criteria should be applied to non-instructional personnel. The law requires that principals and superintendents be evaluated on the basis of expected student progress. We are seeking ways to reduce this principle to procedures.

Another intricate problem is institutional evaluation. Traditional assessment tends to be blind to the perceptions of those employees who are served by others. This is a reflection of the differential rate at which communication moves up or down through the structure of an organization. Higher office can become a refuge from criticism, with the result that failure of the system often reflects most on those who are not primarily responsible. The problem, then is to provide for assessment of administrative and service staff in important part by instructional staff, and, at the same time, preserve and enhance mutual respect and the dignity of higher offices. An obvious way to provide for assessment of administrative and service personnel is to include provision for incorporating the opinions and judgments of those supervised or served. Furthermore, to preserve and enhance mutual respect and the dignity of higher offices, it is necessary to ensure that higher positions are staffed by competent

individuals and that those individuals remain alert to their responsibilities. It is assumed that evaluation can contribute to this end.

A practical problem that recurs is an evident lack of technical adequacy on the part of most school district personnel in writing objectives. Part of establishing standards of expected student progress is stating them as student objectives. Such objectives should reflect the educational program and be directed toward the educational goals of the district. They should also be suitable for use in evaluating employees. The objectives should not be so broad that their achievement would not be expected within the evaluation period, usually one year. Neither should they be directed at the daily-lesson level wherein great catalogs would be required to contain them and excessive energy would be required to monitor them. We also advise against the use of standardized test norms in measuring student progress for these purposes.

The following are items that require particular attention when analyzing an evaluation system. They are not arranged in any order of priority, and the list is not exhaustive.

General

The evaluation procedures apply uniformly to all certificated employees.

All procedures, definitions, objectives, criteria, and measures are formally adopted within the evaluation system.

When required elements are not stated, such as standards of expected student progress, then there are procedures for generating such elements.

When required elements are not included in the evaluation document, reference is made to other documents which are readily accessible, i.e., job descriptions, district goals.

A distinction is made and there is a satisfactory balance between presage, process, and product criteria.

Certificated instructional employees were properly involved in establishing the evaluation system.

The application of standardized norms to student assessment for purposes of evaluating employees has been excluded from the system.

Evaluators are evaluated on the quality of their skills in assessing and evaluating other employees.

Evaluators are evaluated on the quality and results of assistance provided to evaluatees in need of improvement.

There are provisions to balance the power discrepancy between evaluator and evaluatee when arriving at mutual acceptance of standards of expected student progress and other elements of evaluation.

Definitions of terms are clear and accurate.

There are fair procedures for resolving differences regarding mutual agreement to standards of expected student progress and other elements of evaluation.

The guidelines make a distinction between evaluation as a process rather than an event.

There are procedures for institutional accountability (reciprocal, contextual evaluation).

Possibilities for ex post facto evaluation are excluded from the system. (No statements such as "performs other duties as assigned.")

The guidelines encourage a colleagual rather than an authoritarian relationship between administrators, instructional, and service personnel.

Report forms include only the adopted elements of evaluation.

Report forms include all the adopted elements of evaluation.

The use of all report forms is fully explained.

Final evaluation report forms yield a profile typical of the employee's service.

Subjective criteria are excluded if rating scales are used.

The assumption is avoided that evaluation, under the Stull-Rodda Act, is synonymous with adopting a PPB System or utilizing that format.

Deadlines and time allotments for assessment are satisfactory.

Employee evaluation is an integral part of an overall assessment of the educational program, and the results of this assessment are properly reported to the governing board.

There are provisions for inservice training.

There is a procedure for an annual overall review of the evaluation system.

Pupil Progress

Goals related to each certificated employee are clearly stated.

Pupil-progress objectives are stated for each organizational unit of the school system.

Success criteria statements are adopted for each objective.

Procedures are stated for assessing pupil status in relation to each success criteria.

A mechanism has been established for assuring valid application of each method of assessment of pupil progress.

There is a list of conditions or circumstances (mitigating in their effect) to be considered in staff evaluation on the basis of pupil progress.

The list of conditions or circumstances includes minimum standards which should be met before the employee can be held responsible for pupil progress.

There are procedures for assessing these conditions or circumstances.

Criteria are stated for determining adequacy of these conditions or circumstances.

There are allowances for borderline conditions or circumstances where minimum standards have been barely met.

There are procedures for reassessing fundamental standards of pupil progress.

Maintenance of a Suitable Learning Environment and Proper Control

"Suitable learning environment and proper control" is defined for each organizational unit of the school system.

There are procedures for assessing the suitability of the learning environment and control exercised.

There are definitions of unacceptable maintenance of the learning environment and exercise of control.

There are procedures for review and reassessment of definitions of the suitability of the learning environment and of proper control.

Adjunct Duties

There are procedures for determining the load represented by adjunct duty assignments.

There is a statement of objectives and success criteria for each adjunct duty assignment.

There are procedures for assessing the employee's performance of his adjunct duties.

There are procedures for review and reassessment of adjunct duty assignment.

Evaluation of adjunct duties is in proper perspective with respect to the other elements of the evaluation system.

Standards have been established in each area of study.

When predicting student achievement, baseline data has been, or will be, established.

Objectives are not so broad and general that they are of no practical use.

Objectives are not so narrow that they are trivial or become so numerous that they are unmanageable.

Objectives are technically well constructed.

Objectives are directed toward, and consistent with, the educational goals of the district.

In statements of objectives, measurement is to be performed using criterion-referenced measures rather than standardized test publisher's norms.

G E N E R A L P R I N C I P L E S

GARFORD G. GORDON

(Garford G. Gordon is Research
Executive at the California
Teachers Association)

INTRODUCTION

An evaluation is a judgment applied to a situation or to a process. It is often thought of as being applied to an individual. This, however, is a secondary effect, reflecting the evaluation of the process or situation for which the individual is responsbible. It involves application of additional considerations, such as the degree to which the individual has responsibility for or is able to influence the process or situation.

While evaluations can be made in many ways, valid ones require that the judgments be made on as accurate and complete data as possible about factors having meaningful relationships to the situations and processes being evaluated. The meaningfulness of the relationships of any factor to a situation or process depends upon the objectives of the process or the aspects of the situation which are considered significant.

There are two aspects of the function of educational systems. One is the process of education, which is designed to achieve certain objectives in modifying the behavior (in the broadest sense) of the students. The other is the maintenance of certain levels of physical and psychological well-being on the part of students, parents, and community during the time students are subject to the process. It appears, therefore, that valid educational evaluation requires consideration of both situational and process factors; and that judgmental standards should relate both to the objectives of the educational process and to the environment in which it is conducted.

PURPOSES OF SCHOOLS

General. The specific purposes of a school system and of individual units within it--classes, schools, departments--depend upon the society within which it operates. However, certain things are common to all. Education is behavior modification--in other words, those who have completed their schooling or any portion of it should behave differently than they would have behaved if they had never been to school. Such modification of behavior may be trivial or profound, it may reflect conditioning or deep internal changes of perception, philosophy, and emotions. However, only the overt behavior changes that result will show whether or not an observable change has been wrought in the student. Anybody is at liberty to hypothesize internal changes not reflected in outward modifications of behavior; but he has no way to show that change has in fact occurred in the absence of outward manifestations.

A second purpose of the schools is the assumption of responsibility for children and youth over a significant part of their life spans. While it is hoped that activities conducted by the schools during this period will contribute to desired behavior modification; there are other objectives, including maintenance of social and physical conditions the community considers desirable for its younger members. Whether or not meeting the objectives in this area will promote--or even hinder--the attainment of objectives in behavior modification, this function of the schools is an accepted part of their role in modern society.

Any evaluation of a school system or of any units of it must recognize the dual purposes of schools. Although the underlying philosophy and societal demands have not been made explicit, current developments in legislative and other arenas toward requiring greater "accountability" of the schools make the need to recognize the dual purposes clear.

INSTRUCTION

It is true that the purpose of instruction is the modification of behavior. But this must be understood in its broadest sense. Certain segments of the school enterprise are considered primarily to be engaged in modification of specific behaviors. This is chiefly true in the so-called skill areas. However, modern communities expect much broader modifications also--those of the type implied in statements such as, "ability to get and hold a job," "improved self-image," and "appreciation of cultural differences." These broader objectives clearly imply that instruction must not only result in specific cognitive and psychomotor behavioral changes; but that it must operate so as to maintain a physical and emotional environment satisfactory to the community and conducive to achievement of broader objectives.

SUPPORT SERVICES

Support services have two broad purposes. One is the support of instruction by providing facilities, materials, and professional assistance in instructional activities. The second is working directly with students to supply the personal assistance and the environmental conditions expected and required of the schools if their role expectations are to be met.

ADMINISTRATION

The term administration is used rather than "school management" since one major function of administration is to mediate between the community, parents, and instructional and support personnel. The other major function is of course management per se--the securing and distribution of supplies, equipment, auxiliary services, physical facilities, and the conduct of coordinating and liaison activities.

ASSIGNMENT OF RESPONSIBILITIES

General. There are many teaching vice-principals, teaching principals, and even teaching superintendents. The custom of using counselors who teach one or more classes is widespread. Department heads or other lower echelon administrative personnel are rarely free from direct instructional duties. It must, therefore, be recognized that no position title can be assumed to convey the full description of the functions and responsibilities of an individual occupying the position. Valid evaluation of the performance of any individual must be based on an accurate description of the responsibilities assigned to that individual and not on job titles, or even on broad-category statements of job responsibilities--e.g., a general statement of the responsibilities of the position of teacher does not provide sufficient distinction between or definition of the responsibilities of teachers of different student groups, grade levels, and subject areas.

Teaching Responsibilities. While broad guidelines as to what persons in teaching positions are expected to accomplish are desirable, they cannot constitute an adequate basis for evaluation. They must be supplemented by specific statements as to the results expected. This applies not only to pupil progress but to managerial functions such as provision of proper environmental conditions and exercise of control over pupil activities. When the authority or resources needed for any function assigned to a teaching position are not under the control of the person in that position, the responsibility must be clearly assigned to another position.

Support Responsibilities. These may be assigned to any position in a school district. Some positions may be exclusively responsible for supporting instruction by working on the psychological and social problems of pupils. Others may be responsible only for the furnishing of necessary instructional materials. Most will have several responsibilities--e.g., a librarian is typically responsible for providing library materials, for helping students learn to use the library more effectively, and for maintaining order among students using it on an individual basis.

Administrative Responsibility. The responsibility of an administrative position can be very broad indeed. However, no person filling such a position is Superman. Broad statements of responsibility, while theoretically true, are usually unhelpful unless supplemented by specific statements of operational responsibilities. For example, a principal may be said to be responsible for the educational program of his school. Specifically, though, he is responsible for teacher assignments, for maintaining liaison between his staff and the central administration, for seeing that clerical tasks are performed on time, and for supporting the teachers vis-a-vis parents and problem students.

GOALS AND OBJECTIVES

Purpose of Goals and Objectives. The original goal of American schools as stated, was to teach children to read the Bible. In the next two centuries, the preparation of persons equipped to be clerks, accountants, and shipmasters was added. Later, the imparting of common social and political goals and the stimulating of a common patriotism were added. These and other goals were not added for schools in all parts of the country, nor were their implications ever analyzed at the local level. They were sythesized into such generalities as the "Seven Cardinal Principles" at the national, academic level.

The natural consequence of this development was the creation of widely varying individual and local interpretations of what the goals of the schools and of any given school program really were. As a result, any school person was forced to enter into a guessing game and to gamble that his idea of the goals he was to strive for would be the ones valued by whoever evaluated his work. To avoid the chaos caused by this situation, which still exists in great degree, educators preferred to evaluate each other--and be evaluated themselves--according to how they acted, how they appeared, and their experience and academic record.

Research has shown little connection between these presage and process evaluations and the achievement by pupils of any agreed upon levels of success in any area. Which means that evaluation has been a matter of "pleasing the boss" and "keeping your nose clean." The movement towards establishment of objectives which are clearly set forth before school activities commence is an attempt to assure that personnel will be evaluated according to the results they get in progressing toward objectives that are truly relevant to the purposes of the schools.

Objectives vs Goals. There is no precise boundary between goals and objectives. In a pure form a goal is a statement of a purpose of the schools. It gives direction but no specifics. Thus, "to produce better citizens," is a pure goal statement. On the other side, a pure objective is a statement of a specific observable result to be accomplished by a particular educational operation within a specific time span. For example, "by the end of the second week of school to have every pupil write his name in the correct position on each paper he does in class," is a pure objective statement.

Some theorists hold that a sufficient number of pure objective statements in relation to any educational goal will suffice to determine the extent to which the goal is being served by the schools. This is both theoretically and practically fallacious. Theoretically, because it does not provide for the interaction among human experiences in developing skills and knowledge which may, and usually does, make the whole quite different from the sum of the parts. Practically, because any human behavior, even such a clearly defined one as reading, involves far more specifics than can be noted and recorded within the limits of a school day, a school year, or even the entire school career of a student. Problems of record keeping and staff time only serve to reinforce the impracticality of the pure objective approach.

If pure goal statements are too general, and pure objective statements are too limited, what is the answer? The answer, of course, is some form of intermediate statement; or the limitation of objectives to be considered in evaluating a school activity to a feasible number, with the hope--unverified by research and contradicted by most experience--that pupils will somehow fill in the gaps on their own. There is no need to accept the latter suggestion, since there can be a complete range of statements bridging the gaps between those that are "pure objective" and those that are "pure goal."

Because goal statements have typically been too vague for any progress toward the goal to be monitored while instruction is occurring, and because pure objective statements provide for definitive determination of whether or not a specific instructional program has attained the objective stated, many groups and individuals have chosen to use the latter in evaluation. Unfortunately, many have overlooked or forgotten the limitations of specific objective statements and have denied that there can be intermediates between these and general goal statements. This is not true. In the present state of education only a few specific objective statements can be genuinely useful in evaluation. Intermediate statements are a necessity.

Behavioral Objectives. The well-publicized and widely disseminated positions of those espousing specific goal statements have led to a serious misunderstanding among many educators about the nature of objectives. While it is probably true that educational experiences modify the internal psychologies of individuals, even though there is no external manifestation of change, the only way data can be collected for assessing the effects of educational programs is by observation of the behavior of those participating in them. This behavior may be manifest through any type of human activity. Saying certain things in response to a verbal stimulus known as a question, is a form of behavior. So is making marks in specific places on a form in response to a visual stimulus. So is spontaneous reading of a book, or refraining from attacking a smaller person.

It is a fact, then, that all statements of objectives which will contribute to the collection of data relevant to how an educational program is succeeding must be behavioral. The behavior may be such that changes, if any, will be very subtle; or it may be so simple and overt that change, or lack thereof, is easily determined. The fact that certain psychologists, and promoters of instructional systems, have attempted to re-define behavior to include only the latter type, is to be regretted and does not alter this fact.

DEFINITION OF EVALUATION

Evaluation is a judgment. If it is based on data relevant to the purpose of the activity being evaluated, it is good evaluation; if it is not based on data, or is based on irrelevant data, it is bad evaluation. Evaluation is also a judgment based on criteria for satisfaction. If the criteria are realistic, it can be good evaluation; if they are not realistic in terms of the situation being evaluated, it will be bad evaluation. If either the purpose of the activity or the criteria are changed in the course of the activity, it will be *ex post facto* evaluation. Insofar as programs are concerned, being *ex post facto* does not automatically make evaluation bad. However, if the evaluation is to be used to determine the adequacy of the performance of an individual working in the program or responsible for it, then *ex post facto* evaluation cannot be validly applied, and its use is an obvious denial of due process to the individuals involved.

REQUIREMENTS FOR EVALUATION

Objectives. The purposes of an activity must be determined before any judgment can be made as to whether or not, or to what degree, the activity is succeeding. The essential requirement of an objective is that it is possible to determine by some type of observation whether or not it has been achieved. It is an improvement if it can be made possible to observe partial achievement or "near misses." However, quantification of degree of achievement of an objective, or of the wideness of the "misses," is not essential. A false appearance of objectivity can be given by making statements such as, "By the end of the semester 80% of the students will..." The decision that 80% is the objective and that the particular dimension is relevant, are purely judgmental. The use of specific numbers does nothing to make such an objective more useful. This type of quantification is usually more harmful than helpful.

Dimensions. Very simple objectives call for collection of only one kind of data. For example, if the objective were, "to teach pupils how to get the basketball through the hoop," the only data to be collected is the number of times the ball goes through. This can be refined by recording also the number of misses and mathematical manipulation can take place, such as calculating percentage of tries resulting in success, variation among students in percentage of success, and so on.

Most objectives, however, will require collection of more than one type of data. For example, if the objective were to "teach pupils how to score in a basketball game," we would have to add several dimensions to the one of getting the ball through the hoop. Perhaps one dimension would concern the number of fouls committed, another might concern judgment in passing, another might concern ability to judge when to try for a basket, and so on.

A dimension is an activity or a condition, the quality of which determines whether or not an objective has been met. Usually it also determines the degree to which it has been partially met or the closeness of "misses" for those objectives that are not absolute. (An absolute objective is one that is either met or not, and concerning which there is no interest in how close the activity came to meeting it, if it were not met.)

In addition to being logically related to the objective, a dimension must meet the test of practicality. It may be impossible to collect data in a given dimension. For example, an objective might be to "increase student appreciation for American literature." Several dimensions would relate to this objective; but the one that might first come

to mind, "how the students' appreciation actually changed during the course," is one on which no data can be collected; at least not until mind-reading techniques are perfected. Instead, dimensions from which inferences about appreciation of American literature might be drawn must be substituted--e.g., books read, extent of participation in discussion on American literature, opinions expressed, and so on.

Practicality also includes the effort that must be devoted to data gathering and recording. There is a limit both to teacher and student time and, in most cases, to the clerical and other assistance available. Obviously a dimension that greatly escalates the time and resources required for data collection and recording should be considered only in special cases.

Criteria. The collection of data for dimensions relevant to an objective is not evaluation. The summation of the data can be termed an assessment of the results of an educational program relative to the objective. Whether the results are to be judged as satisfactory, as unsatisfactory, or as partially satisfactory, depends on the evaluative criteria chosen.

Criteria must be unambiguous and unequivocal. They need not be numerical. In fact, the use of numerical criteria may be harmful because they give a false appearance of objectivity to a political and subjective process. Some criteria are clearly indicated by the objectives to which the dimensions relate. For example, if it be accepted that one objective of driver training is to enable students legally to drive motor vehicles, then the criterion for success is clear. The objective has been met for any student successfully securing a license and not for one who does not get licensed.

However, most criteria involve some type of quantitative statement, such as, "80% of the students will...," "the student will achieve a score of 10...," "the student will have a markedly increased ability to..." The last is a quantitative criterion even though numbers are not used. The problem in all these cases, is to arrive at some type of consensus as to what degree of success is acceptable. In the numerical cases, these numbers are simply guesswork or are judgments based on experience or are compromises of several such judgments and guesses. In the last example, the definition of, "marked increase" or the judgment that the change is a "marked increase" is likewise found to be either a guess, a judgment based on experience (educated guess), or a combination of several of these.

Any process whereby guesses, opinion, and judgments are compromised and pooled is a political process. However, in many educational situations, the process is more overtly political with staff politics, governing board politics, and community politics all entering into the establishment of criteria.

Validity of Evaluation. The validity of an evaluation is not the sum of the validity of its components, but the product of them and some extraneous factors. For an evaluation process to be valid, it must be based on valid objectives. It must be based on data collected in dimensions relevant to these objectives. The evaluation criteria must be realistic, related to data, and determined before the educational activity begins.

Extraneous factors apply more to the conduct of educational activities by individuals and small groups. The assessment aspects of evaluation may be applied in any circumstances, but criteria for satisfaction and dissatisfaction relating to the specific performance of the programs and of individuals conducting them, cannot be derived for dimensions over which the responsible organizations or individuals have no control. In other words, no success criteria can be validly used in evaluating any educator or educational function on the basis of data in a dimension over which he or it does not have control.

EVALUATION OF PERSONNEL

Evaluation of Positions. The requirements for valid evaluation of educational programs have been discussed above. If a particular position carries direct responsibility for the program and carries no other responsibility, evaluation of the program is evaluation of the functioning of that position. Such situations are rare, if they exist at all. They may be approximated in large educational enterprises where marked specialization is possible.

Typically, educational positions have responsibilities for at least some aspects of several programs. Teachers, even those with the most academic assignments, are charged with advising students from time to time, with ordering or advising on the ordering of supplies and equipment, and with helping determine general school policy on the application of district policies to such matters as discipline. Principals are usually charged with curriculum coordination and personnel administration in their schools. They typically have responsibilities in other fields such as community relations,

coordination with other schools, discipline, and advising the central administration. Central administrators frequently wear so many hats that central office responsibility for various programs is diffuse.

Job Definitions. As stated in the section on the validity of evaluation, valid evaluation must be based on valid objectives. Insofar as the entire school system is concerned, it may be held that the objectives adopted by the community through the governing board are, if properly stated, automatically valid. Be this as it may, objectives for various positions are valid only if they clearly relate to the responsibilities assigned to that position. As a *reductio ad absurdum*, it is clear that an objective pertaining to arithmetic computational skills would be invalid for the evaluation of a teacher of English literature. Unfortunately, the relevance of objectives to the evaluation of specific positions is not always so clear.

The general rule for determining the validity of objectives for evaluation of the functioning of a specific position is simple. It is, "Is there a clear, recognized and formal assignment of responsibility for each aspect of the program to this position?" If the aspect is complex, is there a clear determination of which subdivisions of it belong to this position and which do not? The functioning of the position can only be evaluated when this question has been unequivocally answered for all program elements. Naturally, in most cases the answer for some program elements will be "no." Objectives for these elements will, then, have no relevancy for evaluation of the functioning of the position.

Personnel Evaluation. Evaluation of the functioning of a position is necessary for the evaluation of personnel assigned to the position. But, to use a mathematical metaphor, it is not "sufficient." First, the official assignment of personnel must clearly state that this position is one to which the individual is assigned. In the case of full-time positions this may be simple, but there are many positions manned by one or more part-time persons. This is frequently true in counselling and consultant positions.

Second, and of vital importance, no individual can be validly evaluated in a dimension over which he has no control. Teachers cannot be evaluated on the attainment of objectives where text and other materials are required for instruction, if they have no control over the provision of these materials. Teachers cannot be evaluated on an objective relating to student social behavior if they have no control of disciplinary processes, interrelations with parents, and other elements relevant to such an objective. Similarly, principals cannot be evaluated on the basis of their success in getting supplies to teachers if they have no control over the allotment and distribution of supplies.

SUMMARY

Evaluation. To be fair, to be based on due process, evaluation requires establishment of goals and of objectives specific enough to define dimensions on which data can be gathered. Very narrow goals will generally not make possible an evaluation of the program as a whole. Numerical data is not necessary for valid assessment of educational programs. Criteria for success are derived from guesses, informed judgments, and compromise. In no case can they be considered to be absolutes.

Evaluation is the application of agreed-upon criteria to the data collected for dimensions relevant to the objectives adopted for the educational unit being evaluated.

Job descriptions. To apply the results of an evaluation to a given position, it is essential that the responsibilities of that position in relation to the objectives be unequivocally defined. Where individuals holding positions are to be evaluated, the position description must be accompanied by an unequivocal assignment of authority plus an unequivocal statement of other personnel responsible for those areas where no authority is given.

Individual Evaluation. Individual evaluation can validly be accomplished only by referring the evaluation of programs to the positions responsible, and then applying the results only to those factors for which the individual has both responsibility and necessary authority to act. Authority does not only mean authority over personnel, it also includes authority to secure supplies, modify programs, discipline students, and initiate joint activities with other personnel.

STRATEGY DECISIONS IN

EVALUATION PLANNING

Theodore Bass

(Theodore Bass is a senior member of the consultant staff of the California Teachers Association, with office in San Diego. His B.S.E. degree and M.S. degree were received from the University of Arkansas. His Doctorate was completed at the United States International University.)

The mid-point of any logical problem-solving process is the selection from alternatives, a modeled structure of provisionally projected action, or in brief, a choice of strategy. Seiler points up the issue in system analysis:

> Assuming now, that the possible combinations of action choices have been arrayed, all that is left before implementation is choice or decision. Though less laborious than any other stage of the problem-solving process, the decision state may turn out to be the most difficult. It involves three activities: (1) measuring available choices against goals which have been established; (2) assessing the administrator's and the organization's capacity to implement the decision; and (3) actually making the decision by committing one's self to it.[1]

Seiler does not identify this process as strategy, but it is the synthesizing action. It is strategy.

Strategy theory provides, now, a taxonomy of strategy for the use of the education planner.[2] Application of the taxonomy can be proposed for use in preparation to engage in evaluation.

First is <u>timing</u>. The calendar is inexorable. The law itself specifies time boundaries. Reason governs the remaining time options. The collection of data about pupil needs must come at the beginning of the school term. A five-year plan to develop a defensible evaluation program can be recommended. A short term plan to implement the Stull Bill is required.

Second comes involvement. Fait Accompli precludes involvement. Total democracy involves everyone. Representatives can be designated or elected. Again both law and reason require involvement.

Third in order is knowing. Such concepts as criterion referenced tests, professional judgment, normative evaluation, grade levels, are strategic.

Gaps in data are bridged by theorizing. Provisional models of the future provide imagined alternatives.

Spacing considers approach, choice of media, rearrangement of environment.

Finally, spending allocates scarce resources.

A grand strategy is a combination of all of the foregoing. All of us are strategists by description. The use of deliberate prescription in strategy selection characterizes responsible planning. The complexities of evaluation can be reduced by attention to strategy.

---

[1]John A. Seiler, Systems Analysis in Organizational Behavior. Homewood: Richard D. Irwin, Inc. and the Dorsey Press, 1967.

[2]Theodore Bass, A Taxonomy of Strategy For Use in a System Approach to Education. (Unpublished dissertation, United States International University, San Diego).

SYMPOSIUM IV

SPECIAL PROBLEMS IN EVALUATING MINORITY ACHIEVEMENT

Chairman: Dr. Jack M. Thompson
Director of Instruction and Guidance
Sonoma County Schools

Participants: Yvon Johnson
Administrative Assistant
San Francisco Unified School District

Charles Bustamonte
Director, Project Abraza
Santa Clara County Schools

SYMPOSIUM V

DEVELOPMENT OF INDIVIDUALIZED EVALUATION STANDARDS

Chairman: Dr. John C. Gowan
Professor of Education
California State University, Northridge

Participants: Dale Burklund
Director, Guidance and Vocational Education
Santa Clara County Schools

Ezra Wyeth
Professor, Psychological Foundations Department
California State University, Northridge

INDIVIDUALIZED EVALUATION STANDARDS

E. R. Wyeth

(Ezra Wyeth is Professor of Psychological Foundations, California State University, Northridge. His doctorate was earned at the University of California and his earlier academic work was done in Australia.)

As I write the newspaper lies open near me. Prominent on the sports page is the week's ranking of college football teams. Nebraska is No 1 and UCLA is not even mentioned in the top 20. The season has yet to begin.

For days the paper has exuded gloom. The United States is no longer No 1 in Olympic competition. It might even withdraw from future games. There has been an organized conspiracy and the athletes were beaten by politicians and judges not by their opponents.

Occasionally a voice is raised against this childish devotion to the cult of superiority - the adult echo of 'my dad is better than yours'- but the voice is ignored or drowned in the hurt protests of editors who defend their practices by declaring that they print what the public wants.

So with the connivance of mass media, John Doe can rejoice that his team is No 1 and know on Friday the best player in Saturday's game. And this is the John Doe who recites the Pledge of Allegiance and sings with gusto the National Anthem, thereby attesting to his unswerving belief that all men are equal and he knows in his heart that for 200 years his land has been free of those stupidities about class and rank that plagued those countries from which his ancestors came. And he is a free man. In fact there are some 8 million laws on the statute books in his land - if John Tabori and his book The Natural Science of Stupidity are to be believed - to ensure that he will be free.

Unlike Tabori, John Doe is a mass of contradictions. He has a schizoid personality and it needs the skill of the best of psychiatrists to discover the real John. Whether psychiatry can do anything about his condition is another matter.

You and I have seen something of the real John in recent months in the issue of school busing and the Eagleton matter in which John showed his profound belief in the idea that mentally ill people

are possessed of devils and there is just no way in which they can recover from that.

I do not know who suggested the topic for this discussion but he must be a very brave man and one ready to attack sacred cows. There are cold hard facts that must have given him cause to tread warily. You see he wants to attack one of John's deepest and most cherished beliefs that all men are unequal and society will function best when every man knows his place and each is accorded the honors that go with his position. His children must learn this lesson early in life and John will do his best to ensure that schools and universities play their part - not of course that they will offer much opposition.

Let no one try to understate the enormity of the task involved in getting John to give up his need to know where he stands in relation to others. He has to be treated for a long standing obsession and his very way of life has to be threatened.

There are sporadic attempts to give up report cards in schools. In many cases the attempts are met with opposition from parents. They need those cards badly.

But teachers oppose giving up report cards as well. Their needs are threatened. The teacher - and let me include the professor here - holds great power in his hands. He assigns not grades but status and this gives him a sense of power - almost a divine sense. No wonder he oftens refers to himself as 'we'.

What will these pillars of conservatism - the universities - do if students do not produce evidence of status when they seek entry? Mind you universities have not really placed much reliance on the evidence and some of the more notorious ones have shown their belief that the evidence was wrong by failing nearly half the students in their first year. They might even have to follow the path of that heretical institution in New York that does not ask for report cards and the like. Or the path of Junior colleges that show little interest in where a student stands in relation to his peers and somehow have a much lower failure rate than the prestigious universities.

Let me return to where I started. The idea that there should be individual evaluation - if transferred to the field of sport- would require that the athlete measure his performance against what he is capable of doing and winning a gold medal or being No 1 should be incidental to that. Even in losing there should be great satisfaction if he performed as well as he could.

It also asks that millions of avid fans give up screaming that they are No 1 in the land and begin saying something about winning not being what counts but how the game was played.

SYMPOSIUM VI

THE IMPROVEMENT OF LEARNING THROUGH EVALUATION

Chairman: Dr. Carmen J. Finley
Director of Assessment and Principal Research Scientist
American Institutes of Research
Palo Alto

Participants: Eva L. Baker
Assistant Professor of Education
University of California, Los Angeles

Warren Kallenbach
Professor of Education
California State University, San Jose

IMPROVING INSTRUCTION THROUGH EVALUATION:

PRECEPT OR PERFIDY?

Eva L. Baker

(Eva L. Baker is Assistant Professor of Education at the University of California, Los Angeles. She received her B.A., M.A. and Doctorate degrees from the University of California, Los Angeles.)

The intention to use evaluation data as the primary source of instructional improvement is underscored by the growing prominence, if not acceptance, of accountability models. Little specific analysis is evident regarding the feasibility of instructional improvement based on typical sources of pupil achievement data. The best approximation of data based instructional improvement comes from the field of instructional development, where large-scale programs have been designed and revised against a continued influx of pupil performance data.

Data Gathering

After decisions about goals have been made, hopefully employing some of the refined needs assessment procedures presently available, the improver of instructional programs must decide upon measures and procedures for acquiring relevant data. The trade-offs between well known, frequently used achievement tests and newly developed domain-referenced tests must be examined. Of particular interest beyond the quality of information each test type yields, is the differential cost of materials, not only in initial acquisition, but in terms of the training or orientation required of teachers expected to employ the instruments. The development of methods of regular data acquisition and interpretation must be accelerated. Moreover, it is clear that whatever the data collection ventures adopted, teachers will need to be acclimated to the continued need for instructional hiatus in the name of evaluation.

Criterion Setting

To decide in advance what constitutes a deficiency in pupil performance normally requires a sense of history and purpose not consonant with the tradition of achievement measurement. Pre-set criteria, while satisfying the need for order, may not reflect feasible levels for achievement in the schools. The extent to which criterion statements generated for instructional purposes resurface as political goals needs to be both considered and anticipated in planning. Perhaps an alternative to pre-set criterion statements which reflect percentage of achievement figures might be normative statements for individual teachers or schools. Teachers (or schools) would be compared against themselves, with the simple dictum that

significant improvement is required. Another possibility is to require instructional improvement of only those schools or in only those subject areas where performance is noticeably below state levels. Thus, improvement would not be mandated across the board for all subject areas and all schools or teachers. Individual contractual arrangements would be made according to differential need.

Instructional Improvement

Instructional improvement cycles, even in the presence of interpretable data and adequately defined criteria, require a series of complex decisions for the teacher. Of necessity the teacher must make a judgment regarding which instructional activities are likely at fault for poor pupil performance. Second, the teacher must analyze acceptable alternatives to prior inadequate instruction. Third, the teacher must be able to implement suitable alternatives. These steps demand a large training investment and are additionally hampered by the lack of hard knowledge in the instructional area itself.

Certain guidelines can be roughly delineated for instructional improvement strategies, however. They are based on a two-pronged criterion: 1) to what extent is there work in the research literature which supports the value of the intervention and 2) what are the cost considerations of implementation. Instructional interventions can be conveniently scaled in terms of this dual consideration. Teachers can be taught to analyze their activities so that research-based instructional principles are employed. While training must precede such analyses, the actual cost is low of implementing in a classroom principles such as practice, knowledge of results, certain motivational interventions, and other instructional principles with considerable support in the scholarly literature. The second type of instructional activity change centers about the method of presentation for instruction. For instance, should phonics or linguistic approaches be employed in the teaching of beginning reading? The research-base comparing methods is less strong than for instructional variables, for conflicted findings is usually the case attributable to poor experimental design and controls. In addition, the cost of renovating instructional programs to conform to a particular method is high.

Only where the method alternative is encompassed in a validated instructional program are method alterations likely to pay off in the classroom. The reason for such pessimism derives from the single unquestioned finding in educational research: main effects are rare. Instead, methods are likely to be contingently effective and their utility will depend upon the nature of the learner, the style of the teacher and exigencies of the instructional setting.

Thus the flaw in mandated instructional improvement is the expectation that knowledge about such contingent relationships is available. While the conduct of serious experimental research testing controlled variations would seem to be implied, the question of replicable treatment is most likely to be unsatisfactorily considered in the domain of instruction. Until a sufficient number of validated programs have been developed and implemented which include replicable characteristics, the outlook for painless instructional improvement is grim.

IMPROVING EDUCATION THROUGH EVALUATION

Or

WHO'S GOT DESIGNS ON EDUCATION?

Warren Kallenbach

(Warren Kallenbach is Professor of Education at the California State University, San Jose. He received his A.B. degree at Drury College and both his M.A. and Doctorate degrees at Stanford University.)

Improving education through evaluation? Can we improve education through anything? If we accept the Coleman/Jencks thesis perhaps we should look for little more in schools than "whether the teachers and students find it a satisfying place to be."[1] Jencks found little in Coleman's data to encourage him to believe that the schools can contribute significantly to even adult well-being. He does feel that school reforms do improve the lives of children, however.

Why do we want to improve education anyway?

Because of the Stull act?
Because of "What we're s'posed to do"?
Because we want to do good?
Because we have a Title I, II or III grant?

---

[1] Christopher Jencks et al. Inequality: A Reassessment of the Effect of Family and Schooling in America. New York: Basic Books, 1972, passim.

Should we be talking about changing behavior instead of improving education? If so, whose behavior, for how long, in what direction and for what reasons? Does evaluation or the feedback from evaluation make a real difference to the evaluatee? In other words can we really improve education through research, development and dissemination efforts?

The federal government is getting ready to put some heavy money on just this assumption. It's getting ready "to spend billions where spending only millions has produced disappointments."[2] The National Institute of Education is gearing up, amidst some HEW strife, in hopes of making some if not many significant changes in education. It is assumed there will be many approaches proposed to and carried out by and for the NIE. Most will surely include systems approaches and use accountability procedures.

What help can we expect from currently available evaluation designs? Well, we have just about buried the once the widely-popular do-it-yourself model, namely, get some money to try something, tack on some easily-available standardized tests or get some equally easily-available evaluation expert--after we commence our project, of course. Some more desirable models are available, specifically, those of Stufflebeam, Guba, EPIC, CSE, or Provus, respectively, to name several of the most productive ones. If we accept the thesis that the major purpose of evaluation is to provide "useful information for judging decision alternatives,"[3] the above designs can be most useful to us.

There are some research and development centers, some regional educational laboratories, and some ESEA Title III projects that can provide exemplary models for change. There are some others that do not. Some programs or projects that have yielded evidence of changes in learner behavior are (1) the Hawaii English Project, (2) the Far West Laboratory for Educational Research and Development minicourses, (3) the Individualized Mathematics System, (4) the University of Wisconsin R and D Center IGE/MUS-E Project, and (5) the Southwest Educational Regional Laboratory Reading Programs. These programs were carefully designed and carried out.

---

[2] James Brann. "NIE: New Life or Research," Saturday Review of Education, September 16, 1972, Volume IV, p. 43.

[3] Daniel L. Stufflebeam et al. Educational Evaluation and Decision Making. Itasca: F.E. Peacock Publishers, Inc., 1971.

Their evaluative data indicate positive significant changes in learner behavior following application and dissemination of the project treatment procedures. Keep in mind though that the most widely-disseminated current program in education--the School Mathematics Study Group --" never had any formal design for development and dissemination." [4]

What model do we follow for instructional improvement? A highly-structured one? A loosely-organized one? The section by Eva Baker offers some explict recommendations on both these and other approaches.

---

[4] Personal Communication to the Author, September 1972.

SYMPOSIUM VII

COMPETENCY BASED CURRICULUM

Chairman: Dr. William Watts
Associate Professor
Department of Education
University of California, Berkeley

Participants: Newton Metfessel
Professor of Education
University of Southern California

Leo P. Ruth
Supervisor of Teacher Education
and English
Department of Education
University of California, Berkeley

BEHAVIORAL OBJECTIVES AND THE DANGERS OF SYSTEMTHINK IN EDUCATION

Leo Ruth

(Leo Ruth is Supervisor of Secondary Education (English-Reading), School of Education, University of California, Berkeley. He received his B.A. degree from Chico State College and his M.A. degree from the University of California at Berkeley.)

A new style of planning and problem solving, heralded as the way to salvation from error in managing Big Education, is sweeping the nation. Business and governmental agents are pressing "systems analysis" (including behavioral objectives) upon educators, using the force of money and legislation to win quick, unquestioning commitment to various forms of so-called rational decision making, still largely untested in education. (Significantly, through adoption of ACR No. 98 [the Dunlap-Vasconcellos Resolution on PPBS] in Summer 1972, the California Legislature halted the pell-mell rush to install PPBS in California until further hearings are concluded on the matter.)

In California, the loss of a humanistic orientation in educational planning was forecast as early as 1964 in the "Little Report" (THE EMERGING REQUIREMENTS FOR EFFECTIVE LEADERSHIP FOR CALIFORNIA EDUCATION):

> We were struck by the rate at which pedagogy is becoming transformed into educational technology. Increasingly, the design of educational programs involves an integration of facilities, personnel, and processes. Thus pedagogical considerations are affecting the way schools are designed and the way teachers, materials, and pupils are "programmed." (Italics added) (p. 15)

Of course, this unwitting (?) failure to distinguish teachers from materials as "operating units" within a design restricts the freedom and responsibility of the teacher. The "new utopians" of education in their search for planning strategies to shape a world free from human imperfections are promoting mechanical planning models which diminish the scope of human responsibility within operating structure ostensibly concerned with people, but actually concerned with people-substitutes--computer hardware, system procedures, programming, automated tasks, operating unit approaches.

Currently, much educational planning is gripped by a kind of "prosaic mentality" which promotes a dehumanized, "teacher-proof" "learning package" and "management by objectives." This "systemthinker" is attracted to mechanical processes, and to the facts. Facts are the abstractions which can be assigned numbers. Especially "real" in this 1984 world are the events that are quantifiable, measurable. The prosaic mind likes things sharply defined, praises objectivity, practices method, loses itself in technique. Though the systemthinker uses the language of science, he understands little of science, perpetrating fallacies and rigidities of method that real scientists renounce.

The burden of this paper is to demonstrate the specific dangers of this type of systemthink in education. Some of the most evident dangers are these:

1. Reality and truth get replaced by systems models, which then assume a reality apart from the actual world they purport to reflect.
2. Simplifications inherent in even the most sophisticated models have reductionist effects upon conceptions of knowledge, learning, and the learner.
3. An image of man as a reactive being rather than a creative being dominates these mechanistic systems models.
4. Related to this robot image of man is a simplistic conception of teaching/learning which evaluates its "output" using such criteria as the following: speed of response; number of errors or trials; number of animals (men) responding; etc.
5. Preoccupation with planning technique fosters acceptance of illusory simplicities and generates confidence in mere changes of structure as the way to improve operation.
6. Techniques of formulating "behavioral" objectives related to system goals are specific (and rigid), but they offer no basis for distinguishing meaningful and meaningless substance. Indefinite numbers of objectives can be generated that meet systems goals, but they may be trivial, irrelevant, or even destructive of true understanding in a field of learning.
7. Systems procedures inaugurate, strengthen, and perpetuate a strong centralizing bias in the allocation of power for decision making.
8. Systems procedures may very often serve chiefly political ends of reallocation of power under the guise of bringing about "improvement," but usually such procedures perpetuate existing states of knowledge, existing values, established rather than emergent situations.
9. Systems models in education tend to be efficiency oriented rather than humanitarian in aim.
10. This desire to be efficient, precise, and predictable restricts productiveness and creative, unexpected response. How do outcomes not preconceived by the systems designer take place?
11. In view of the lack of any theory to predict the consequences of widespread application of a PPB System to all sectors of government, or of any great body of research, current incautious promotion reflects irresponsible endorsement of as yet incompletely evaluated systems approaches in education.
12. The frequent disallowance of alternatives to systems approaches to organization of education constitutes a contradiction of a first principle of systems thinking, evaluation of alternatives.
13. The lack of any data base whatsoever for such widespread imposition of the PPB System upon education reflects disregard of another basic principle of rational decision making, that of basing analysis, planning, and decision making on data.
14. The current hard sell of PPBS in education consequently is a matter of ideology resting largely on faith.
15. The power of these monolithic conceptual schemes to structure people and events in education reduces individuals to "rule-followers" and "role-players" in the system, denying both pupils and teachers freedom to

explore and express their own purposes. These systemthink procedures in their worst forms constitute methods of manipulation and control which bring ever closer George Orwell's "science-fiction" anti-humane world of 1984.

SYMPOSIUM VIII

EVALUATION OF ADMINISTRATIVE AND SUPERVISORY PERSONNEL

Chairman: Ernest A. Poore
County Superintendent of Schools
Fresno

Participants: John N. Davis
Administrative Assistant
San Jose Unified School District

Robert A. Hansen
Administrative Assistant
Fresno Unified School District

EVALUATION OF ADMINISTRATIVE & SUPERVISORY PERSONNEL

Robert A. Hansen

(Robert A. Hansen is Administrative Assistant to the Superintendent, Fresno Unified School District in Fresno, California. He received his B.S. degree from Drake University and his M.S. degree from the University of Southern California.)

Evaluation of personnel, as a process, can only be defended if the ultimate goal is to "improve" the evaluatee. The entire concept of improvement requires the identification of at least what is better, and preferably what is ideal.

In education, historically we have been guilty of expending vast amounts of resources to improve teachers, in most cases without even making an attempt to identify what was a better teacher. We have been willing to have extensive inservice training classes; we have been willing to provide opportunities for teachers to take special courses at the local university; we have had released time, we have had supervisors and coordinators visit the classrooms; we have given much advice and training; but we have now matured to the extent that we are willing to at least identify for good, for bad, for weak, or for strong, what we feel should be required for any job. Of course, this is the key to the whole process of evaluation-- we can only evaluate a person's effectiveness if we have clearly identified what is required of him.

The simplest form of evaluation, then, of administrative and supervisory personnel, is to consider their capabilities as they relate to their job description. Clearly then it is required, at the very least, to have for every position in the school district a recently reviewed and adopted job description. It can be as detailed or as generalized as the district desires it to be; however, it must be recognized that if it is going to be used for evaluation, then it should be adequately defined and explicit enough to provide for a comparison between what is and what should be. Just using the job description would be a minimum evaluation.

The Stull Bill provides all of us with a charge to evaluate effectiveness as it relates to our real product, the student. When we consider the evaluation of supervisory and administrative personnel in the school district, it follows that we should use, rather we must use, the minimum requirements of the Stull Bill. Article 5.5, Section 13487 of the Education Code says: "The governing board of each school district shall develop and adopt specific evaluation and assessment guidelines which shall include but shall not necessarily be limited in content to the following elements: (a) The establishment of standards of expected student progress

in each area of study and of techniques for the assessment of that progress. (b) Assessment of certificated personnel competence as it relates to the established standards. (c) Assessment of other duties normally required to be performed by certificated employees as an adjunct to their regular assignments. (d) The establishment of procedures and techniques for ascertaining that the certificated employee is maintaining proper control and is preserving a suitable learning environment."

One form that we are trying this year in Fresno with supervisory personnel is to use a process which provides that initially, in the fall, each evaluatee takes a form and fills in his own job description in a manner which could be construed to be behavioral objectives. He also fills in his personal goals and objectives professionally for the year, and then, in addition to the job description, if there are some unique objectives or assignments that he has been given, they are also filled in. This is done by the evaluatee. The evaluatee and the evaluator discuss what he has put on the form, and the evaluator either agrees, amends, or deletes any of the items and so signifies on the form. During that initial interview the development of the objectives might require that the evaluator, during the school year, make some observations, in effect monitor some of the activities that have been purported to be accomplished. If so, the form provides that he will note when he has made the observations.

At the close of the school year the two meet again, but prior to the meeting the evaluatee does a self-evaluation of each of the items on a simple ranking scale. He then brings the form to a meeting with the evaluator. Together they look over the original set of objectives, any monitoring that has been done by the evaluator, the self-evaluation of the evaluatee, and then the evaluator marks on the form his own evaluation of each of the objectives. If any of the objectives require improvement, he marks on the form the kind of improvement expected and the kinds of help that will be given. The completed form is signed by both parties; the evaluatee retains one, one is placed in his file and one is retained by the evaluator.

Of course, the process of evaluation is as old as time. We do it all day, every day. We hire people; we fire people; we reprimand people; and we commend people daily based on our subjective evaluation. It seems to me the only difference here is that we are ordering our process to the extent that each person will clearly understand under what conditions he is being evaluated, what is being required of him, what he will be held accountable for, and then we are providing at least the opportunity for clear discussion of the discrepancy between his demonstrated performance and his expected performance. Also, the district is required to help him "close the gap." Rightly used, implementing the design of the law (Stull Bill) will: (a) provide the public with the knowledge that "education" is holding personnel accountable; (b) assure improved performance of all levels of supervision; and (c) provide internal security to the individual, knowing that his job is defined and he will be helped if needed.

Now let's make it work.

SYMPOSIUM IX

STUDENTS EVALUATING THE EDUCATIONAL SYSTEM

Chairman: Dr. Daniel Freudenthal
Berkeley Unified School District

Participants: Beth Barmack
Teaching Assistant
East Campus, Berkeley High School

Victor Ichioka
Teaching Assistant
East Campus, Berkeley High School

STUDENTS AND TEACHERS, PARTNERS IN SELF-EVALUATION

Beth Barmack

(Teacher, student, poet, leader in women's group work.)

Much of this conference is devoted to the "hard data" of evaluation and testing. Great stock is placed in the vast and remunerative achievement testing enterprise.

Little weight is given here or is often given to the role of students and teachers in evaluating themselves, each other, and the educational system of which they are the critical parts. We submit that great weight should be put upon their self-assessment; for in the classroom, the student is the consumer, and the teacher is the prime mover.

Today, public education is in the doldrums, the butt of vast criticism. From all quarters come cries for accountability, year's growth for a year in school, make 'em learn. Into this caldron is thrown the million dollar question of teacher evaluation, as if this were a simple matter of numbers, pluses, minuses, statistics, manipulated with utter sophistication, interpreted with utter naivete.

Therefore, as a panel, we must ask such critical questions as:

1. What is being evaluated anyway?
2. What is to be learned?
3. Can we make 'em learn?
4. Make 'em learn for what?
5. Finally, what about teachers and teaching?
   Students and learning?

Today, your chairman has brought with him two uniquely equipped and successful teachers of the "newer" and "freer" school, who are indeed with their students. Both have urgent messages which point up the student's role in teacher evaluation, evaluation of the whole educational system, not to mention the social and political system of which education is an instrument. At the same time they point up the teacher's simultaneous role in the whole process of evaluation. They focus on a partnership in self-evaluation.

STUDENTS AND TEACHERS, PARTNERS IN SELF-EVALUATION

Victor Ichioka

(Artist, teacher, student)

I came to this conference mainly to share questions about our school system. My main experience with public schools has been going through them. I am not a professional, and, perhaps, that may have its advantages. At any rate, I am still more a student than a teacher. I intend to look at schools from that point of view.

I think any inquiry into the worth of evaluation finally ends up with questions about the worth of the school system itself. I think questions like: What is taught? What is learned? Who benefits from the schools? These underlie the whole concept of evaluation. These and other such questions should be raised and talked about.

I think also that any discussion of evaluation has to include a look into exactly what controls a student and his family have over his educational destiny. What kinds of choices does a student have? Are the schools serving the needs of the student and his family? His community? Who determines the student's needs? Who determines whether these needs are fulfilled?

I have no answers. But I believe that if we are to investigate the value of schooling, we should ask the most basic questions we know how.

Here, then, are the makings of lively and deep discussion, calculated to raise basic questions on teacher evaluation.

SYMPOSIUM X

COST EFFECTIVENESS

Chairman: Dr. Wallace R. Muelder
Associate Dean
School of Education
University of Southern California

Participants: John Stallings, Chairman
Department of Educational Administration
University of Southern California

Richard B. Horne
Assistant Superintendent
Business and Financial Services
Los Angeles County Schools

COST EFFECTIVENESS IN EDUCATION

John W. Stallings

(John W. Stallings is Professor of Educational Administration at the University of Southern California, Los Angeles. He received his B.S., M.S., and Doctorate degrees from the same university.)

Many an educator, facing problems old and new, must agree often with the advice offered by Machiavelli nearly 50 years ago, "There is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things." A new order of things is needed if educators are to respond to the demands of the legislature and of taxpayers for accountable programs. And leaders are a necessity if the California system is to take this new order in hand. A new order of things in this instance must include a decision making process for public education based on analysis of alternative approaches to resource allocation in terms of cost effectiveness. This analysis reflects an aspiration shared by all educators: the ability to relate the outputs of schools--the educational product--with the inputs, measured in terms of financial resources, human resources, time and material resources.

Elements of Cost Effectiveness. What are the elements necessary for the application of cost effectiveness in education: First, it is necessary to find a workable definition for cost effectiveness.

Cost effectiveness analysis was defined in the Planning, Programming, Budgeting System Manual for California School Districts as a method of determining the most efficient mix of activities to achieve a specific objective. Total costs are related to effects. Costs are measured in dollars, and effectiveness is expressed in terms other than dollars.

Cost effectiveness differs from cost benefit analysis. Cost benefit analysis seeks to establish a ratio of costs to benefits when both costs and benefits are measured in monetary terms. The inherent difficulty of placing a dollar value on educational outcomes makes cost benefit analysis extremely difficult and useless for practical purposes in school administration. Cost effectiveness analysis, however, does offer some hope for potential value

in helping the decision maker arrive at the best answer when various alternatives are under consideration.

One of the initial requirements for gaining significant cost effective comparisons is to hold constant either the objectives or the cost. If you were given the task of buying the best automobile at the cheapest price, you would have an impossible assignment. Your fulfillment of the job could vary from locating a car whose owner would give it to you free if you would tow it away to finding an excellent luxury limousine that sells for tens of thousands of dollars. The free car might satisfy the "cheapest price" portion of the assignment (unless you could find someone that would pay you to remove a junk car) and the luxury limousine might satisfy the "best automobile" part of the project.

It is essential to fix or hold constant one of the variables. The assignment then becomes "Purchase the best automobile you can for less than $2,000" (cost is constant), or "Purchase the best 1932 four-door Model A Ford regardless of the cost" (object is constant).

In education, more and more districts are fixing the educational objectives. Limited, specific, measurable objectives are established for each educational program. The objectives may be 50 per cent of the third grade pupils shall increase reading achievement by one year. . .; or 80 per cent of the eighth grade pupils shall score 75 per cent or higher in mathematics. . . etc.

By holding the objective constant, different methods of achieving the objectives can be compared in terms of cost. When one method is less expensive than other methods and produces equal or better results, a strong case exists for utilizing the least expensive method. Likewise, when the cost is held constant and unequal educational results are obtained, the alternative producing the greatest educational outcome deserves first consideration.

Utilizing analysis of this nature to compare alternative methods of achieving educational objectives is cost effectiveness analysis. In concept, it is deceptively simple; in practice, data collection for accurate analysis can be quite difficult.

Measurement. Another major component essential to an operative cost effectiveness analysis is measurement. At least four key steps are important in evaluation:

1. Determining what is to be accomplished;
2. Agreement on the degree of accomplishment;
3. Concurrence on the measuring instruments and methods; and
4. Deciding on how to analyze the results.

Much has been said and written by professional educators about the inadequacy of academic achievement tests. While agreeing whole-heartedly to the short comings, we should none the less recognize that achievement testing is still a valuable source of data and should be one of the devices

utilized to measure school outputs. Other measures such as teacher-made tests, opinion scales, check lists and questionnaires also should be used. A functional management system utilizing performance objectives requires that the measuring device be identified along with a statement of anticipated accomplishments.

Obviously, there are some problems related to measurement. The productivity of a school or a district cannot be quantified completely on the basis of test scores. Many non-cognitive segments of the educational program will likely remain unverified. Behavior of pupils, attitudes of students, school's social atmosphere, and a host of other areas may depend upon professional judgments of staff for evaluation.

Further, the tests themselves lack perfection. Their content may not parallel the curriculum or they may be unfair to the students because of the students' language or cultural background. Despite these acknowledged deficiences, since measurement is essential, educators must use what data is available as a starting point and work to improve the measuring devices.

Nor is cost effectiveness analysis designed to replace the decision maker. Value judgments still must be made. Cost effectiveness simply provides more basis, more data for decisions.

Program Costs. A further requirement of cost effectiveness analysis in education is program costs.

In the past, school district budgets have been developed around budget categories according to functions-administrative, instruction, health services, etc. Changes from one budget to the next were generally incremental, reflecting pupil enrollment increases or decreases, cost of living fluctuations, revenue changes, and political appeal. The present system of school budgeting (PPBS) requires allocation of resources by program--reading, mathematics, English, etc. Budgeting and accounting by program is necessary for cost effectiveness analysis.

Program budgeting and accounting would need to be done uniformly throughout the state if comparisons are to be made between school districts. And a uniform school accounting system should be maintained throughout the nation to draw comparisons between and among the states.

Thus, it is desirable that a uniform school budgeting and accounting practice be utilized in the fifty states of the U. S. and that this sytem be on the basis of program budgeting and accounting.

Variable Elements. The last component of cost effectiveness analysis is a set of variable elements. In fact, practically all inputs into the analysis are variable elements. Certainly students differ from one group to the

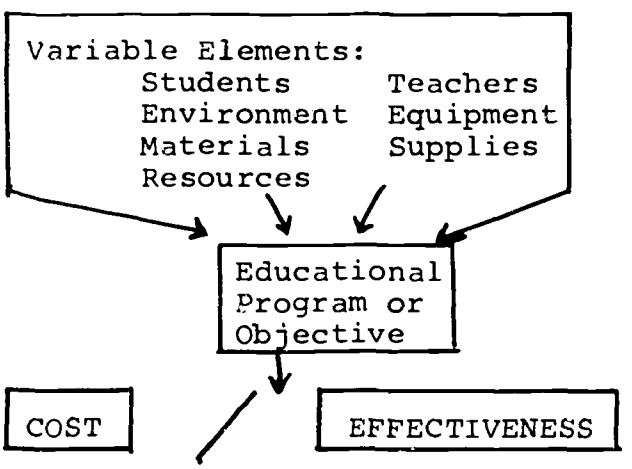
next; teachers vary in ability and skills; resources are allocated in various ways; and the learning environment changes.

Many of the changeable elements can be determined by decisions. The choice may be made to buy new desks for the room instead of textbooks. One teacher may have 35 pupils while another teacher has 25 pupils with a teacher aide. One school building may be old and delapidated, perhaps ex-army barracks, while an adjoining building is new with carpeting, air conditioning, and all modern design. There is no reasonable limit to the number of possibilities for variations that could exist in the classrooms of a district.

Cost effectiveness analysis simply takes whatever variation that exists and utilizes the program cost established for that situation and relates that cost to the effectiveness achieved. Table I depicts this process.

TABLE I

COST EFFECTIVENESS ANALYSIS



Cost Effectiveness Relationship. An operative system as has been discussed here would provide information for performing effectiveness analysis. The relationship between cost and effectiveness may be very simply shown by using a formula such as

$$\text{Cost Effectiveness} = \frac{C}{I\ (N)\ (P_2 - P_1)}$$

where C is Cost of the objective; I is the importance of the objective rated on a scale of 0 to 1.00; N is the number of pupils; P_1 is the per cent of pupils achieving the objective before instruction and P_2 the per cent of pupils achieving the objective after instruction.

For example, suppose the following information were known about two first grade classes whose reading objective was to read at second grade level on a standardized achievement test by the end of the year:

A. Lincoln School using Reading Program X
28 pupils in class
2 pupils reached objective pre-test
14 pupils reached objective post-test
$1,000 cost
Reading judged .95 importance

B. Washington School using Reading Program Y
24 pupils in class
1 pupil reached objective pre-test
16 pupils reached objective post-test
$1,500 cost
Reading judged .98 importance

To compute the cost effectiveness for each school for the first grade reading programs, substitute the data measured and determined into the formula:

A. Lincoln School-Program X

$$\text{Cost Effectiveness} = \frac{\$1,000}{.95\ (28)\ (.50-.07)}$$

$$CE = \frac{\$1,000}{11.438} \qquad CE = \$87.43$$

B. Washington School-Program Y

$$\text{Cost Effectiveness} = \frac{\$1,500}{.98\ (24)\ (.67-.04)}$$

$$CE = \frac{\$1,500}{14.818} \qquad CE = \$101.23$$

The analysis shows that Lincoln School using reading program X is better in terms of cost effectiveness. Program Y at Washington School costs more per unit of reading accomplishment.

A study of the cost effectiveness analysis shows that the variable elements (student population, teachers, etc.) were converted to: 1) costs of reading program by the budgeting and accounting system; and 2) effectiveness of reading program by holding the objective constant and measuring results. The final step establishes a relationship between the cost and the effectiveness of the first grade reading.

Cost effectiveness analysis may be used in a variety of ways to help decision makers. In the example, different reading programs were compared. The analysis may be

used to compare the effectiveness of the same programs for different groups of pupils--minority pupils with anglo; or handicapped pupils with regular pupils. The analysis can be used to compare one district with another; one school with another; one grade level with another; or one classroom with another. Cost effectiveness can be utilized with support programs such as transportation and cafeteria as well as with instructional programs.

Cost effectiveness analysis is one more method to provide additional data to the educator for effective decision making.

STATUS OF COST EFFECTIVENESS AS A METHOD OF ANLAYSIS IN EDUCATION

Richard B. Horne

(Richard B. Horne is Assistant Superintendent-Business and Financial Services, Office of Los Angeles County Superintendent of Schools. He received his A. B. degree from the University of California, Los Angeles, and his Doctorate from the University of Southern California.)

For purposes of illustrating the status of the concept of cost effectiveness as a method of analysis for aiding decision-making, let us look at the two words--COST EFFECTIVENESS--separately. Take the word cost and imagine a four point rating scale with the letter "C" being the lowest point of achievement and "T" the highest point of achievement.

C / O S T

Obviously, there is no objective way of determining exactly where we are along the continuum of developing widely accepted and used methods of determining the real cost of educational programs. However, in my opinion, we are somewhere between the "C" and the "O" on our hypothetical scale. Someone might quickly reply, "you mean to tell me we are only one-quarter way there in our ability to determine the cost of an educational program. After all, it is only a matter of adding up the dollars of each program by the objects of accounting (salaries, supplies, equipment, etc.)." I wish it were that simple--my answer is yes and that estimate may be overly optimistic.

Cost is a measure of the resources that go into an educational program. An effective cost system must have at least three essential elements.

(1) A program oriented budgeting and accounting system using educational programs as the primary element for developing a budget and determining costs.

(2) A program budgeting and accounting system uniformly applied and administered by qualified personnel at all levels of the educational organizational structure (district, intermediate, state, and federal.)

(3) An electronic data processing system to facilitate the timely and accurate collection of data with the capability of manipulating the data to produce meaningful information.

Let us look at the California scene as a case in point. California, like most states, still has a budgeting and accounting system based upon the functions of education (administration, instruction, food services, transportation, etc.) not programs. A uniform program budgeting and cost accounting system has been developed and was scheduled for field testing during 1972-73 and operational status during 1973-74. However, it has become embroiled in the politics of the legislature and the State Board of Education and has been delayed at least a year. This represents valuable time lost because any effective cost accounting system applied on a uniform statewide basis needs two or three years of operation to work out the problems. In the meantime, at the local level some school districts are experimenting with program budgeting and accounting systems, but the practice is not widespread. The competence of the personnel involved in the various districts varies and considerable upgrading is necessary, not to mention the additional problems associated with far too many small administratively inefficient school districts.

Coming back to my basic point--we are only one-quarter of the way toward developing and implementing an effective cost system.

Assuming we had a viable cost system implemented, what about the other word effectiveness? I am sorry to report the degree of progress is even less. If we take the word effectiveness and treat it in the same manner as cost and imagine a thirteen point rating scale with the first letter "E" being the lowest point of achievement, in my opinion we are somewhere between the two "F's". In other words, only two thirteenths of the way along a thirteen point scale.

E F / F E C T I V E N E S S

Adding this "effectiveness" dimension to our method of analysis really complicates the concept far beyond the complexity of a cost system. Whereas cost is a measure of the resources that go into an educational program, effectiveness measures what comes out or the result--to what extent were the program objectives achieved. Even if we develop a successful method of cost analysis, it will provide barren data rather than meaningful information for decision-making. We shall be able to say with a high degree of confidence that a particular educational program has a cost of blank dollars, but we shall not be very sure how effective the program is in terms of educational performance.

Educational research has an important role on the effectiveness side of our two word method of analysis. An all out effort should be made toward developing instruments and methods for measuring educational performance. Since at the present time we are only spending about three tenths of one per cent on educational research and development, I would like to support an immediate substantial increase in the amount of money spent for that purpose even at the expense of curtailing other programs. In my opinion, a high level strategic decision should be made in that direction and the long term results will justify the higher priority and cost.

E V A L U A T I O N

A B I B L I O G R A P H Y

Prepared by the Research Library, California Teachers Association

I GENERAL REFERENCES

1. Alkin, Marvin C. "Evaluation theory development," UCLA Evaluation Comment, 2:2-7, Oct. 1969.

2. Allen, James E., Jr. "How is education being held accountable?" College Board Review, #82, pages 8-9+. Winter 1971-72.

3. Association of California School Administrators. Thrust for Education Leadership, v.1, May 1972.
   Among the contributors are Henry S. Dyer, John Vasconcellos, Wilson Riles.

4. Browder, Lesley H., Jr., ed. Emerging Patterns of Administrative Accountability. McCutchan Publishing Co., Berkeley, Cal., 1971. 571 pages.
   Readings on the foundations, politics, economics and applications of accountability.

5. California. Joint Committee on Educational Goals and Evaluation. Education for the People. 2 vols. California State Dept of Education, Sacramento, Cal., 1972.
   "Guidelines for total community participation in forming and strengthening the future of public elementary and secondary education in California."

6. California. Joint Committee on Educational Goals and Evaluation. The Way to Relevance and Accountability in Education. The Committee, Sacramento, Cal., 1970. 35 pages.
   Recommends the means for identifying educational goals and objectives and the means for developing a statewide plan of assessment and evaluation.

7. Dyer, Henry S. "School evaluation: a realistic response to accountability," North Central Association Quarterly, 46:390-396, Spring 1972.
   Some of the pitfalls and fallacies of evaluation as it has been practiced, and how evaluation can be a "rational and realistic response" to

accountability with the help of accrediting bodies such as the North Central Association.

8. Dyer, Henry S. and Elsa Rosenthal. State Educational Assessment Programs, An Overview. TM Report No. 6, ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, N. J., 08540. Dec. 1971.

9. Education Commission of the States. Compact, v.6, #1, Feb. 1972.
Entire issue on National Assessment and measuring American education. Some of the contributors are Ralph W. Tyler, George B. Brain, Carmen J. Finley, and Frank B. Womer.

10. "Educational Evaluation" is the theme of Review of Educational Research, v.40, April 1970. 320 pages.
"Objectives, priorities, and other judgment data," by Robert E. Stake. "Politics and research: evaluation of social action programs in education," by David K. Cohen. "Curriculum evaluation," by Ian Westbury. "Values, goals, public policy and educational evaluation," by Harold Berlak. "Evaluation of instruction," by Barak Rosenshine. "Measurement techniques in evaluation," by Douglas D. Sjogren. Bibliography with each article.

11. Educational Testing Service. State Educational Assessment Programs. ETS, Princeton, N. J., 1971. 83 pages.
Detailed information about educational assessment programs and plans in all of the states.

12. Eisner, Elliot W. "Emerging models for educational evaluation," School Review, 80:573-590, Aug. 1972.
Three ideas are developed which the author believes hold promise for improving the process of educational evaluation---educational objectives, expected outcomes, and evaluation methods.

13. "Evaluation for Administrative Action," theme of Journal of Research and Development in Education, vol. 3, Summer 1970. 108 pages.
Several writers explain what evaluation is and how it is conducted.

14. Georgia. Dept of Education. Div. of Planning, Research and Evaluation. Goals for Education in Georgia. The Dept., Atlanta, Ga., 1970. 48 pages
The Georgia Assessment Project (GAP) is designed "to provide statewide measurement of the progress of Georgia's children and youth toward achievement of those qualities necessary to live successfully in Georgia and the U.S...."

15. House, Ernest R. "The conscience of educational evaluation," Teachers College Record, 73:405-414, Feb. 1972.
Some of the problems involved in evaluation.

16. Indiana University. School of Education Bulletin, Viewpoints, v.48, July 1972.
Four views of formative evaluation in instructional development.

17. Lessinger, Leon M. and Ralph W. Tyler, eds. Accountability in Education. Charles A. Jones Publishing Co., Worthington, Ohio, 1971. 85 pages.
Discusses accountability as seen in the 70's and in historical perspective.

18. NEA. Association for Supervision and Curriculum Development. Evaluation as Feedback and Guide, edited by Fred T. Wilhelms. ASCD, Washington, D.C., 1967. 283 pages.
How to use evaluation as "a positive force toward better teaching, better learning, and a better balanced curriculum."

19. NEA. Association for Supervision and Curriculum Development. Improving Educational Assessment & An Inventory of Measures of Affective Behavior, edited by Walcott H. Beatty. ASCD, Washington, D.C., 1969. 164 pages.
Ralph W. Tyler, Robert E. Stake, Daniel L. Stufflebeam and Walcott H. Beatty write on the improvement of educational assessment.

20. Norton, Robert E. "The relationship between evaluation and accountability," American Vocational Journal, 47:61-65, Feb. 1972.

21. Ornstein, Allan C., ed. 'Teacher accountability," Nation's Schools, 89:45-68, May 1972.
Special report, with articles by educators Robert Havighurst, Scott D. Thompson, David Selden, Mario D. Fantini, W. James Popham, Henry S. Dyer and Hulda Grobman.

22. Palardy, J. Michael and James E. Eisele. "Competency-based education," Clearing House, 46:545-548, May 1972.
CBE assumes detailed descriptions of expected outcomes, provision for individual differences, and provision for opportunities for the learner to pursue his personal goals.

23. Phi Delta Kappa National Study Committee on Evaluation. Daniel L. Stufflebeam, Chairman. Educational Evaluation and Decision Making. F. E. Peacock Publishers, Itasca, Ill., 1971. 368 pages.
Proposes to expose five problem areas (definition, decision making, values and criteria, administrative levels, and the research model), to identify and assess approaches to deal with these, to synthesize a new definition and methodology of evaluation resulting from the assessment, and to provide operational guidelines for implementing the proposed new approach.

24. Roberson, E. Wayne, ed. Educational Accountability Through Evaluation. Educational Technology Publications, Englewood Cliffs, N. J., 1971. 107 pages.
The evaluation scheme presented contains four phases: planning, implementation, product and recycling.

25. Sawin, Enoch I. Evaluation and the Work of the Teacher. Wadsworth Publishing Co., Belmont, Cal., 1969. 298 pages.
Discusses the implications of principles of teaching and learning for evaluation, the foundations for techniques of educational evaluation,

evaluation instruments and procedures, and some uses of evaluation results.

26. U. S. President's Commission on School Finance. How Effective is Schooling? A Critical Review and Synthesis of Research Findings. Rand Corp., Santa Monica, Cal., 1972. 222 pages.
The Commission's answer to the question,"What does the research tell us about educational effectiveness?"

27. Van Deren, Richard H. A Plan for Effective Learning Management. Soquel Union Elementary School District, Capitola, Cal., 1970. 31 pages.
A learner-oriented management process designed to "facilitate responsible instructional innovations and efficient allocation of resources that will result in measurable improvement of learning."

28. Wiley, David E. Design and Analysis of Evaluation Studies, From the Proceedings of the Symposium on Problems in the Evaluation of Instruction, UCLA, CSE Report No. 28. Center for the Study of Evaluation of Instructional Programs, University of California, Los Angeles, May 1969.

29. Wittrock, M.C. "The evaluation of instruction: cause and effect relations in naturalistic data," UCLA Evaluation Comment, 1:1-7, May 1969.

30. Wynne, Edward. Politics of School Accountability: Public Information about Public Schools. McCutchan Publishing Co., Berkeley, Cal., 1972. 291 pages.
Analyzes and forecasts the interactions of accountability and learning measurement, the professionals engaged in education research, schoolmen, the public and public spokesmen, the media, general public expectations about the relation of research and policy, and the students.

II EVALUATION OF TEACHERS AND ADMINISTRATORS

1. Bledsoe, Joseph C., Iva D. Brown, and Arthur D. Strickland. "Factors related to pupil observation reports of teachers and attitudes toward their teacher," Journal of Educational Research, 65:119-126, Nov. 1971.

2. California. Dept of Education. California State Board of Education Guidelines for School Districts to Use in Developing Procedures for Evaluating Certificated Personnel. The Dept., Sacramento, Cal., 1972. 11 pages.
Designed to help districts in the implementation of the Stull Bill.

3. Coats, William D., Lloyd Swierenga, and Jack Wickert. "Student perceptions of teachers--a factor analytic study," Journal of Educational Research, 65:357-360, April 1972.

4. Combs, Arthur W., Donald L. Avila, and William W. Purkey. Helping Relationships: Basic Concepts for the Helping Professions. Allyn and Bacon, Boston, 1971. 300 pages.
Roles and responsibilities of teachers and others in the helping professions.

5. Dempsey, Richard A., and Norman L. Breyer. Staff Development and Evaluation. A. C. Croft, Swarthmore, Pa., 1971. 53 pages.
Assessment based on behavioral specifications.

6. Flanders, Ned A. Teacher Influence, Pupil Attitudes, and Achievement. U.S. Office of Education, Washington, D.C., 1965. 126 pages. (OE-25040 Cooperative Research Monograph No. 12)

7. Goldman, Harvey. "Evaluation of administrative behavior at the building level," National Association of Secondary School Principals Bulletin, 54:70-79, Sept. 1970.
Proposes measures providing opportunities for meaningful evaluation in terms of the specific goals of individual principals, as well as taking into account the limitations to which the principals are subject.

8. Greene, Robert E. Administrative Appraisal: A Step to Improved Leadership. National Association of Secondary School Principals, Washington, D.C., 1972. 48 pages.
Weaknesses in many of the current appraisal systems are pointed out, followed by general guidelines for designing or redesigning a goal-oriented system for a district's particular needs.

9. Jones, Anthony S. "A realistic approach to teacher evaluation," Clearing House, 46:474-481, April 1972.
Includes teacher evaluation forms.

10. Klein, Stephen P. and Marvin C. Alkin. "Evaluating teachers for outcome accountability," UCLA Evaluation Comment, 3:5-11, May 1972.
Describes possible ways to implement teacher-evaluation laws.

11. Lambert, M. Dale. "Refocusing teacher evaluation: a process of guided self-analysis," Thrust for Education Leadership, (ACSA) 1:41-44, Feb. 1972.

12. McNeil, John D. "Concomitants of using behavioral objectives in the assessment of teacher effectiveness," pages 47-53 of Current Research on Instruction, edited by Richard C. Anderson and others. Prentice-Hall, Englewood Cliffs, N. J., 1969. 396 pages.

13. Marshall, Max S. "Reverse grading," Educational Leadership, 28:663-665, March 1971.
The writer feels that student appraisal of teachers will not bring progress or peace.

14. NEA. Research Div. "The evaluatee evaluates the evaluator," ERS

Circular #5, 1970. 52 pages.
A survey of "client-oriented" evaluation.

15. _____. "Evaluating administrative/supervisory performance," ERS Circular #6, 1971. 60 pages.

16. _____. "Evaluating teaching performance," ERS Circular #2, 1972. 60 pages.
Summarizes the results of an Educational Research Service survey.

17. Ober, Richard L., Ernest L. Bentley, and Edith Miller. Systematic Observation of Teaching, An Interaction Analysis-Instructional Strategy Approach. Prentice-Hall, Englewood Cliffs, N. J., 1971. 236 pages.
Accountability, both for the actions of the teacher and the continuous interactions within the classroom, by means of systematic observation.

18. Owens, Mary Seymour. "Evaluation of teaching competence by three groups of educators," Journal of Experimental Education, 40:77-82, Winter 1971.
A study "to compare the perceptions of teacher competence within and among three groups of educators (teachers, college supervisors, and public school administrators) as measured by the interview scales of IOTA."

19. Pi Lambda Theta. The Evaluation of Teaching. A Report of the Second Pi Lambda Theta Catena. Washington, D.C., 1967. 259 pages.
Different criteria for evaluating teaching, including pupil outcomes.

20. Plowman, Paul D. Behavioral Objectives: Teacher Success Through Student Performance. Science Research Associates, Chicago, 1971. 188 pages.

21. Popham, W. James. Designing Teacher Evaluation Systems. Instructional Objectives Exchange, Los Angeles, Cal., Dec. 1971. 49 pages.
"A series of suggestions for establishing teacher assessment procedures as required by the Stull Bill (AB 293), 1971 California Legislature." (Subtitle)

22. Redfern, George B. "Client-centered evaluation," School Administrator, pages 7-10, March 1972.

23. _____. "Evaluating administrative productivity: can it be done?" School Administrator, pages 15-16, July 1971.
The author says "yes" provided certain prerequisites exist.

24. Roberson, E. Wayne. "Teacher self-appraisal: a way to improve instruction," Journal of Teacher Education, 22:469-473, Winter 1971.

25. Rosenshine, Barak. "Enthusiastic teaching: a research review," School Review, 78:499-514, Aug. 1970.
A review of the attempts to assess the relationship of enthusiasm to pupil achievement and to specify the behavioral components of enthusiasm.

26. Rosenshine, Barak, and Norma Furst. "Research on teacher performance criteria," pages 37-72 of Research in Teacher Education: A Symposium, edited by B. Othanel Smith for the American Educational Research Association. Prentice-Hall, Englewood Cliffs, N. J., 1971. 166 pages.

27. Rosenshine, Barak. "Teaching behaviors related to pupil achievement: a review of research," pages 51-98 of Research Into Classroom Processes: Recent Developments and Next Steps, edited by Ian Westbury and Arno A. Bellack, Teachers College Press, N. Y., 1971.

28. _____. Teaching Behaviours and Student Achievement. National Foundation for Educational Research in England and Wales, London, 1971. 229 pages. (Available from Fernhill House Ltd., 303 Park Ave. South, New York, N.Y. 10010).
A summary of more than 50 studies in which some measure of observed teacher behavior was related to one or more measures of student achievement.

29. Schalock, H. Del. "The focus: knowledge, teaching behavior, or the products?" pages 43-49 of Performance-Based Certification of School Personnel, edited by Joel L. Burdin and Margaret T. Reagan. ERIC Clearinghouse on Teacher Education, Washington, D.C., 1971. 140 pages.

30. Simpson, Ray H. Teacher Self-Evaluation. Macmillan, N. Y., 1966. 100 pages.

31. U.S. Office of Education. Do Teachers Make a Difference? A Report on Recent Research on Pupil Achievement. USOE, Washington, D.C., 1970. 181 pages.
Contributors: Alexander M. Mood, James W. Guthrie, Henry M. Levin, Eric Hanushek, George W. Mayeske, Stephen Michelson, Robert M. Gagne, and James S. Coleman. (See also item #32.)

32. _____. How Teachers Make a Difference. USOE, Washington, D.C., 1971. 166 pages.
An assessment of the present state of the art of teaching. Papers by Alexander M. Mood, Philip W. Jackson, N.L. Gage, Dan C. Lortie, Barak Rosenshine, Ned A. Flanders, and Lawrence M. Stolurow. (See also item #31.)

33. Veldman, Donald J. "Pupil evaluation of student teachers and their supervisors," Journal of Teacher Education, 21:165-167, Summer 1970.

34. Washington, Eva. Expert Teacher Action Program. Fearon Publishers, Belmont, Cal., 1971. 113 pages.
Based on a set of standards, 25 variables, that leads to a definition of what constitutes expert teaching.

III MEASUREMENT OF STUDENT ACHIEVEMENT

1. Beall, Lewis L. "Pupil progress: one measure of teaching," California School Boards, 31:10-13, May 1972.
   Includes the hazards as well as the advantages.

2. Block, James H. "Criterion-referenced measurements: Potential," School Review, 79:289-298, Feb. 1971.
   The author examines the "alleged flaws" pointed out by Ebel (see item #6) and defends the use of such measurements because of their potential "to promote the learning of all."

3. Bloom, Benjamin S., J. Thomas Hastings, and George F. Madaus. Handbook on Formative and Summative Evaluation of Learning. McGraw-Hill, N. Y., 1971. 923 pages.
   The intent of this book is to help the teacher bring students up to mastery levels of learning, use evaluation for instruction decisions, understand and use evaluation techniques for cognitive and affective objectives, and cooperate with specialists and others in evaluation. The second part of the book deals with evaluation in each of the major subject fields and levels of education.

4. Bloom, Benjamin S. "Testing cognitive ability and achievement," pages 379-397 of Handbook of Research on Teaching, a project of the American Educational Research Association, edited by N. L. Gage. Rand McNally & Co., Chicago, 1963. 1218 pages.

5. Darlington, Richard B. "Another look at 'cultural fairness'," Journal of Educational Measurement, 8:71-82, Summer 1971.
   Four definitions of "cultural fairness" are examined.

6. Ebel, Robert L. "Criterion-referenced measurements: Limitations," School Review, 79:282-288, Feb. 1971.
   The author states that "the use of criterion-referenced measurements cannot be expected to improve significantly our evaluations of educational achievement." (See also item #2.)

7. Green, Donald Ross. Racial and Ethnic Bias in Test Construction. CTB McGraw-Hill, Del Monte Research Park, Monterey, Cal. Adapted from Racial and Ethnic Bias in Test Construction, Final report of the U.S. Office of Education, Contract No. OEC-9-70-0058 (057).

8. Husek, T. R. "Different kinds of evaluation and their implications for test development," UCLA Evaluation Comment, 2:8-10, Oct. 1969.

9. John, Vera P. and Vivian M. Horner. "Testing and evaluation procedures," pages 142-163 of Early Childhood Bilingual Education, by Vera P. John and Vivian M. Horner, Modern Language Association of America, N. Y., 1971. 187 pages.

10. Johnson, Geraldine F. "Metropolitan tests: inappropriate for ESEA pupils," Integrated Education, 9:22-26, Nov/Dec. 1971.
The author rejects the assertion that Metropolitan Achievement Tests provide valid reading ability measurement of ESEA Title I students.

11. Karmel, Louis J. Measurement and Evaluation in the Schools. Macmillan/Macmillan-Collier, London, 1970. 492 pages.
In addition to the technical aspects of testing and measurement applications, the contemporary issues and problems of testing are discussed, such as IQ, heredity and environment, minority testing.

12. Khatena, Joe. "Some problems in the measurement of creative behavior," Journal of Research and Development in Education, 4:74-82, Spring 1971.

13. Klein, Stephen P. "Evaluating tests in terms of the information they provide," UCLA Evaluation Comment, 2:1-6, June 1970.

14. _____. "The uses and limitations of standardized tests in meeting the demands for accountability," UCLA Evaluation Comment, 2:1-7, Jan. 1971.

15. Mathis, Harold I. "The disadvantaged and the aptitude barrier," Personnel and Guidance Journal, 47:467-472, Jan. 1969.
Aptitude testing is contrasted with achievement testing, particularly in reference to underlying assumptions, validation procedures, and interpretation of test scores.

16. Mayo, Samuel T. "Mastery learning and mastery testing," NCME Measurement in Education, 1:1-4, March 1970. (National Council on Measurement in Education, Special Report series).
Mastery learning can be characterized as a "method of organizing discrete educational objectives which are meaningful and useful to the individual." The importance of mastery testing of mastery learning is discussed in this article.

17. Moreno, Steve. "Problems related to present testing instruments," El Grito, 3:25-29, Spring 1970.
Problems of validity and reliability of English test scores on Spanish-speaking or bilingual children, and of availability of valid tests for Mexican-American children. (El Grito: A Journal of Contemporary Mexican-American Thought. P. O. Box 9275, Berkeley, Cal. 94709)

18. Moxley, Roy A. "A source of disorder in the schools and a way to reduce it: two kinds of tests," Educational Technology, Teacher & Technology Supplement, 10:S3-S6, March 1970.
Norm-referenced and criterion-referenced tests are viewed as to variability, construction, methods of instruction, objectives and the product.

19. Popham, W. James, ed. Criterion-Referenced Measurement: An Introduction. Educational Technology Publications, Englewood Cliffs, N. J., 1971. 108 pages.

Includes "Instructional technology and measurement of learning outcomes: Some questions," by Robert Glaser (Reprinted from American Psychologist, 1963), as well as papers by Popham and Husek (see item #20), Garvin, and Cox.

20. Popham, W. James, and T. R. Husek. "Implications of criterion-referenced measurement," Journal of Educational Measurement, 6:1-9, Spring 1969.
Norm-referenced, and particularly criterion-referenced, measurements are examined with respect to variability, item construction, reliability, validity, item analysis, reporting, and interpretation. (See also item #19.)

21. Shoemaker, David M. "Criterion-referenced measurement revisited," Educational Technology, 11:61-62, March 1971.
The author says it might be more appropriate to refer to "criterion-referenced" tests as "achievement" tests. Their use should be judged according to the amount of useful information they provide the classroom teacher.

22. Spence, Allyn G., and others. "Home language and performance on standardized tests," Elementary School Journal, 71:309-313, March 1971.

23. Stone, Chuck. "Testing and the educational power struggle," Integrated Education, 9:4-9, July/Aug. 1971.
The author, who is a staff member of the Educational Testing Service, discusses the testing experience as it affects the Black community.

24. Thorndike, Robert L. "Concepts of culture-fairness," Journal of Educational Measurement, 8:63-70, Summer 1971.

25. Thresher, B. Alden. "Uses and abuses of scholastic aptitude and achievement tests," pages 24-40 of Barriers to Higher Education, College Entrance Examination Board, N. Y., 1971. 151 pages.

26. Williams, Frank E. "Assessing pupil-teacher behaviors related to a cognitive-affective teaching model," Journal of Research and Development in Education, 4:14-22, Spring 1971.

IV EVALUATION OF CURRICULUM AND EDUCATIONAL PROGRAMS

1. Bloom, Benjamin S., ed. Taxonomy of Educational Objectives. Handbook I: Cognitive Domain. Longmans, N. Y., 1956. 207 pages.
Includes illustrative tests.

2. "Evaluating educational programs: a symposium," Urban Review, 3:14-22, Feb. 1969.
Evaluation of Title I. Contributors: J. Wayne Wrightstone, James S. Coleman, David G. Hawkridge and Albert B. Chalupsky, Henry S. Dyer, John Mann, Martin Mayer, Edward A. Suchman, Peter H. Rossi, Edward Wynne, and Michael Scriven.

3. Forehand, Garlie A. "Curriculum evaluation as decision-making process" Journal of Research and Development in Education, 3:27-37, Summer 1970.

4. Gooler, Dennis D. and Arden D. Grotelueschen. "Curriculum development accountability," Educational Leadership, 29:165-169, Nov. 1971.
The role of evaluation in curriculum development.

5. Harbeck, Mary B. "Instructional objectives in the affective domain," Educational Technology, 10:49-52, Jan. 1970.
The author says usually educational goals in the affective domain have been stated in the school philosophy and have been forgotten in the planning of the actual instructional planning. How to state and measure the outcomes is a major problem in this domain.

6. Krathwohl, David R., Benjamin S. Bloom, and Bertram B. Masia. Taxonomy of Educational Objectives. Handbook II: Affective Domain. David McKay Co., N. Y., 1964. 196 pages.
The classification of educational goals.

7. Mager, Robert F. Preparing Instructional Objectives. Fearon Publishers, Palo Alto, Cal., 1962. 60 pages.

8. Ontario Institute for Studies in Education. Curriculum Theory Network Monograph Supplement, "Curriculum evaluation: potentiality and reality," 8/9, 1971-72. Edited by Joel Weiss. Ontario Institute for Studies in Education, 252 Bloor St., West, Toronto 181, Ontario, Canada. 256 pages.
Evaluation of materials, instruction, and outcomes of curriculum programs.

9. Wight, Albert R. "Beyond behavioral objectives," Educational Technology 12:9-14, July 1972.
Evaluation of programs based on behavioral objectives.

Lois Wing
CTA Research Library
10-5-72

A P P E N D I X

This paper was delivered at last year's conference, the 23rd Annual State Conference on Educational Research, on November 19, 1971. It was not published in the Proceedings (*Research Resume* Number 46), but we are including it at this time because of its appropriateness to this year's theme of *Evaluation*.

ACCOUNTABILITY IN INSTRUCTIONAL

PRODUCT DEVELOPMENT

Harry Handler
Chairman, Department of Instructional Technology
School of Education
University of Southern California

A discussion of accountability in instructional product development will encounter as much if not more ambiguity as have other efforts to apply the concept of accountability to existing activities in the educational community. As Dick Harsh has already indicated: "The fascinating thing about labels is that they are widely used by persons having diverse understandings and attitudes toward the entity which the label represents."

Each member of the audience will interpret the comments which follow in terms of his past experiences and anticipated needs. For example, last week four students were discussing the results of the State Testing Program. One of the students was a sociology major, another an ecology major, the third a journalism major, and the fourth had not selected a major.

> The sociologist: Well, just as we have been predicting, IQ's are decreasing in urban areas.
>
> The ecologist: You see, students who live in smog-filled environments are losing IQ points.
>
> The journalist: Wow! Relationship found between decreasing IQ's and increasing smog.
>
> The non-major: Why be surprised? Anyone who continues to live in a setting where he has to breath the stuff can't be too bright.

In the Rand Publication, Accountability, Program Budgeting, and the California Educational Information System: A Discussion and Proposal, Farquhar identifies the following three areas for which educational personnel are held accountable:

1. Education planners are judged on (a) whether or not the chosen goals are desired by the school board, commun-

ity, and state, and (b) whether or not the programs established effectively meet these objectives.

2. Educational Administrators are judged on their ability to monitor and administer the programs so as to bring about successful outcomes (if objectives and program design are sound).

3. Teaching personnel and related administrators are judged on their ability to bring about the behavioral and educational outcomes stated in the objectives and implicit in the educational program structure.

Accountability in instructional product development is concerned with all three of the above areas. As the staff of the Southwest Regional Laboratory has demonstrated, instructional product development includes:

1. The production of research-based systems to accomplish specified instructional outcomes under natural conditions.

2. The operational use of a program, without further direct assistance from the developer, in all the appropriate classes in a school district.

3. The continued use of a product to consistently obtain the instructional outcomes at a specified level of effectiveness.

This presentation will partially respond to the concerns expressed by Harmes when he raised the following question: "Why should educators, particularly public school educators, be singled out and be asked to guarantee results from instructional programs which are unreliable to varying degrees?"

Four basic subsystems and related evaluation criteria in terms of subsystem objectives required to successfully accomplish specified instructional outcomes are presented. In addition, three distinguishable phases of the development process are reviewed.

The content of this paper draws heavily on the work of the staff of the Southwest Regional Laboratory and will be available in greater detail in the forthcoming publication Instructional Product Development. Baker, Robert L. and Richard E. Schutz, editors. Southwest Regional Laboratory for Educational Research and Development.

Four Subsystems and Related Evaluation Criteria

in Terms of Subsystem Objectives

Instructional systems. An instructional system refers to the research-based methods and materials prepared to accomplish specified instructional outcomes under natural conditions. The instructional system includes everything needed by the schools to obtain specified outcomes. A system, therefore, includes specification of instructional outcomes, student materials, instructor procedures, delivery mechanism for delivering the student material to the student, performance indicators which are used for evaluating student progress, for diagnosing student learning difficulties, and for instructor and instructional system accountability.

Instructional systems evaluation criteria in terms of system objectives:

1. Statement of anticipated observable student outcomes.
2. Criterion measures to determine the accomplishment of the outcomes.
3. Student instructional materials.
4. Statement of learner prerequisites in terms of initial proficiency the learner must exhibit.
5. Statement of the teacher's instructional responsibilities.
6. Evidence that the system has yielded dependable results.
7. Data concerning instructional time and study time requirements.
8. Statement of direct and indirect costs.

It is unreasonable for the developer to expect that a given instructional system will achieve all of the intended objectives during the first field tryout. As Deterline has stated, "We must rely on data. Too many courses are considered to be 'finished, polished, and perfect,' because a group of subject matter experts or instructional specialists pronounced them so. In their best judgment, the course SHOULD teach. An empirical approach requires more than that. The subjective, intuitive and experiential judgments of the experts are interesting, encouraging and nice to have. But the final judges are the students. The data they produce as they proceed through the various components of the course, and the evaluation data, and any available follow-up data tell the story. We are accountable for the data; if the data are not adequate in comparison with the data we had hoped for, we have not met the requirements."

The developer must collect the data essential to provide information related to the system objectives along with data which provide direction for appropriate revisions. Although these data are frequently collected, inadequate planning many times does not allow for the time, resources, and related costs required to conduct the subsequent iterations.

Unfortunately, the impression exists that when the developer has established some arbitrary criterion level of acceptable performance that he is willing to abandon the critical task of effecting learning for those who do not meet the criterion. This is not the position taken by the responsible developer. However, he has learned that there is a direct relationship between the levels of difficulty and mastery desired and the development costs.

In the November-December, 1967 issue of the Harvard Business Review, Abraham Zaleznik discussed the relationship between the individual's ability to cope with disappointment and related success in government and industry. The instructional product developer must also learn to manage disappointment in a constructive manner and to avoid the pitfalls inherent in the use of rationalization, projection, or compensation. If the objectives of an instructional system are not attained, he is not being accountable if he resorts to such statements as: 1) "It doesn't make any difference, the people who are demanding the data won't understand it anyhow"; 2) "The materials are fine, the schools goofed it up"; or 3) "The results may not be too good, but we have learned how to produce the materials cheaper than anyone else".

If the data do not support the objectives, the developer must responsibly analyze the sources of weakness and assume responsibility for the improvement of the product.

Training systems. A training system refers to the materials and procedures required to train persons who are in direct contact with the pupil and have the ultimate responsibility for various phases of instruction. In addition to the teacher, these human resources include parents, tutors, aides, etc. The training system must be designed in such a way as to effectively train all personnel involved in the teaching process in a specified manner consistent with the instructional requirements of the instructional system.

Training systems evaluation criteria in terms of system objectives:

1. Statement of personnel requirements and interrelationships (e.g., teachers, tutors, parents,

aides, etc.).

2. For each personnel category identified, proficiency anticipated at the end of training and in conducting the instructional program.
3. Measurement procedures to determine the accomplishment of the training outcomes.
4. Instructional materials for each category of trainees.
5. Instructional materials for trainers.
6. Feasible alternative plans for sequencing and scheduling training for each trainee category.
7. Evidence that the system has yielded dependable results.
8. Statement of direct and indirect costs.

The measurement procedures to determine the accomplishment of the training outcomes are not the same as the measurement procedures to determine pupil performance. Pupil performance measures are not a direct means of determining training system outcomes. Training system evaluation criteria should be directly related to the unique objectives listed.

The economic and political contraints of the school community should be considered in the development of the training requirements. Training systems which require instant and dramatic modifications of existing scheduling practices, teaching procedures, and personnel policies will be less effective than those which recognize existing competencies and build on associated strengths.

Deterline has strongly stated the point as follows: "So we can conclude that both teachers in educational settings and instructors in training settings might be expected to cry 'foul and unfair' when faced with the notion of accountability. And I don't blame them a bit. Unless someone else accepts accountability for teaching those teachers relevant skills beyond those they already possess, and unless the conditions that limit their effectiveness can be changed, then there is no justification for expecting them to do better, or for holding them accountable for doing so. They need more effective techniques, procedures, methods, materials and other components of instruction. And then, when new components are added, no matter how innovative, how effective, there will be a limit to the overall effectiveness of that totality, probably still short of that ideal of all students learning everything. That new limit will have to be determined empirically, and then additional research and development efforts can be aimed at raising that ceiling."

Installation systems. An installation system consists of the procedures and materials required by a local educational agency to effectively introduce an instructional program. It provides for those considerations related to the administrative authority and responsibility for supervisory service, curriculum, and pupil personnel services. This system includes briefing information about the instructional program, procedures for providing various categories of agency personnel with pupil performance data in a form they find manageable and useful, and materials which can be used for public information purposes.

Installation systems evaluation criteria in terms of system objectives:

1. Materials for use by installing agency to describe a program at a "public information" level suitable for at least the following audiences: governance group, operating staff, general public.
2. Statement of feasible procedures for procuring and storing instructional and training system materials.
3. Statement of anticipated required revisions in prevailing personnel and/or administrative policies.
4. Statement of feasible procedures for required personnel assignment and scheduling.
5. Statement of feasible alternatives for training of training-supervisors.
6. Statement of anticipated administrative requirements for maintaining the program.
7. Statement of alternative options where the new program interfaces with remaining extant programs.
8. Evidence that the system has yielded dependable results.
9. Statement of direct and indirect costs.

The requirements for the maintenance of program outcomes are of particular importance. Although the installation of a program may be initially successful, without knowledge of the requirements to sustain program performance, the user may find that in a very brief period the program will become fragmented, diluted, and void of its initial integrity and effectiveness.

Accountability systems. An accountability system includes all of the mechanisms required to maintain maximum pupil performance and the procedural adequacy of each of the three above-mentioned systems: the instructional system, the

training system, and the installation system. Differentiating assignments and establishing responsibilities for which development personnel will be held accountable are anchored to specified program and/or procedural outcomes.

Accountability systems evaluation criteria in terms of system objectives:

1. Assessment devices for all program outcomes for which accountability is to be maintained, related both to pupil proficiency and to systems procedural adequacy.
2. A human resources analysis which may be used as a basis for differentiating and assigning accountable responsibility.
3. A data collection and reporting procedure for maintaining accurate and timely bench mark information.
4. Statement of consequences or alternatives to be effected if performance at a given bench mark test is determined to be unacceptably low.
5. Available forms and/or equipment to operate and report the bench mark tests.
6. Evidence that the system has yielded dependable results.
7. Statement of direct and indirect costs.

The functions of the accountability system are related to process evaluation as discussed by Stufflebeam et al. Since the instructional product developer relies heavily on the application of formative evaluation strategies, he must also provide a means for the formative documentation of his work. A systematically prepared documentation program includes procedures related to the planning, reporting, and reviewing requirements at each stage of the development effort.

Similarly, Harmes has written: "Defining the objectives of a process performance contract takes on dimensions that are different from those of product contracts. As the best type of objective for product contracting defines how it will be determined that students have learned, the best type of objective for process contracting defines how it will be determined that a process has been, or is being, operated as intended."

"Special attention must be given, however, to the methods by which performance will be monitored. Products can be and are

measured at a point in time. Processes, on the other hand, take place over a span of time. Unless one wishes to continuously monitor an instructional process, a very time-consuming and expensive thing to do, he must devise techniques whereby the process is sampled at points in time or during short time spans. From this sampling, it is then inferred that the process taking place between samples is similar to the samples of it."

The Development Process

Progress in the development of a product can be described in terms of three distinguishable phases: Pre-License, Nonexclusive License, and Distribution License.

Pre-License. Emphasis during the Pre-License phase is on the development of instructional and training systems. The developer is heavily involved in prototype and component testing. Tryouts are of short duration, one day to several weeks. The developer provides intensive human and financial support to school personnel during these early stages of product development. It is during this period that uncertainty related to product specifications, instruction parameters, and instructional effectiveness is reduced. The materials are not developed at the expense of the schools.

Nonexclusive License. The Nonexclusive License phase provides the opportunity for larger scale tryouts of longer duration. Emphasis is on the development of installation and accountability systems. Quality assurance procedures are developed for use by the schools in monitoring product performance. This is a period of shared cost responsibility between the developer and the schools.

The Nonexclusive License phase also provides the developer with an index of user interest and product utility. Efforts during this period involve interface activities between completed systems and educational programs already operational. The assembly of systems into programs and installing the program on a wide-scale basis provide new areas for investigation that heretofore have received no research and development attention.

Distribution License. The Distribution License phase involves assisting the private sector, within the constraints of the developers mission and resources, to install the instructional and support systems on a national basis. This is the category of the private marketplace and public school finance. Although this is not a research and development

category, specific attention must be given during the earlier research and development phases to advance planning of procurement alternatives if the results of a programmatic development effort are ever to be implemented. An important axiom is, "Don't build it out of reach of the consumer."

To the individual teacher, without resources, working with hand-me-down materials, or materials he has purchased on his own, it can be said: "Go to it, more power to you, if you can identify what society wants students to learn, how you will teach it, and how you will know if your students have mastered the content, you are accountable."

To the task force, given fifty thousand dollars to accomplish what others have failed to accomplish given five times as much, working day and night to improve the quality of instruction, assisting teachers in meeting the objectives of the instructional program, forced to use crude measures to determine program effectiveness and constantly operating in an atmosphere of expediency, it can be said: "Tell it as it is, you have nothing to hide, don't oversell, and at the same time don't underestimate the contribution you are making."

And, to the institutions, both public and private, who are in position to bring together the resources necessary to develop research-verified, reliable, cost-effective, and cost-efficient instructional products, it can be said: "Unfortunately, not even you can do it alone."

Stephen Barro has written, "Each participant in the educational process should be held responsible only for those educational outcomes that he can affect by his actions or decisions and only to the extent that he can affect them."

If we accept Barro's position, then if follows that the instructional product developer is directly responsible for ensuring that his product leads to the pre-specified outcomes for which it was designed. But, the further removed the materials and procedures are from his control, the less he can be held accountable for the effectiveness of the product. Therefore, he must accept responsibility for the development of the product and be committed to affecting a continuous chain of responsibility/accountability for each of the agencies, systems, and individuals who use the product. Without this sequential, programmatic, sharing of responsibility, it is doubtful that the educational community will be able to respond to the existing demands for accountability.

Bibliography

Baker, Robert L. and Richard E. Schutz, editors. Instructional Product Development. Southwest Regional Laboratory for Educational Research and Development. New York: Van Nostrand Reinhold Co., 1971.

Barro, Stephen M., "An Approach to Developing Accountability Measures for the Public Schools," Phi Delta Kappan, 52: 196-205, 1970.

Deterline, William A., "Applied Accountability," Educational Technology, 11: (1) 15-20, 1971.

Farquhar, J.A., "Accountability, Program Budgeting, and the California Educational Information System: A Discussion and a Proposal," The Rand Corporation, R-637-CC/RC, April, 1971.

Harmes, H.M., "Specifying Objectives for Performance Contracts," Educational Technology, 11: (1) 52-56, 1971.

Harsh, J. Richard, "Biopsy of an Educational Growth," CACER Research Resume: Proceedings of the 23rd Annual State Conference on Educational Research, 46: 3-14, 1971.

Stufflebeam, Daniel, L. et al., Educational Evaluation and Decision Making, Itasca, Illinois: F.E. Peacock Publishers Inc., 1971.

Zaleznik, Abraham, "Management of Disappointment," Harvard Business Review, 45: (6) 59-70, 1967.

CALIFORNIA ADVISORY COUNCIL
ON EDUCATIONAL RESEARCH

Sponsored By

CALIFORNIA TEACHERS ASSOCIATION

1972-1973