DOCUMENT RESUME

ED 075 464 TM 002 542

AUTHOR	Goolsby, Thomas M., Jr., Comp.
TITLE	Invitational Conference on Measurement in Education; Proceedings (of the) Eleventh Southeastern Conference (December 8-9, 1972). Session II.
INSTITUTION	Georgia Univ., Athens. Coll. of Education.
SPONS AGENCY	Harcourt Brace Jovanovich, Inc., New York, N.Y. Test Dept.
PUB DATE	9 Dec 72
NOTE	66p.; Papers presented at Session II of Southeastern Invitational Conference on Measurement in Education (11th, Athens, Georgia, December 9, 1972)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Conference Reports; *Ethics; *Human Dignity; *Psychological Testing; Speeches; *Testing

ABSTRACT

Two papers concerned with the ethics of testing and one paper that is a proposal for alternatives to psychological testing, presented at Session II of the Invitational Conference on Measurement in Education, are provided. The papers and their authors are: "Ethical Issues and Questions about Testing Confronting Test Publishers" by Thomas J. Fitzgibbon, "Ethical Issues in the Use of Humans for Research" by W. L. Bashaw, and "Alternatives to Psychological Testing" by Donald N. Bersoff. (For related documents, see TM 002 522-539, 541, 543-549.) (DB)

PROCEEDINGS

Invitational Conference on Measurement in Education

Eleventh Southeastern Conference -- December 8-9, 1972

Compiled and Organized by

Thomas M. Goolsby, Jr.
Conference Chairman
University of Georgia

Session II
Saturday Morning
December 9, 1972
University Motor Inn
University of Georgia
Athens, Georgia

Papers

Ethical Issues and Questions About Testing Confronting Test Publishers - Thomas J. Fitzgibbon

Ethical Issues in the Use of Humans for Research
W. L. Bashaw

Alternatives to Psychological Testing
Donald N. Bersoff

Sponsors

College of Education, University of Georgia

Test Department, Harcourt Brace Jovanovich, Inc.

Foreword

The Saturday Morning Session was devoted to ethics of testing and a proposal for alternatives to psychological testing. This programming was in great measure a response to recent emphases on ethics by the American Psychological Association, U. S. Supreme Court decisions concerning denial of employment opportunities through testing, and a call from some quarters for alternatives to testing.

The presentations were well received by the audience with lively discussions. There seems to remain substantial disagreement among authorities, school personnel, and the lay public concerning testing. In addition to the topics presented here, there needs to be some "in depth" presentations and discussions concerning:

1. Testing in public schools, i.e., for special education purposes, achievement, ability, and so forth (testing mainly for counseling and program planning purposes).
2. "One Shot" testing concerned with major decisions about a person's future, i.e., GRE, MCAT, LSAT, and so forth.
3. Testing to deny and/or offer employment opportunities.

Thomas M. Goolsby, Jr.
Athens, Georgia
February 20, 1973

Invitational
Conference on
Measurement
in Education

ELEVENTH SOUTHEASTERN CONFERENCE

December 8-9, 1972
University Motor Inn,
University of Georgia
Athens, Georgia

Sponsors:

College of Education, University of Georgia
Test Department, Harcourt Brace Jovanovich, Inc.

ELEVENTH SOUTHEASTERN INVITATIONAL CONFERENCE

ON MEASUREMENT IN EDUCATION

Pre-Conference Session
in Conjunction with

The National Council on Measurement in Education
The Association for Measurement and Evaluation in Guidance

Friday, December 8, 1972
Main Dining Room
University Motor Inn
Athens, Georgia

Ira E. Aaron, University of Georgia, Presiding

9:00 - 9:30 a.m.	Legislative Accountability for Public Schools - Goals for Georgia Sam A. Nunn, United States Senator
	Open Discussion
9:30 - 10:00 a.m.	Public Schools in Transition - One Student Body, Its Needs and Directions George R. Rhodes, Jr., United States Bureau of Equal Educational Opportunity
	Open Discussion
10:00 - 10:00 a.m.	Coffee Break.
10:30 - 11:00 a.m.	Man to Mankind: The International Dimension of Teacher Education David Imig, American Association of Colleges of Teacher Education
	Open Discussion
11:00 - 11:30 a.m.	Measurement and the "Right to Read" Ira E. Aaron, University of Georgia
	Open Discussion
11:30 - 12:00 noon	Measurement for Purposes of Evaluation Dennis E. Henkle, Virginia Polytechnic Institute and State University
	Open Discussion

Friday, December 8, 1972 E. Paul Torrance, University of Georgia, Presiding

8:00 a.m.-1:00 p.m.	Registration Grace Wray
1:00 - 1:30 p.m.	Introducing the Conference
Welcome:	Joseph A. Williams, Dean, College of Education, University of Georgia
Greetings:	Philip I. Clark, Test Department, Harcourt Brace Jovanovich, Inc.
1:30 - 2:15 p.m.	Problems in Evaluation Studies of Educational Programs for Minorities Jay A. Davis, Educational Testing Service, Raleigh, North Carolina
2:15 - 3:30 p.m.	
Session I: Main Dining Room	Perspectives of School Desegregation in the Southeast Morrill M. Hall, Milton Hill, and Harry B. Williams, Center for Educational Improvement, University of Georgia
Session II: Large Gold Room	Issues in the Testing of Indian Children with Some Emphasis on Behavior Modification and Other Operative Programs Joseph D. Blanchard, U. S. Bureau of Indian Affairs
Session III: Small Gold Room	A Systems Approach to Reading for Migrant Children Muriel M. Abbott, Test Department, Harcourt Brace Jovanovich, Inc.
Session IV: Bronze Room	The Prediction of Achievement Means of Schools from Non-School Factors Through Criterion Scaling Thomas C. Innes, State Testing Bureau, University of Tennessee
3:30 - 3:45 p.m.	Coffee Break

ALL ACTIVITIES ARE SCHEDULED IN MAIN DINING ROOM UNLESS OTHER WISE NOTED.

3:45 - 5:00 p.m.

Session I: Large Gold Room

Mental Retardates: Children With Different Abilities

Measurement of the Self Concept of Mentally Retarded Children
Jack Shultz

Affective Characteristics of the Mentally Retarded
Bert Richmond

Motor Development in Mentally Retarded Children
Ernest Bundschuh

Evaluating Achievement of the Mentally Retarded - A Comprehensive Process
Andrew Shotick
Mental Retardation Center, University of Georgia

Session II: Main Dining Room

The Effects of Item Analysis, Frequency Distributions, and Profile Analysis on Instruction in Two Programs for the Disadvantaged

Information Systems and Economically Deprived Children
William F. White, University of Georgia

Evaluation of Follow-Through Programs
Frances Cox, Atlanta City Schools

Session III: Small Gold Room

Identification of Gifted and Creative Children and Youth Among Black Disadvantaged Groups
E. Paul Torrance and Catherine B. Bruch, Department of Educational Psychology, Measurement and Research, University of Georgia

Session IV: Bronze Room

Who Needs Adult Basic Education?
Joseph E. Fuller, Atlanta City Schools and Fulton County Schools

6:30 - 7:30 p.m. Social Hour

7:30 p.m. Dinner

Saturday, December 9, 1972 Clemmie W. Brower, Atlanta City Schools, Presiding

8:00 - 9:45 a.m. Davis House Cafeteria	Steering Committee Breakfast Meeting
9:00 - 9:45 a.m.	Ethical Issues and Questions About Testing Confronting Test Publishers Thomas J. Fitzgibbon, Director, Test Department, and Vice-President, Harcourt Brace Jovanovich, Inc. Open Discussion
9:45 - 10:30 a.m.	Ethical Issues in the Use of Humans for Research W. L. Bashaw, University of Georgia Open Discussion
10:30 - 11:00 a.m.	Coffee Break
11:00 - 11:45 a.m.	Alternatives to Psychological Testing Donald N. Bersoff, University of Georgia Open Discussion
11:45 - 12:00 noon	Respondent Warren G. Findley, University of Georgia

Planning Committee

Thomas M. Goolsby, Jr., University of Georgia (Chairman)
Clemmie W. Brower, Atlanta City Schools
Robert Duby, Harcourt Brace Jovanovich, Inc.
James M. Dunlap, North Carolina Department of Public Instruction
Robert B. Frary, Virginia Polytechnic Institute and State University
Thomas C. Innes, State Testing Bureau, University of Tennessee
Thomas H. Parry, Clemson University
Robert H. Smoot, Harcourt Brace Jovanovich, Inc.

Table of Contents

*Refers to number at the bottom of pages.

ETHICAL ISSUES AND QUESTIONS ABOUT TESTING FOR TEST PUBLISHERS

Dr. Thomas J. Fitzgibbon, Vice President
Director, Test Department
Harcourt Brace Jovanovich, Inc.

IT SEEMS TO BE MY FATE LATELY THAT I'M ASKED TO TALK ABOUT THINGS FROM A TEST PUBLISHER'S POINT OF VIEW. "GIVE US YOUR VIEWPOINTS ABOUT CRITERION REFERENCED TESTING," THEY'LL SAY; OR "HOW DOES A TEST PUBLISHER FEEL ABOUT THE TEST MORATORIUM ISSUE," OR, "DO TEST PUBLISHERS THINK THAT THE NATIONAL INSTITUTE FOR EDUCATION WILL REALLY COME OUT WITH GOOD, USEFUL SUGGESTIONS FOR THE MODIFICATION OF STANDARDIZED TESTING?" I HAVE ANSWERS TO ALL THESE QUESTIONS OF COURSE. TO THE FIRST, ABOUT CRITERION REFERENCED TESTING, I TELL THEM I HAVE WRITTEN A PAPER CONTAINING THESE VIEWS AND WILL BE MORE THAN HAPPY TO SEND THEM AN AUTOGRAPHED COPY. TO THE SECOND QUESTION REFERRING TO HOW I FEEL ABOUT THE MORATORIUM ISSUE I SAY, "VERY NERVOUS," AND TO THE THIRD WHICH RELATES TO NIE HELPING US IMPROVE STANDARDIZED TESTING, I PUT ON MY HOPEFUL FACE AND ANSWER, "GOD, I HOPE SO." ANOTHER WAY THE QUESTION MAY BE PHRASED THESE DAYS IS, "WHAT DOES THE TEST INDUSTRY THINK ABOUT X?" THE FIRST TIME I FIELDED ONE OF THESE I COULD ONLY SAY, "WHAT TEST INDUSTRY? I DIDN'T KNOW WE HAD ONE."

ADDRESS GIVEN AT THE ELEVENTH SOUTHEASTERN INVITATIONAL CONFERENCE ON MEASUREMENT IN EDUCATION HELD IN ATHENS, GEORGIA, DECEMBER 8, 1972

THEN WILLIAM F. BRAZZIEL, IN A SPEECH FEBRUARY LAST IN WASHINGTON, D.C. , TALKED ABOUT A 300-MILLION DOLLAR TESTING INDUSTRY AND I SAT THERE ENTRANCED -- AND INFURIATED. ENTRANCED BECAUSE I HADN'T KNOWN THERE WAS THAT MUCH MONEY AROUND, AND INFURIATED AT MY SALES MANAGER BECAUSE HE HAD NEVER TOLD ME ABOUT IT. TO MAKE MATTERS WORSE, I HAVE RECENTLY READ WHERE RALPH NADER INFORMED THE NEW JERSEY EDUCATIONAL ASSOCIATION THAT, INDEED, THERE IS A 300-MILLION DOLLAR TESTING INDUSTRY. NOW I'M SURE IT MUST BE SO AND THUS AM IMMEDIATELY DECLARING TO ALL,THAT THERE IS A SALES POSITION OPEN ON MY STAFF. BUT BACK TO THE QUESTIONS. WHETHER ADDRESSED TO ME AS A PUBLISHER OR AS A MEMBER OF THE TESTING INDUSTRY, WHAT I MUST ALWAYS SAY FIRST, IS THAT I DO NOT REPRESENT ALL TEST PUBLISHERS; I DON'T WANT TO AND I EXPECT THEY DON'T WANT ME TO. THEY HAVE THEIR OWN VIEWS AND VOICES. AND SO I SAY AGAIN TODAY--I'M SPEAKING ONLY FOR MY HBJ TEST DEPARTMENT COLLEAGUES AND MYSELF. AND SOMETIMES ONLY FOR MYSELF. COME TO THINK OF IT, MOST OF THE QUESTIONS I RECEIVE ARE RATHER DULL. WHY DOESN'T SOMEBODY ASK ME WHAT TEST PUBLISHERS THINK OF SEX? OR IS IT TRUE THAT NEW YORK CITY MASSAGE PARLORS ARE SECRETLY FINANCED BY THE TEST INDUSTRY? OR, EVEN, DOESN'T THE VERY EXISTENCE OF A RAQUEL WELCH REFUTE THE THEORY OF A NORMAL CURVE? THINGS LIKE THAT.

I'M SURE WE COULD SAFELY PREDICT THAT THIS MORNING WE WILL NOT BE ABLE TO FACE AND DISCUSS, ALL ETHICAL ISSUES WHICH SURFACE WHEN EDUCATIONAL TESTS ARE DEVELOPED AND USED. HOWEVER, IF WE DEFINE ETHIC AS A "PRINCIPLE OF RIGHT OR GOOD CONDUCT, OR

A BODY OF SUCH PRINCIPLES" AND REFLECT UPON OUR OWN EXPERIENCES, WE CAN AT LEAST LAY OUT FOR DISCUSSION CERTAIN SITUATIONS WHICH ARE IMPORTANT TO US. I PROPOSE TO DO THIS IN MY REMAINING TIME AND I INVITE YOU TO THINK ALONG WITH ME. WHEN WE GET THROUGH, WE WON'T HAVE SETTLED ANYTHING, BUT WE MAY HAVE RAISED SOME POINTS THAT WILL MAKE SENSE TO PURSUE IN OTHER PLACES AT OTHER TIMES.

WE CAN START WITH A VERY BROAD QUESTION WHICH NO ONE HAS EVEN ASKED ME BUT I THINK IT SHOULD BE POSED, I.E., IS IT RIGHT TO SELL ANYTHING IN SOCIAL SERVICE OR HELPING PROFESSION AREAS? SHOULDN'T EVERYTHING BE FREE? WELL, OBVIOUSLY IF THE ANSWER TO THAT WAS "YES," OUR AMERICAN TRADITIONS WOULD UNDERGO SOME DRAMATIC CHANGES. ALL OF THE FACULTY AT THE UNIVERSITY OF GEORGIA WOULD DONATE THEIR SERVICES, AND HBJ WOULD GIVE AWAY ITS TESTS UNTIL THE STOCKHOLDERS ROSE UP IN ARMS. I DON'T THINK EITHER ONE OF THOSE THINGS ARE GOING TO HAPPEN, BUT A SERIOUS QUESTION STILL REMAINS. WHEN ONE SELLS, EITHER HIS EXPERTISE AS A FACULTY MEMBER OR HIS EXPERTISE IN PREPARING A TEST INSTRUMENT, ONE PAYS CLOSE ATTENTION TO DEMAND, AND DEMANDS CAN DICTATE A RATHER NARROW RESPONSE TO THE BROAD SPECTRUM OF HUMAN NEEDS. FOR EXAMPLE, IT IS FAR SAFER TO LISTEN TO DEMAND AND PUBLISH A NEW READING TEST, THAN IT IS TO PUBLISH A NEW AFFECTIVE DEVICE OF SOME KIND. WITH ALL OF THE TALK ABOUT THE AFFECTIVE DOMAIN, WE DON'T RECEIVE MANY REQUESTS FOR NEW INSTRUMENTATION IN THIS AREA. "CREATING A MARKET" AS IS OFTEN DONE IN OUR SOCIETY, SEEMS ALMOST REPREHENSIBLE IN THE HELPING PROFESSIONS. YET, HOW MANY PERSONS HERE TODAY FEEL THAT THE SCHOOL

ENVIRONMENT NEEDS MORE THAN COGNITIVE ORIENTATION AND THAT SOMETHING SHOULD BE DONE ABOUT IT. QUESTION TO YOU: SHOULD WE CREATE A DEMAND? IS THAT RIGHT CONDUCT?

ISSUES HAVE A WAY OF BEING RELATIVE. WHAT IS ONE MAN'S ISSUE MAY BE ANOTHER MAN'S CONVENIENCE. WE ARE ALL AWARE THAT TESTS IN THE PUBLISHER'S CATALOG RANGE FROM RECENT PUBLICATIONS TO ONES WHICH MAY BE OF FAR EARLIER VINTAGE. IT WOULD SEEM THAT THE THING TO DO IS TO SET UP A SCHEDULE IN WHICH TESTS ARE PERIODICALLY EVALUATED AS TO THEIR CURRENT USEFULNESS AND ACTION IS TAKEN ACCORDINGLY. AS A MATTER OF FACT, THIS IS WHAT IS USUALLY DONE. SALES ARE A GOOD BAROMETER TO WATCH, BY THE WAY. IF A TEST IS MEETING SOME MEASUREMENT NEED, IT WILL BE SOLD AND USED. HOWEVER, THIS SHOULD BE ONLY ONE OF THE INDICES TO WHICH ONE ATTENDS. OTHERS WOULD INCLUDE NEW TECHNICAL DEVELOPMENTS WHICH WOULD MAKE THE TEST EVEN MORE EFFECTIVE, CHANGING SCHOOL CONDITIONS WHICH MIGHT DICTATE A DIFFERENT FORMAT, AND SO ON. WHAT IS "RIGHT AND GOOD CONDUCT" ON THE PART OF THE PUBLISHER HERE CAN BE OBSCURE AS FAR AS THE TEST USER IS CONCERNED, AS WITNESS THE FOLLOWING STATEMENTS. THE FIRST IS TAKEN FROM GUIDELINES TO THE TEST SELECTION COMMITTEE BY THE SCHOOL BOARD OF ONE OF OUR LARGEST CITIES AND IT READS:

"THE COMMITTEE WILL NOT ADOPT ANY TEST WITH A COPYRIGHT PRIOR TO 1970."

THE SECOND IS A RECENT PERSONAL COMMUNICATION FROM A HIGH SCHOOL PRINCIPAL WHO WROTE TO THE PRESIDENT OF MY COMPANY. HE SAID IN PART:

DEAR SIR:

WE HAVE BEEN USING THE AM, BM, AND CM FORMS OF THE RE-USUABLE METROPOLITAN ADVANCED READING TESTS, PUBLISHED BY YOUR FIRM, TO GIVE ANNUAL READING TESTS TO OUR STUDENTS. WE GIVE A DIFFERENT FORM EACH YEAR IN ORDER THAT EACH TEST WILL BE A NEW EXPERIENCE FOR OUR YOUNGSTERS. THUS, WE USE EACH FORM EVERY THIRD YEAR.

MUCH TO OUR DISMAY, WE HAVE JUST LEARNED THAT THE 1958 EDITION OF THESE TESTS, WHICH WE HAVE BEEN PURCHASING AND RECEIVING UP TO THE PRESENT, HAS NOW BEEN REPLACED BY A NEW 1970 EDITION, AND THAT YOU DO NOT PLAN TO CONTINUE PRINTING THE DIGITEK ANSWER SHEETS FOR THE 1958 EDITION.

THIS DECISION, IF CARRIED OUT, MEANS THAT OUR WHOLE INVENTORY OF 1958 TESTS IS WORTHLESS AND MUST BE DISCARDED. YOU ARE AWARE, OF COURSE, THAT IN THESE DAYS OF TIGHT EDUCATION BUDGETS WE CAN NOT AFFORD TO WASTE ALL THIS MONEY AND THAT SUCH A WASTE WOULD BE VERY HARMFUL TO OUR BOYS AND GIRLS.

IT SEEMS TO US THAT WHEN A PUBLISHER SELLS <u>REUSUABLE</u> TESTS WITH SEPARATE ANSWER SHEETS HE IS OBLIGATED TO CONTINUE PRINTING THE ANSWER SHEETS UNTIL THE TEST PAPERS ARE NO LONGER USUABLE. (AND THEN HE CAME ON WITH THE UNKINDEST CUT OF ALL):

I MIGHT ADD, INCIDENTALLY, THAT WE SEE LITTLE DIFFERENCE IN THE QUALITY BETWEEN THE 1958 AND THE 1970 EDITIONS OF THE TEST.

SINCERELY YOURS,
PRINCIPAL X

MY QUESTION IS: WHAT IS "RIGHT AND GOOD CONDUCT" ON THE PART OF THE PUBLISHER HERE? TO MAKE IT EASIER, LET ME REPORT THAT WE HAD NOT DISCONTINUED PRINTING THE ANSWER SHEETS; SOMEONE ON HIS STAFF HAD MISINFORMED HIM. ALSO, EVEN IF THEY HAD BEEN WE WOULD HAVE LICENSED HIM AND HELPED HIM TO DEVELOP HIS OWN. BUT THE POINT IS, HE WANTED "OLD MATERIALS" AND THE SCHOOL BOARD PREVIOUSLY CITED WANT "NEW MATERIALS." HOW DOES ONE MAINTAIN THE PROPER BALANCE BETWEEN THE TWO?

A CANADIAN GUIDANCE ORGANIZATION ONCE ASKED ME TO PUT TOGETHER A PAPER DISCUSSING ETHICAL AND LEGAL ISSUES INVOLVED IN DISSEMINATION OF TEST INFORMATION. BEING AN EX-COUNSELOR, AND STILL ONE AT HEART, I ROSE TO THE OCCASION AND PRODUCED A PIECE ENTITLED, "THE ETHICAL AND LEGAL POSITION OF THE COUNSELOR IN DIVULGING TEST INFORMATION," WHICH LATER WARREN FINDLEY WAS KIND ENOUGH TO PUBLISH IN THE FIRST JOURNAL OF MEASUREMENT AND EVALUATION IN GUIDANCE IN SPRING, 1968. THE RESPONSE FROM BOTH THE CANADIAN AND AMERICAN EXPOSURE WAS HEAVY AND CONTINUOUS FOR OVER A PERIOD OF TWO YEARS. I'D LIKE TO SAY THAT THIS WAS DUE TO MY DEATHLESS PROSE, BUT I'M QUITE SURE IT WAS NOT. IT WAS DUE, I THINK, TO A FEELING ON THE PART OF TEACHERS, COUNSELORS, SCHOOL ADMINISTRATORS, PARENTS, AND CONCERNED OTHERS THAT THE RIGHT TO HUMAN PRIVACY IS OF EXTRAORDINARY VALUE IN OUR SOCIETY AND THAT IT MUST BE PRESERVED. WITH THAT, FOR MANY OF US AT LEAST, ALSO COMESTHE REALIZATION THAT TECHNOLOGY CAN OPEN PRIVATE DOORS AS NEVER BEFORE. TESTING (MEANING DEVELOPERS, RESEARCHERS,

AND USERS) HAS A TREMENDOUS STOCK IN THIS ISSUE FOR IT YIELDS INFORMATION WHICH CAN BE CONSIDERED EXCEEDINGLY PRIVATE IN NATURE. ON THE OTHER HAND, TESTING IS SOMETIMES USED TO GATHER INFORMATION FOR THE EXPRESS PURPOSE OF MAKING SUCH INFORMATION PUBLIC AS WITNESS NATIONAL ASSESSMENT, STATE ASSESSMENT PROGRAMS, AND DIVERSE SCHOOL SYSTEMS REPORTING READING SCORES TO THEIR COMMUNITIES. IN MOST INSTANCES, THESE ARE NOT SCORES OF INDIVIDUALS; CERTAINLY THE NATIONAL ASSESSMENT PROGRAM IS FAR FROM SUCH A POSITION, BUT MY POINT IS THIS--TESTING IS NOT ALWAYS PRIVATE, SOMETIMES IT IS DELIBERATELY PUBLIC. AND SOMETIMES, IN MY OPINION, THIS IS VERY JUSTIFIABLE, AS A PARENT HAS A RIGHT TO KNOW ABOUT THE STATUS AND WELFARE OF HIS YOUNGSTER; SO DOES THE COMMUNITY HAVE THE RIGHT TO KNOW ABOUT THE STATUS AND WELFARE OF ITS SCHOOLS. PERMIT ME A LONGISH EXCERPT AT THIS TIME FROM THE PREFACE TO GUIDELINES FOR THE COLLECTION, MAINTENANCE AND DISSEMINATION OF PUPIL RECORDS. THIS IS A REPORT OF A CONFERENCE SPONSORED BY THE RUSSELL SAGE FOUNDATION ON ETHICAL AND LEGAL ASPECTS OF SCHOOL RECORD KEEPING:

> ...EVERY ONE OF US WILL ASSESS DIFFERENTLY THE TOLERABLE LIMITS OF INTRUSION UPON OUR PRIVACY. AND EVEN A SINGLE INDIVIDUAL WILL ASSESS THOSE LIMITS DIFFERENTLY AT DIFFERENT TIMES AND IN VARYING SETTINGS. NOR ARE COMMUNITY NEEDS STATIC. THEY, TOO, DIFFER IN IMPORTANCE, OR IN URGENCY, AT DIFFERING TIMES AND UNDER DIFFERING CIRCUMSTANCES.

THUS, THE CONFLICT BETWEEN THE NEEDS FOR PERSONAL PRIVACY AND THE CLAIMS OF THE COMMUNITY IS ALWAYS A CONTEST AMONG MANY POWERFUL AND CONSTANTLY VARYING FORCES. THE MODERN VERSION OF THIS ENDURING STRUGGLE BETWEEN THE INDIVIDUAL AND HIS ENVIRONMENT IS NOT SIMPLY A MATTER OF URBAN AND SUBURBAN CROWDING, OR EXPLOITATION AND PREJUDICE, OR A POLLUTED ENVIRONMENT, AND A DEFILED LANDSCAPE. MORE SUBTLE, BUT BY NO MEANS LESS IMPORTANT, IS THE INCREASINGLY INTRUSIVE CAPACITY OF MODERN TECHNOLOGY, THE REALITY OF INSTANT COMMUNICATION AND THE AVAILABILITY OF COMPUTERIZED MEMORIES.

MODERN SCIENCE HAS INTRODUCED A NEW DIMENSION INTO THE ISSUES OF PRIVACY. THERE WAS A TIME WHEN AMONG THE STRONGEST ALLIES OF PRIVACY WERE THE INEFFICIENCY OF MAN, THE FALLIBILITY OF HIS MEMORY, AND THE HEALING COMPASSION THAT ACCOMPANIED BOTHTHE PASSING OF TIME AND WARMTH OF HUMAN RECOLLECTION. THESE ALLIES ARE NOW BEING PUT TO ROUT. MODERN SCIENCE HAS GIVEN US THE CAPACITY TO RECORD FAITHFULLY, TO MAINTAIN PERMANENTLY, TO RETRIEVE PROMPTLY, AND TO COMMUNICATE BOTH WIDELY AND INSTANTLY, IN AUTHENTIC SOUND OR PICTURES OR IN SIMPLE WRITTEN RECORDS, ANY ACT OR EVENT OR DATA OF OUR CHOICE. TECHNOLOGY CAN NOW TRANSFORM WHAT PARTICIPANTS BELIEVED WERE PRIVATE EXPERIENCES INTO PUBLIC EVENTS.

IN CONSIDERING THE CHALLENGE OF THESE DEVELOPMENTS, IT IS WELL TO REMIND OURSELVES THAT NOT EVERY LAYING ON OF HANDS IS ASSAULT; NOR IS EVERY TECHNICAL TRESPASS ACTIONABLE; NOR EVERY TAKING OF PRIVATE PROPERTY FORBIDDEN. SO, TOO, NOT EVERY AFFRONT TO INDIVIDUAL DIGNITY IS OBJECTIONABLE.

THE RANGE OF TOLERABLE INVASIONS OF PRIVACY IS AMPLE. THEY INCLUDE THOSE THAT ARE PURELY TECHNICAL, OR INSUBSTANTIAL, OR WHOLLY INADVERTENT. THEY ALSO ENCOMPASS THOSE INVASIONS DEEMED DESIRABLE IN ORDER TO FURTHER A PARAMOUNT PUBLIC PURPOSE. THE CRITERIA FOR SEPARATING THE PERMISSIBLE FROM THE IMPERMISSIBLE INVASIONS ARE NUMEROUS AND COMPLEX. THEY MUST BE CHOSEN WITH SPECIAL ATTENTION TO THE PARTICULAR SETTING IN WHICH THE ISSUE ARISES.

I RECOMMEND THE GUIDELINES TO YOU FOR THEY DO ATTEMPT TO HELP US ALL IN SEPARATING THE PERMISSIBLE FROM THE IMPERMISSIBLE. THEY CAN HELP US AS WE SEARCH FOR THE PROPER COURSE OF ACTION OF "RIGHT OR GOOD CONDUCT."

THE QUESTION "WHEN IS A TEST COMPLETE ENOUGH TO SELL" SHOULD BE ANSWERED "NEVER," IF BY THE QUESTION ONE IMPLIES THAT NO TEST INSTRUMENT SHOULD BE USED UNTIL EVIDENCE IS FURNISHED DOCUMENTING ITS EFFICACY UNDER ALL CONDITIONS, AND ALL POSSIBLE SITUATIONS, IN WHICH IT MAY BE EMPLOYED.

EVEN OUR STANFORD ACHIEVEMENT TEST, WHICH NEXT YEAR WILL SEE ITS 50TH YEAR OF USE IN AMERICAN SCHOOLS, COULD NOT MEET THIS STANDARD. IF, HOWEVER, THE QUESTION IMPLIES THAT CRITERIA FOR COMPLETION SHOULD BE EXPLICATED AND EVIDENCE OFFERED TO THE POTENTIAL USER THAT CRITERIA HAVE, OR HAVE NOT BEEN SATISFIED, THEN THE ANSWER BECOMES OPTIMISTICALLY MANAGEABLE. I WON'T GO INTO THESE CRITERIA AT THIS TIME, THEY ARE ADDRESSED IN THE APA-AERA-NCME STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTS AND MANUALS, NUMEROUS TEST AND MEASUREMENTS TEXTS, AND SO ON. I WOULD LIKE TO SUGGEST AN ADDITION TO THE QUESTION, "WHEN IS A TEST COMPLETE ENOUGH TO SELL?" WHICH ASKS, "AND WHEN SHOULD IT BE ABANDONED?" THE APPARENT, AND RIGHT ANSWER TO THAT PART OF THE QUESTION IS "WHEN IT DOESN'T WORK FOR YOU IN YOUR SITUATION; WHEN IT DOESN'T SATISFY YOUR MEASUREMENT NEEDS." WHEN SUCH IS THE CASE, I.E., IT DOESN'T "WORK RIGHT" ANYMORE, IT CAN MEAN, FOR ONE THING THAT THE DEVELOPER HAS NOT KEPT THE TEST UNDER CONTINUING RESEARCH SCRUTINY. BEFORE WE BECOME TOO ROMANTIC ABOUT THE QUEST FOR ETERNAL TRUTH, HOWEVER, LET'S FACE UP TO BUDGETARY RESTRAINTS. NO DEAN OF THE FACULTY, NO DIRECTOR OF AN OE R&D CENTER, NO NATIONAL INSTITUTE FOR EDUCATION, AND NO TEST PUBLISHER CAN EVER AVOID OR FORGET THEM. SO THE SITUATION CAN BE, NOT DERELICTION OF RESPONSIBILITY BUT ONE OF ALLOCATING AND SCHEDULING DOLLAR AND PERSONNEL RESOURCES OVER A WIDE VARIETY OF REQUESTS. BUT DIFFICULTY OF PROBLEM RESOLUTION SHOULD NOT, AND DOES NOT HAVE TO, LEAD TO A PUBLISHER COP-OUT. IN THE FIRST PLACE, SOME CONTINUING RESEARCH

FUNDS CAN BE MADE AVAILABLE; SECONDLY, THERE ARE A VARIETY OF WAYS IN WHICH USER RESEARCH CAN BE AIDED, CONSIDERED IN RELATIONSHIP TO INSTRUMENT MODIFICATION, AND DISSEMINATED TO OTHER USERS. IT SEEMS TO ME THAT THIS PUBLISHER FUNCTION IS MORE AND MORE IMPORTANT AS TIME GOES ON. TODAY'S SCHOOLS, FOR EXAMPLE, ARE FACED WITH CHARGES THAT COMPENSATORY EDUCATION IS A FAILURE AND THAT WHAT HAPPENS IN THE CLASSROOM DOESN'T MAKE MUCH DIFFERENCE IN SCHOOL ACHIEVEMENT. IT MAY WELL BE THAT MANY OF THESE CONCLUSIONS ARE MORE CLOSELY RELATED TO THE INSTRUMENT USED THAN THE TREATMENT EMPLOYED. WHO KNOWS? THE TIME SPAN BETWEEN THE RESEARCH QUESTION AND THE RESEARCH CONCLUSION IS OFTEN SO SHORT THAT I DESPAIR OF ANY MEANINGFUL ANSWER. BUT, IF THE PUBLISHER IS WATCHING, HELPING, COLLECTING, PONDERING, AND DISSEMINATING, A STABILITY OVER TIME CAN BE INTRODUCED WHICH SHOULD BEAR UPON LATER RESEARCH CONCLUSIONS. I THINK THAT IS RIGHT CONDUCT.

NOW I WANT TO CHANGE PACE AND SIMPLY LIST SOME REQUESTS WHICH HAVE COME TO US FROM TEST USERS:

> FOR THIS SPECIAL PRINTING DROP THE DON'T KNOW RESPONSE. IT'S KEEPING OUR KIDS FROM GUESSING AND IT MAKES OUR SYSTEM-WIDE AVERAGES LOOK TOO LOW IN THE NEWSPAPERS.
>
> HERE'S 2,000 DOLLARS OUT OF MY RESEARCH BUDGET. I DON'T KNOW WHAT TO DO WITH IT YET, BUT I HAVE TO SPEND IT BEFORE X DATE. HOLD IT FOR ME, THEN WE'LL FIGURE OUT SOMETHING.

PLEASE SEND ME 50 PACKAGES OF FORM P FOR USE IN EARLY MAY. (TJF NOTE: THE CITY-WIDE PROGRAM USING THE SAME FORM WAS SCHEDULED FOR MID-MAY)

OUR KIDS DON'T KNOW HOW TO TAKE TESTS, SEND US SOME PRACTICE TESTS.

AS I LISTEN TO THESE REQUESTS, AS I LISTEN TO THE ATTACKS UPON "TESTING," AND AS I LISTEN TO THE DEBATE RAGING AROUND "ACCOUNTABILITY," I'M REMINDED OF A RECENT JAMES RESTON COLUMN IN THE NEW YORK TIMES HEADED "THE TYRANNY OF WORDS." IT HAS TO DO WITH SLOPPINESS IN DEFINITION LEADING TO SLOPPINESS IN THINKING. HIS CONTEXT IS POLITICS, NOT EDUCATION, BUT IT IS APPROPRIATE TODAY BECAUSE WHO AMONG YOU WOULD LIKE TO ARGUE THAT TESTS DON'T BECOME INVOLVED IN POLITICS? OR THAT SLOPPY THINKING IS FOREIGN TO EDUCATIONAL CIRCLES?

RESTON WRITES, THAT IN THE BRUTAL LANGUAGE OF POLITICS, NOT ONLY MEN AND WOMEN, BUT WORDS LOSE THEIR REPUTATIONS. HE CITES THE WORD "APPEASE" AS A CASUALTY OF THE SECOND WORLD WAR, AND THE WORDS "PERMISSIVE" AND "PERMISSIVENESS" AS VICTIMS OF A TIME WHICH HAS SEEN THEM COME TO MEAN A SLACKNESS OF HUMAN CHARACTER.

WHAT WE FIND IS THAT THEIR ORIGINAL MEANINGS HAVE BEEN PERVERTED. "APPEASE" MEANS "TO BRING PEACE, TO PLACATE, SOOTHE, TO SATISFY OR RELIEVE" -- IN OTHER WORDS TO DO WHAT EVERY SENSIBLE FAMILY DOES TO HOLD THINGS TOGETHER. BUT SINCE BRITISH PRIME MINISTER CHAMBERLAIN AT MUNICH BEFORE THE LAST WORLD WAR,

"APPEASEMENT" HAS COME TO MEAN MAKING DISHONORABLE CONCESSIONS TO EVIL MEN TO SAVE ONE'S HIDE FOR A LITTLE WHILE.

AND "PERMISSIVE," SAYS RESTON, IS GOING THROUGH THE SAME TRANSFORMATION FROM MEANING "LENIENT, TOLERANT, PERMITTING DISCRETION..." TO MEANING THE DEFIANCE OF ALL TRADITIONAL VALUES AND AN INVITATION TO MORAL AND POLITICAL CHAOS.

THE FOREGOING HAS RELEVANCE FOR US. "TESTING," WHICH SURELY MEANS THE GATHERING OF MEANINGFUL INFORMATION IN ORDER TO IMPROVE THE LOT OF INDIVIDUALS AND OUR SOCIETY, IS BY SOME NOW BEING CALLED "TYRANNICAL," "INSIDIOUS," "STULTIFYING." THE WORD "ACCOUNTABLE," WHICH MEANS "TO BE ANSWERABLE FOR" (SURELY SOMETHING TO BE SUPPORTED BY THOSE WHO BELIEVE IN A RESPONSIBLE SOCIETY), HAS BECOME A SEMANTIC BATTLEGROUND OVER WHICH A FIERCE STRUGGLE FOR CONTROL OF AMERICAN EDUCATION IS BEING WAGED. TO PARAPHRASE RESTON: IN THE BRUTAL LANGUAGE OF EDUCATION, NOT ONLY MEN AND WOMEN BUT WORDS LOSE THEIR REPUTATIONS. THIS FACES US WITH THE NEED TO BE AS PROFESSIONALLY PRECISE AS POSSIBLE IN AN EDUCATIONAL MILIEU WHICH IS EACH DAY BECOMING LESS AND LESS THE TIGHT LITTLE EDUCATIONAL WORLD WE HAVE KNOWN AND IS MORE AND MORE PERMEATED BY THOSE WITH LITTLE EDUCATIONAL TRAINING BUT STRONG EDUCATIONAL DESIRES.

AS SOMEWHAT OF AN AFTERTHOUGHT I SHOULD EXPLAIN I DON'T MEAN THE NIGGLING PRECISENESS THAT LED EDWIN G. BORING TO ACCUSE A COLLEAGUE OF "MEASURING TRIFLES WITH SANCTIMONIOUS PRECISION" BUT, RATHER, A PRECISENESS WHICH IS CRISP, YET RELATIVE TO THE NEEDS OF THE PARTICIPANTS OF THE MOMENT. FOR EXAMPLE -- WE HAVE THE AIRLINE CAPTAIN CALLING THE TOWER PRIOR TO LANDING IN ORDER TO ASCERTAIN THE RIGHT TIME. THE TOWER RESPONSE IS "WHAT AIRLINE ARE

YOU?" AND THE CAPTAIN, SOMEWHAT SURPRISED ASKS, "WHAT IN THE WORLD DIFFERENCE DOES THAT MAKE?" "WELL," SAYS THE TOWER, "IF YOU'RE TWA IT'S 1800 HOURS. IF YOU ARE AMERICAN, IT'S 6:00 P.M. IF YOU'RE WITH OZARK, THE LITTLE HAND IS ON THE 6 AND THE BIG HAND IS ON THE 12; BUT IF YOU ARE WITH SOUTHERN -- IT'S TUESDAY.

THIS, THEN, TAKES ME TO THE QUESTION OF PUBLISHER RESPONSIBILITY FOR THE PRECISENESS ABOUT WHICH I HAVE JUST SPOKEN. IN 1971 OUR SALESMEN CONSULTANTS (THESE ARE PEOPLE WITH AT LEAST A MASTERS DEGREE) AND OUR R&D PEOPLE (MASTERS OR DOCTORATE) CONDUCTED OVER 1000 TRAINING WORKSHOPS OR SESSIONS. THIS YEAR THE TOTAL WILL BE HIGHER THAN THAT. THESE ARE NOT SALES CALLS; THEY ARE EDUCATIONAL SESSIONS AND THE STUDENTS ARE PRIMARILY TEACHERS, COUNSELORS, AND ADMINISTRATORS.

WE HAVE INSTITUTED AN ENTIRELY NEW KIND OF TRAINING PROGRAM IN CONJUNCTION WITH OUR DIAGNOSTIC READING TESTS. IN THIS INSTANCE, WE RETAINED THE SERVICES OF A HIGHLY SKILLED UNIVERSITY TEACHER-TRAINER IN READING, AND THEN PUT TOGETHER A WORKSHOP IN WHICH OUR PEOPLE, WHO KNOW THE TEST WELL AND SHE, WHO KNOWS READING INSTRUCTION WELL, CAN JOIN FORCES. THIS JOINING OF FORCES HAS NOW RESULTED IN IN-SERVICE TRAINING MATERIALS WHICH TAKE COGNIZANCE OF THE FACT THAT MANY CLASSROOM TEACHERS NEED A REFRESHER COURSE IN HOW TO TEACH YOUNGSTERS TO READ.

USING OUR STAFF MEMBERS, WE ARE TEACHING A SPECIAL CLASS IN TEST AND MEASUREMENTS IN A NEW YORK CITY COMMUNITY DISTRICT FOR WHICH THE TEACHER-STUDENTS RECEIVE DISTRICT SALARY CREDIT. IT IS GOOD FOR US BECAUSE IT IS TOO EASY TO DRIFT INTO

ALOOFNESS FROM TEACHERS AND KIDS, AND IT'S GOOD FOR THEM BECAUSE WE CAN TAILOR THE COURSE TO THEIR DISTRICT NEEDS.

WE HAVE JUST RECENTLY COMPLETED A TRAINING SESSION OVER CLOSED-CIRCUIT TV FOR 1400 TEACHERS IN PITTSBURGH. IT IS YET TO BE EVALUATED BUT I SUSPECT MOST OF THOSE TAKING PART WILL SAY IT WAS HELPFUL. OUR STAFF WILL BE LOOKING AT THE VIDEO TAPE IN THE NEXT WEEK OR SO.

WE HAVE AUTHORS CONDUCTING TRAINING SESSIONS, STAFF AND AUTHORS WORKING ON TRAINING PACKAGES INCLUDING AUDIO-VISUAL APPROACHES, AND STAFF WRITING AND DISSEMINATING PUBLICATIONS SUCH AS "MEASUREMENT PROBLEMS ASSOCIATED WITH ACCOUNTABILITY" AND "HOW TO REPORT ACHIEVEMENT TEST SCORES TO THE COMMUNITY." WE WOULD EVEN TRY SKY WRITING IF WE THOUGHT IT WOULD DO ANY GOOD.

THE TRUTH OF THE MATTER IS THAT WE AREN'T DOING ENOUGH AND NEVER WILL BE ABLE TO DO ENOUGH. THE TASK IS TOO LARGE FOR US -- WE DON'T SEE EVERYONE WHO USES TESTS AND EVERYONE WHO USES TESTS DOESN'T WANT TO SEE US; THEY MAY PREFER THEIR LOCAL UNIVERSITY, TEST CONSULTANT, OR ANOTHER TEST PUBLISHER. AND THE RESULTS OF NOT DOING ENOUGH ARE EVERYWHERE. THE REPORT OF THE TENTH NATIONAL CONFERENCE ON CIVIL AND HUMAN RIGHTS IN EDUCATION (THE CONFERENCE AT WHICH THE NEA TESTING MORATORIUM FIRST SURFACED) INDICATES A GREAT NEED FOR THOSE PROFESSIONALLY TRAINED IN EDUCATIONAL MEASUREMENT TO TAKE PART IN SUCH DELIBERATIONS. THE ACCOUNTABILITY CONCEPT WITH ALL THE ANXIETY

IT GENERATES IS ANOTHER MOVEMENT WHICH HAS RAISED QUESTIONS ABOUT TESTING WHICH LARGELY GO UNANSWERED AT THE LOCAL SCHOOL LEVEL. TOO MANY PEOPLE ARE CALLED UPON TO KNOW THINGS ABOUT TESTING WHICH THEY SIMPLY DO NOT KNOW.

SO WHAT IS A TEST PUBLISHER'S RESPONSIBILITY IN ALL THIS -- WHAT IS THE COURSE OF RIGHT CONDUCT? I WOULD SAY IT IS TO SPEAK OUT AS I AM DOING NOW AND CALL FOR A STUDIED APPROACH TO THE PROBLEM OF IGNORANCE ABOUT TESTING. IGNORANCE ON THE PART OF THOSE WHO ARE LITERALLY FORCED TO USE THEM -- WHETHER BY THE STATE, THE COMMUNITY, THE SCHOOL, OR BY THEIR OWN NEED-TO-KNOW FOR BETTER CARE AND INSTRUCTION OF CHILDREN. I ASK NCME, AMEG, AND AERA TO ADDRESS THEMSELVES TO THIS PROBLEM. I ASK THOSE OF YOU WHO HAVE COLLEGE OR UNIVERSITY RESPONSIBILITIES FOR THE TRAINING OF TEACHERS, COUNSELORS, AND ADMINISTRATORS TO CONSIDER WHAT CAN AND SHOULD BE DONE. SHOULD MORE EVALUATION AND MEASUREMENT COURSES BE TAUGHT; SHOULD THEY BE CONDUCTED, NOT AS A SEPARATE SUBJECT, BUT AS A BODY OF KNOWLEDGE AND A SERIES OF TECHNIQUES IMBEDDED IN THE COURSE OF INSTRUCTION IN A SKILL OR CONTENT AREA? (THIS IS THE APPROACH THAT SOME TEACHER TRAINERS IN READING HAVE USED SO EFFECTIVELY.) SHOULD NIE CONSIDER THE SOLUTION OF THIS PROBLEM AS ONE OF ITS GOALS? WHATEVER THE SOLUTION, AND THERE MUST BE ONE, IT IS VERY EVIDENT THAT SOPHISTICATION IN TEST DEVELOPMENT FAR OUTSTRIPS SOPHISTICATION IN TEST USE. THIS IMBALANCE IS UNHEALTHY AND EVEN DANGEROUS.

HOW SHALL WE CORRECT THE IMBALANCE? IT IS RIGHT CONDUCT THAT WE DO SO.

REFERENCES

1. Brazziel, William F. "Bias in Testing." Invited address given at the Tenth National Conference on Civil and Human Rights in Education, Washington, D.C. February 18-20, 1972.

2. Nader, Ralph. Speech before New Jersey Education Association Annual Convention, November 2, 1972.

3. Personal correspondence from High School Principal to President of HBJ., November, 1972.

4. Fitzgibbon, Thomas J. "The Ethical and Legal Position of the Counselor in Divulging Test Information." Measurement and Evaluation in Guidance. Spring, 1968.

5. Guidelines for the collection, maintenance and dissemination of pupil records. Convened by Russell Sage Foundation at Sterling Forest Conference Center. Sterling, N.Y. May 25-28, 1969.

6. "Standards for Educational and Psychological Tests and Manuals." (prepared by a joint committee of APA, AERA, and NCME) American Psychological Association, Inc., Washington, D.C., 1966.

7. Reston, James, "The Tyranny of Words." New York Times, November 23, 1972.

Ethical Issues in the Use of Humans for Research

W. L. Bashaw

University of Georgia

Introduction

We are now in an era of great public concern about and awareness of ethical issues in human research. There have always been outstanding researchers who have carefully considered the rights and welfare of their human subjects, but much famous research indicates that many researchers consider as clever technique what other researchers consider to be an affront to human dignity and welfare. Let's look at some of these famous examples to get a feel for the problem:

Tuskegee Project (recent newspaper reports) -- In which cures for syphillis were withheld from experimental subjects so that the long range effects of syphillis could be studied.

Birth Control (See Welsh, Educational Researcher, 1972) -- In which placebos were given instead of birth control pills, resulting in unwanted pregnancies.

Rosenthal and Jacobsen (Pygmalion in the Classroom) -- In which teachers were given false information concerning the abilities of their children and such false information was predicted to change inappropriately the teachers' methods of handling specific children.

Hartshorn and May (Studies in Deceit) -- In which children were placed in situations encouraging them to be deceitful.

My concern for ethical issues in research is relatively new. It has developed only in the last few years. The APA draft of "Ethical Standards for Research with Human Subjects", first published in July, 1971, for membership critique, spurred my concern to the point of including ethical considerations as a topic in almost all of my classes. Thus, this document has had at least one positive effect. At least, I am convinced that the effect

is positive and I will continue to encourage my colleagues and my students to be concerned with ethical considerations in their choice of research problems and methodologies. I will review the specific principles presented in the APA Guidelines.

More recently, I have been appointed to the Research Ethics Committee of University of Georgia. Our purpose is to screen proposals before they are approved by the Institution for submission to HEW. I will discuss the University of Georgia policy.

I have gone through my own textbook collection to see what standard research texts for educators say about ethical matters. This survey will be discussed.

Finally, there exists a wide research literature on specific research methodologies involving ethical matters such as subject expectations, deceit, volunteerism, reactive effects of experimentation, and reactive effects of measurement. I will briefly discuss points made in some of these papers.

The APA Principles

Introduction

The APA Ethical Principles are based on empirical evidence of projects that gave rise to ethical problems. In my own history, I have had a few such opportunities to consider the ethics of particular research projects. In one of my uncompleted projects I did what you all do -- I asked for permission to gather data in a school. No one asked the children if they wanted to participate. A promise to me that teachers would act as monitors was not kept. It was clear to me that I was not getting student cooperation -- they were enjoying the tasks too much for me to believe that they were seriously performing, so I discarded the data without looking at it. What issues were involved?

Certainly I could have analyzed, interpreted, and published, but chose not to do so. But moreover, the biggest issue, as I understood it then and now, was the selection of a sample by coercion, a sample that did not know that research subjects are supposed to play a role for science. They knew only that no one was watching who could influence their school records. As a student researcher, I served as monitor on other similar projects -- similar in that subjects seemed to reach exhaustion points in their willingness to continue to serve science. To fail to try was the only way that they were allowed to express hostility towards coercive participation. One wonders how many students in how many projects not only fail to try assigned tasks, but actually fake responses in an attempt to create trouble for the researcher. (This is not rare, as we will see later in this paper.)

Students are quick to identify bad research. I once had a neighbor who said that her graduate class was involved in a study. She was supposed to respond to stimulus cards, but could not see them clearly due to the manner in which they were being held by the researcher, so she marked randomly. Yet, University policy invariably requires that all faculty do research, even if the use of the time of human beings (as subjects) is wasted on incompetent techniques.

Another personal example involved my younger brother. He was an undergraduate psychology student who participated in a pain threshold study conducted by a psychologist who is famous for his work in pain thresholds. My brother possibly didn't know what "threshold" meant, and he does have a hearing loss, but in any case, he did not understand the directions. He has a burn scar on his back, because, as an All-American type boy, he thought the task was to see how much heat one could stand. So he gritted his teeth for

science. What was the responsibility of the institution in this case? Free medical service at the University clinic, of course.

The Development of the Principles

The development of the APA statement of ethics is rather unique. Most such statements are developed by a panel of senior experts and formally adopted as policy for an organization. APA chose to do it the slow, hard way by encouraging full participation of its membership. A guarantee of relevance to real experience was possible by the basing of the Standards on reported incidents involving questionable methods. The responsibility for the Standards was given to an ad hoc Committee on Ethical Standards, chaired by Stuart W. Cook. The sample of APA members who were asked to describe such incidents consisted of 18,000 psychologists (67% of APA membership). These problems were studied and organized and formed the basis of the July 1971 APA Monitor report. This report was presented to the membership with the request that it be carefully studied by researchers and students who were to provide the committee with critical feedback. Indeed, critical feedback was obtained. Opinions ranged over all extremes. The draft Standards were drastically revised and the revision appeared in the May, 1972, APA Monitor. Membership opinion was also obtained in open meetings at both the 1971 and 1972 APA Annual Meetings. The final, approved version appeared in the January, 1973, APA Monitor.

A Review of the Standards

Let's begin our brief review of the APA Principles by quoting the purpose of them as stated in the May Monitor:

"The purpose of this examination of ethical considerations and principles that arise in the actual course of planning and conducting psychological research with human beings is to promote explicit attention to these

ethical issues in their full complexity, and thus to foster responsible decisions." (APA Ethical Standards, p. II).

The ten principles make-up of one-third of one of the nineteen pages of newsprint required to present fully the committee's report. The bulk of the report is an elaboration and defense of these ten principles along with examples of problems that have occurred related to each. The ten principles are quoted fully in an appendix to this paper.

You might find my paraphrasing of the ten principles to be useful:

1. The investigator is responsible for making decisions relevant to ethical issues. If he ever believes he must deviate from one of the APA principles, he should seek advice from others and insure tight safeguards.
2. The investigator is responsible for what happens.
3. The research must be explained to the subject, to the extent of answering any questions the subject raises. Confidentiality and the protection of the welfare and dignity of the subjects must be maintained.
4. Deception must be followed by an adequate explanation to the subject so that the subject understands the reasons for the deception and has his faith in behavioral research somewhat restored.
5. The subject has the right to decide if he wishes to participate and has the right to drop out, if he wishes.
6. There should be a clear contract between the researcher and the subject. Promises should be kept.
7. Subjects should not be subjected to any form of disconfort, harm, danger, or mental stress except if they were well informed of risks and have agreed to participate, and if the researcher is taking special precautions.

8. After the study, subjects should be informed as to what was done and why. Misconceptions must be cleared up.
9. The researcher is responsible for the consequences for the participant, including long-range consequences.
10. Data should be kept confidential. The possibility of a loss of confidentiality must be explained prior to the study. (Note: a researcher must yield data to the courts.)

The University of Georgia Policy

The Role of Government

The role of the government in ethical practice is also relatively new. A few years ago, while I was on the University of Georgia' Research and Development Center staff, we received a directive on testing. If we needed any data not part of the regular school program, then the USOE demanded the right to approve such data requests.They were specifically concerned with issues of rights-to-privacy and freedom from embarrassment. Harry Anderson, my R & D colleague, decided to screen instruments empirically. A behavioral maturity rating scale developed for R & D Center studies by Yung Ho Kim, Anderson's student, was submitted to a panel of teachers who rated each question on three criteria: (1) Do you know your children well enough to respond accurately? (2) Will the information embarrass the child or his family? And (3) Does the question violate the child's right-to-privacy? On the basis of this information, some items were deleted from the scale before it was ever used for rating children. This story has two points. One is that the researcher can develop ways of handling some ethical issues. The second was that the USOE had decided at that time to police research at the Washington level.

More recently, HEW has decided to get partially out of the ethics-watch-dog business. The primary screening task has been passed down to the institutions who have the responsibility of screening research proposals for ethical problems before they are submitted to Washington. Thus, a local panel of reviewers is selected to critique proposals before the institution releases them. Washington has taken the leadership in protecting the public from ill-advised research techniques, while at the same time, allowing local institutions to make the judgments.

University of Georgia Policy

Let's look at the University of Georgia policy. To my knowledge, this institution has taken no steps to extend its review system to projects not supported by Washington. University of Georgia has publicly admitted its responsibility for what its faculty does in all research projects, and is committed to ". . . provide for the facilities and professional attention required for subjects who may suffer physical, psychological, or other injury as a result of participation in an activity." It is clear, to me at least, that accepting this commitment for HEW projects must be interpreted as accepting it for any University-sponsored project. It is doubtful that such a commitment can be met without introducing more general screening of projects not supported by HEW.

It is clear that the University of Georgia policy does not cover all of the problem areas listed in the APA draft guidelines, and many statements in it are far from being defined clearly. At present, the definition is left up to the panel members reviewing a proposal. The thirteen principles are paraphrased by me below.

1. The research should be well designed.
2. The researcher should be competent.

3. The value to science must exceed the risk to the subject.

4. Personality manipulation is dangerous.

5. Rights to privacy and confidentiality must be maintained.

6. The researcher is responsible for what happens.

7. The project and risks must be known by the subjects, or the researcher had better have a good explanation of why not.

8-9-10. Free consent must be obtained, and if possible, in writing.

11. The personal integrity of the subject must be respected.

12. Subjects must be able to withdraw.

13. The researcher should stop if danger to the subject is apparent.

The actual University of Georgia policy statement appears as an appendix at the end of this paper.

It would be nice if I could tell you stories about the projects I have reviewed as a committee member. However, I have reviewed only one. Whether this is due primarily to the shortage of funds for grants, or some other cause, I cannot say. The one proposal I have seen involved deceit. The proposing investigator had obviously prepared his proposal with the University of Georgia policy in his hand. The committee reviewing this proposal did require that some additional safeguards be added.

Research Textbook Survey

One of my concerns is whether or not ethical issues are handled adequately in research training. The degree to which textbooks reflect concern with research ethics is evidence that this topic is probably handled badly in formal research course instruction. Only a handful of available texts discuss ethics as a specific topic. Some have relevant related topics,

but do not discuss specifically ethics as such. Some discuss design problems in a way that suggest questionable procedures as if they were clever techniques.

I could find only a few standard texts including specific topics on ethics and each has only a few pages on the topic. One book of readings contains three papers dealing specifically with ethics (38 pages) and other papers that certainly deal with ethical issues, although not specifically.

The few books having sections on ethical consideration were members of the population of research methods texts that I possess. On the negative side, 12 of 17 of my collection make no specific reference to ethical and unethical research practices.

As mentioned previously, some of these do have content that can be considered as ethical considerations, if handled as such by the instructor. In any case, it is evident that instruction in ethics will need to come from the instructor, rather than the text, if teachers of research methodology are to influence future researchers in regard to ethical considerations in research.

I must point out that most texts do deal with some questions of ethics -- those that are rather obvious, such as plagarism, honesty in reporting, integrity in analysis and interpretation, and the obligation to conduct sound research.

Research Related to Ethics

There is a considerable literature on research artifacts associated with the ethical principle so far discussed. It is interesting to consider, however, the difficulty in studying the effects of unethical practice without engaging in the same! For example, how can you study the effects of deceitful research practice without being deceitful in order to set up the experimental situation? Perhaps you could follow around behind another researcher who is

Known for d[illegible] practices, and thus, get a n

known for deceitful practices, and, thus, get a natural situation to study! For another example, how can you study the effects of volunteerism vs. coercion without having a coerced group? In spite of these difficulties, perhaps the gains from such research are important enough to allow departure from strict interpretation of the APA Standards.

I have provided a brief bibliography that includes some relevant research reviews. The Rosenthal and Rosnow book is perhaps one of the best sources of research artifacts, just as the famous Webb, et al. Unobtrusive Measures book is one of the best sources of information on measurement artifacts.[1] Both of these works do discuss ethical considerations. Rosenthal and Rosnow do in depth, while Webb, et al., do only in their preface. Both suggest and recommend procedures that we might judge to be highly unethical practices; as well as contribute information that might lead to more ethical practices.

The Weber and Cook (1972) review illustrates the various roles that a subject assumes. These include the "apprehensive subject", the "good subject", the "negativistic subject", and the "faithful subject". A "Faithful Subject" is one who believes docility is required. He believes he must scrupulously follow directions. There are two versions of the faithful subject -- a passive and an active version. The passive version is one who assumes that he should be uninvolved, he apathetically follows instructions. The active version assumes he must contribute to science and be extremely honest so that conclusions will be valid. The "Good Subject" is one who wants to confirm the researcher's hypothesis. The "Negativistic Subject" is one who

---

[1] Katz (1973), was received after the presentation of this paper. It is highly recommended as possibly the best source document for ethical problems with human subjects.

wishes to disprove t
tility at being expl
Subject" is one who
All of these ro
reflects the anxiety
anxiety can evidentl
to the subject, whic
Incidentally, this a
threat to the validi
and Cook also presen
exerts strong influe
not include a stateme
istic research as an
subjects obviously w
Principles will some
Oakes (1972) re
subjects. He conclu
is the major problem
these differing popu
with type of subject.
as any other, it is j
others. This again,
plement laboratory st
A more complete
this presentation. P
terest in some of the
the ethics of researc

the researcher's hypothesis (due possibly to his hos-
ploited or controlled by others). The "Apprehensive
o is anxious about the use of data to evaluate him.
roles are relevant to this paper. The apprehensive role
ty associated with the fear of being evaluated. This
tly be alleviated somewhat by having the research explained
ich is what the APA ethics Principles say one should do.
anxiety interacts with treatment, thus posing an internal
dity of the research as well as an external one. Weber
ent evidence that the knowledge of the research hypothesis
uence on the subject, so the pre-experiment briefing should
ement of expectations. The authors call for more natural-
in answer to these concerns. The "good" and "negativistic"
will contaminate conclusions. Perhaps the use of the
newhat alleviate these effects of roles.
reviewed the research on the use of volunteers and coerced
ludes, in general, that external validity (generalization)
m here. Highly contradictory evidence is obtained from
pulations. Here exist obvious interactions of effects
t. He points out that the use of any type is just as valid
just that results from one group do not generalize to
, suggests a concern for realistic field studies to com-
studies.
e review of the relevant literature is impossible for
Perhaps these few cited will lead to your increased in-
he problems that are caused, or uncovered, by considering
rch with humans.

Recommendations

I would like to conclude by listing some of the questions to which our attention might be given.

1. Who has the right to place students in a research project -- the principal, the teacher, the parent, the child?

2. Has our history of researcher-deceit created an environment of complete distrust in the directions presented by the researcher? Donald Campbell (in Rosenthal and Rosnow, p. 372) states clearly the concern over our long history of deception of human subjects: ". . . we are doomed to wear out our laboratories. For this reason, we are leaving the college in favor of the high school, the grammar school, and the street. Publicity will eventually contaminate these laboratories, too. . . and public anger over deception (will) not be reduced by debriefing. . ."

3. APA guidelines state that the subject must benefit for participation. What is the reward to the public school subject? You can't argue that participation in the research is educational, as is the case with psychology sophomores. Is it sufficient to reward the principal with the promise of greater school-university shared benefits?

4. Can any research be planned to avoid the need for deception? We need to develop new, creative paradgyms rather than assuming old deceit models are necessary. (Eg., there is some evidence that role-playing yields similar results to normal experiments, i.e., pretend that you have just received an IQ test score for a child that indicates an IQ 20 points lower than you believed it would be. How do you now react to this child?)

5. Is it possible that some research is unnecessary from the point-of-view that any conclusion will lead to no change in behavior? (E.g., if one can show that protein deprivation does not lead to lowered intelligence, will

we quit trying to feed poor children? This example is from a real proposal to study the intelligence of starving African children by conducting a true experiment.)

6. Does the university dissertation requirement and publish-or-perish policies lead to enormous wasting of human subjects' time in poor research? (Ans.: Yes). Isn't this an ethical issue? No subject should be asked to give time and effort to a study that is not well defined and competently planned.

So, in summary, let me point out that statements of ethical principles, if followed, do lead to interesting methodological problems. However, these same principles, when considered carefully, also uncover methodological problems that were always there. We need new designs, fresh approaches, a higher concern in having sound designs, and a concern that results of research will be useful. But the difficulties should not stop competent research. As the APA guidelines state: "Since research is the scientist's distinctive way to contribute to human welfare, he has an obligation to pursue it to the best of his ability." (APA Ethical Standards, draft, p. II). And in another place, "The first and most important ethical obligation of the researcher -- to conduct meaningful research." (APA Ethical Standards, draft, p. IX).

Selected References

1. Carlson, Rae. Where is the person in personality research? Psychological Bulletin, 75, 1971, 203-219.
2. Ethical standards for psychological research (proposed principles). APA Monitor, 2, July 1971, 9-28.
3. Ethical standards for research with human subjects (draft). APA Monitor, 3, May 1972, insert following p. 3.
4. Katz, Jay. Experimentation with Human Beings. N.Y.: Russel Sage Foundation, 1973.
5. Kelman, Herbert C. The rights of the subject in social research: an analysis in terms of relative power and legitimacy. American Psychologist, 27, 1972, 989-1016.
6. Oakes, William. External validity and the use of real people as subjects. American Psychologist, 27, 1972, 959-962.
7. Revised Ethics Principles adopted. APA Monitor, 4, January 1973, 2.
8. Rosenthal, Robert and Rosnow, Ralph L. Artifact in Behavioral Research. New York: Academic Press, 1969.
9. Weber, Stephen J. and Cook, Thomas D. Subject effects in laboratory research. Psychological Bulletin, 77, 1972, 273-295.

Ethical Principles in the Conduct of Research with Human Participants

(APA Monitor, January, 1973)

The decision to undertake research should rest upon a considered judgment by the individual psychologist about how best to contribute to psychological science and to human welfare. The responsible psychologist weighs alternative directions in which personal energies and resources might be invested. Having made the decision to conduct research, psychologists must carry out their investigations with respect for the people who participate with concern for their dignity and welfare. The Principles that follow make explicit the investigator's ethical responsibilities toward participants over the course of research, from the initial decision to pursue a study to the steps necessary to protect the confidentiality of research data.

1. In planning a study the investigator has the personal responsibility to make a careful evaluation of its ethical acceptability, taking into account these Principles for research with human beings. To the extent that this appraisal, weighing scientific and humane values suggests a deviation from any Principle, the investigator incurs an increasingly serious obligation to seek ethical advice and to observe even more stringent safeguards to protect the rights of the human research participants.

2. Responsibility for the establishment and maintenance of acceptable ethical practice in research always remains with the individual investigator. The investigator is also responsible for the ethical treatment of research participants by collaborators, assistants, students and employees, all of whom, however, incur parallel obligations.

3. Ethical practice requires the investigator to inform the participant of all features of the research that reasonably might be expected to influence

willingness to participate, and to explain all other aspects of the research about which the participant inquires. Failure to make full disclosure gives added emphasis to the investigator's abiding responsibility to protect the welfare and dignity of the research participant.

4. Openness and honesty are essential characteristics of the relationship between investigator and research participant. When the methodological requirements of a study necessitate concealment or deception, the investigator is required to ensure the participant's understanding of the reasons for this action and to restore the quality of the relationship with the investigator.

5. Ethical research practice requires the investigator to respect the individual's freedom to decline to participate in research or to discontinue participation at any time. The obligation to protect this freedom requires special vigilance when the investigator is in a position of power over the participant. The decision to limit this freedom gives added emphasis to the investigator's abiding responsibility to protect the participant's dignity and welfare.

6. Ethically acceptable research begins with the establishment of a clear and fair agreement between the investigator and the research participant that clarifies the responsibilities of each. The investigator has the obligation to honor all promises and commitments included in that agreement.

7. The ethical investigator protects participants from physical and mental discomfort, harm and danger. If the risk of such consequences exists, the investigator is required to inform the participant of that fact, secure consent before proceeding, and take all possible measures to minimize distress. A research procedure may not be used if it is likely to cause serious and lasting harm to participants.

8. After the data are collected, ethical practice requires the investigator to provide the participant with a full clarification of the nature of the study and to remove any misconceptions that may have arisen. Where scientific or humane values justify delaying or withholding information, the investigator acquires a special responsibility to assure that there are no damaging consequences for the participant.

9. Where research procedures may result in undesirable consequences for the participant, the investigator has the responsibility to detect and remove or correct these consequences, including, where relevant, long-term after-effects.

10. Information obtained about the research participants during the course of an investigation is confidential. When the possibility exists that others may obtain access to such information, ethical research practice requires that this possibility, together with the plans for protecting confidentiality, be explained to the participants as a part of the procedure for obtaining informed consent.

The University of Georgia/Office of the Vice President for Research/
Graduate Studies Research Center/Athens, Georgia 30601

PART ONE OF A GENERAL INSTITUTIONAL ASSURANCE

The University of Georgia will comply with the policy for the protection of human subjects participating in activities supported directly or indirectly by grants or contracts from the Department of Health, Education, and Welfare. In fulfillment of its assurance:

The University of Georgia will establish and maintain a committee competent to review projects and activities that involve human subjects. The committee will be assigned responsibility to determine for each activity as planned and conducted that:

The rights and welfare of subjects are adequately protected.

The risks to subjects are outweighed by potential benefits.

The informed consent of subjects will be obtained by methods that are adequate and appropriate.

The University of Georgia will provide for committee reviews to be conducted with objectivity and in a manner to ensure the exercise of independent judgement of the members. Members will be excluded from reviews of projects or activities in which they have an active role or a conflict of interest.

The University of Georgia will encourage continuing constructive communication between the committee and the project directors as a means of safeguarding the rights and welfare of subjects.

The University of Georgia will provide for the facilities and professional attention required for subjects who may suffer physical, psychological, or other injury as a result of participation in an activity.

The University of Georgia will maintain appropriate and informative records of committee reviews of applications and active projects, of documentation of informed consent, and of other documentation that may pertain to the selection, participation, and protection of subjects and to reviews of circumstances that adversely affect the rights or welfare of individual subjects.

The University of Georgia will periodically reassure itself through appropriate administrative overview that the practices and procedures designed for the protection of the rights and welfare of subjects are being effectively applied and are consistent with its assurance as accepted by the Department of Health, Education, and Welfare.

Date:_____ Robert C. Anderson

/_____

Vice President for Research

Enclosure: Implementing Guidelines, Part Two of a General Institutional Assurance

UNIVERSITY OF GEORGIA

Policy

The policy stated below is designed to protect the rights and welfare of human subjects involved in research and training programs in this University. The policy applies to all investigations and training programs in the health sciences and social sciences which involve the use of human subjects in any way.

1. Research involving human subjects must conform to the moral and scientific principles that justify such research and should be based on scientifically established facts.

2. Research involving human subjects should be conducted only by scientifically qualified persons and under the supervision of a senior scientist.

3. Research involving human subjects should not be carried out unless the importance of the objective is in proportion to the inherent risk to the subject.

4. Special caution should be exercised by the scientist in performing research on human subjects in which the personality of the subject may be altered or disturbed by experimental procedures.

5. Research and training programs involving human subjects must respect the rights of privacy of subjects and assure that maximum confidentiality of personal information will be maintained.

6. In the purely scientific application of research carried out on human subjects, it is the duty of the investigator to protect the life, health and welfare of the subjects, within the context of the research situation.

7. The nature, the purpose and the risk of the research must be explained to the subject by the investigator. If, for any reason, the nature and purpose of the research cannot be explained to the subject, the investigator must demonstrate the necessity of this approach to a committee of his peers, as well as indicate the provisions for safeguards that may be needed.

8. Research on a human subject must not be undertaken without his free consent after he has been informed of the risks involved.

9. The subject of research should be in such a mental, physical and legal state as to be able to exercise fully his power of choice and his grant of consent.

10. Consent should, as a rule, be obtained in writing. However, the responsibility for the care and protection of subjects in research always remains with the research worker; it never falls on the subject even after consent is obtained.

11. The investigator must respect the right of each individual to safeguard his personal integrity, especially if the subject is in a dependent relationship to the investigator.

12. At any time during the course of research the subject or his guardian should be free to withdraw consent for the research to be continued.

13. The investigator or the investigating team should discontinue the research if in his or their judgment, it may, if continued, be harmful to the individual.

Procedures

1. Individual investigators should request the appointment of a committee of institutional associates prior to beginning any investigation involving human subjects in which a grant or an award has been made by the U.S. Public Health Service. The Vice President for Research will then appoint a committee of three or more institutional associates of the investigator from the membership of a panel of such associates. This panel shall consist of two clinical psychologists, two social psychologists, three sociologists, four political scientists, two social workers, two professional counselors, two biological scientists, a lawyer, and four other faculty members from other disciplines, all of whom, if possible, shall be tenured members of the Graduate Faculty. The Vice President for Research may, when unique competency is needed, supplement the panel with other persons.

2. The investigator will submit to the committee his plans for protecting the rights and welfare of the individuals involved, assuring the use of appropriate methods to obtain informed consent, and with regard to the risks and potential medical benefits of the investigations, and his plans for making certain that facilities and professional attention necessary for the health and safety of the subjects have been or will be adequately taken care of. The report of the committee will be forwarded to the Vice President for Research for his approval.

3. At the time of the initial meeting of the committee for each project, the committee will decide what appropriate methods of further review will be necessary, and what records of group review and decision and of informed consent will be necessary, and make a recommendation on those matters as a part of their report to the Vice President for Research. In each project, however, there will be at least one additional meeting of the committee during the course of the investigation, at which time a review of the procedures of the investigator, insofar as they relate to the rights and welfare of the human subjects, will be made. A report of that meeting will be furnished the Vice President for Research, including documentary evidence of consent by the subjects of the investigation. Following any and all meetings of the committee, a report (including the advice of the committee to the investigator) will be furnished the Vice President for Research, and a copy of that report will be furnished the investigator.

4. If any changes in investigative protocol which may affect the subjects are contemplated, the investigator will call the proposed changes to the attention of the committee before they are made. The investigator will also call to the attention of the committee any emergent problems of investigation which may affect the subjects when such problems occur. The chairman of the committee will call a special meeting of the committee in either of these events, and the advice of the committee will be given to the investigator and transmitted to the Vice President for Research for his approval.

5. The investigator will give assurance to the Vice President for Research that he is following or will follow the advice of the committee.

ALTERNATES TO PSYCHOLOGICAL TESTING

or

HOW WE'VE TURNED A SILK PURSE INTO A SOW'S EAR[1]

Donald N. Bersoff

University of Georgia

In mid-1972 two staff members of Hunter College in New York were denied continued employment because they failed their civil service examinations. One of the test takers was an audio-visual technician, the other a graphic artist. Both had been functioning well in their positions. Both were women. They flunked the test because they could not lift a 25-pound barbell over their heads nor successfully maneuver through an obstacle course. In an editorial that could have been written by Gagné or Glaser, the New York Times ridiculed the testing process which so absurdly rejected these employees by saying:

> Standards are best protected when the tests are geared with the greatest possible precision to the qualities needed for the particular job or activity. Testmakers who fail to comprehend this necessity should be notified that they have flunked their own basic test and should be reassigned to less sensitive duties, even if they can they can pass the weight-lifting requirement (New York Times, June 22, 1972, editorial page).

In another context, many psychologists, particularly those who are engaged in personality and intelligence assessment, have been spending more time in the courts lately than in the testing room. The education of psychologists involved in testing in school settings is more likely to be concerned with learning about such famous names as Hobson, Stewart, Spangler, the Guadalupe Organization, Covarrubias, Larry P. and Diana than it is with Anastasi, Binet, Terman, Cronbach or Thorndike.[2]

For forty years, beginning with World War I, psychological testing was perceived as the vehicle by which major decisions about people's lives were made in industry, the military, hospitals, mental health clinics, and the schools. Scores derived from psychometric instruments were used to classify, segregate, track, advance, employ, institutionalize and educate people. Now IQ testing is outlawed in San Francisco, personnel selection tests are declared illegal unless directly relevant to employment, group intelligence measures are banned in the New York City schools, a whole profession which has distinguished itself from psychiatry primarily because its practitioners can test has been declared moribund, and school psychologists in Boston have been declared incompetent. In the last ten years what was once a silk purse has been transformed into a sow's ear.

Who is to be held accountable for this psychological alchemy? The answer--two brands of psychos--psychoanalysts and psychometricians. Psychoanalysts are to blame because they have perpetrated a fraudulent (Freudulent?) theory of personality and perpetuated its myth. Psychometrists, the test constructors, are to blame because they have forgotten their historical antecedents and have become overly concerned with psychometric esthetics to the neglect of validity.

Psychoanalysts and other psychodynamically-oriented practitioners hold to a theory of response predispositions viewing behavior as relatively independent of the situation in which the person is immersed. Personality is thus perceived as sets of needs, drives, or trans-situational traits which initiate and guide behavior. These dispositions are considered to characterize a person and become translated into classificatory descriptions such as, "He is retarded," "She is schizophrenic," or "They are obsessive-compulsive neurotics." Because these traits are seen as underlying overt behavior, special instruments, primarily projective tests, are used to ferret

out the hidden dynamics concealed within. Projective testing continues despite the fact that the assumptions on which they are based have been challenged and demonstrated to have little substance (Murstein, 1965).

Modern theorists concerned with personality assessment (Epstein, 1966; Fiske & Pearson, 1970; Fulkerson, 1965; Mischel, 1968; Peterson, 1968; Wallace, 1966) have questioned the validity of psychological testing within a psychodynamic model and have recognized the "organism error" on which the tests have been interpreted. As Murstein (1965) has shown, test responses are not only a function of the characteristics of the organism to whom the test is administered, but also of the stimulus properties of the test and the background variables (such as setting, nature of examiner) in which the test is administered. It is because behavior is the product of the previously learned capabilities of the person and the specific stimulus situations surrounding it (Wallace, 1966) that tests administered in a typical clinical or classroom setting have limited validity in assessing or predicting overt behavior.

Psychometrists have been equally culpable in this regard. Peterson (1968) has described the situation accurately:

> Psychologists customarily study large samples of subjects, place them all in one or a very limited number of situations, and then exercise close restraint in generalizing to the larger subject population, while they make irresponsibly loose generalizations to a larger universe of situations with perfect serenity (p. 32).

Test developers, it seems, have lost sight of their original purpose, to gather samples of behavior which can be used to differentiate individuals, and instead have subjected their tests to more and more elaborate statistical manipulation, at the same time stressing temporal economy and ease of administration and interpretation. But, as Wesman (1968) has stated,

"Efficiency is certainly desirable--but validity is crucial. How tests were constructed is interesting and even germane; how they work is the critical issue (italics in the original) (p. 272)."

Newer conceptions of validity assert that no method of instrument of assessment is invalid a priori. The validity of a measuring device must be evaluated in terms of the purpose for which it is intended. Thus, in comparing the validity of assessment procedures the purpose of assessment should be specified first. Increasingly, with the rise of behaviorally-oriented approaches to therapeutic intervention and instructional evaluation, the goal of assessment is seen as the acquisition of relevant information that will contribute to decisions about desired changes in behavior (Bijou & R. Peterson, 1971; D. Peterson, 1968; Wittrock, 1970). As Peterson (1968) states, "The only legitimate reason for spending time . . . in assessment is to generate propositions which are useful in forming decisions of benefit to the persons under study (p. 32)."

Within that definitional framework, psychological tests are generally not helpful in the acquisition of knowledge relevant to the construction of intervention strategies for those assessed. This indictment is especially true as these strategies concern instruction and its evaluation. Three statements by widely known authors involved in the evaluation of instruction will serve to make a predominating judgment clear:

Robert Glaser (1970)--"Standard psychometrically developed tests, as a result of the way in which they have been validated and evaluated. . . lack utility. . .with respect to instructional decisions (p. 77)."

Robert Gagné (1970)--"Dependence upon imprecise techniques of mental testing for the measurement of concepts and principles has led to the ignoring of the requirements for control procedures that have come to be

the standard features of measurement techniques used with other simpler kinds of learning outcomes. The single item used in traditional achievement testing constitutes an uncontrolled, ambiguous measure that can only in rare instances be shown to be related directly to the learning outcome of interest (p. 123)."

M. C. Wittrock (1970)--"The student's scores on standardized tests of interests, abilities, and achievement. . . do not enable us to make rigorous inferences about what the students have learned nor about the role of environments and intellectual processes in producing the learning (p. 10)."[3]

Testimonials are not the most convincing scholarly evidence, however. Two items will be more pertinent: A discussion of what I have called the current functioning myth (Bersoff, 1971c), and a brief report of a study by Hollands & Richards (1965).

First, the current functioning myth. When psychologists report the results of an intellectual evaluation they usually write something like, "Billy's level of intelligence is within the 'Average' range of intellectual functioning," or "Millie is currently functioning within the 'Average' level of intelligence." When the term "functioning" is used in this regard it implies that the person called "Average" typically operates or performs (functions) at an average level when intellectual abilities are called for. However, what has been measured is functioning under optimal, not actual, conditions. Nevertheless, more often than not, the obtained score is translated into a measure of present expectation: "Thus, he/she can be expected to achieve at an average level in the classroom." These kinds of statements, made on the basis of an individually administered intelligence test with its implication of current functioning overlooks the

fact that test responses do not occur in an organismic vacuum.

It is very clear from the reading of the Stanford-Binet and Wechsler manuals that conditions are very much different when intelligence is assessed in the individual testing situation and in the environment in which the person actually functions. This is most evident in the assessment of children. The authors of the Binet and WISC manuals specifically instruct testers to establish rapport, to be sympathetic and understanding, to watch for fatigue, to minimize physical defects, and more, all in order to "maintain both high motivation and optimal performance during administration (Terman & Merrill, 1960, p. 50)." Intelligence testing of this sort is conducted in a one-to-one situation in which a carefully observing examiner is on the alert for situations or behaviors which will tend to lessen the evocation of optimal performance. Only rarely, if ever, does a similarly ideal condition exist in the real life situations in which the child will exhibit intelligent behavior.

In the functioning environment of the classroom conditions for the evocation of optimal performance are greatly reduced. There are more children, less individual observation, more numerous and frequently occurring stimuli that may be distracting, and less chance for the teacher to reinforce the child's appropriate intellectual behavior. In a study just completed (Rigsby, 1972), for example, it was found that school psychologists during administration of the verbal portion of the WISC emitted an average rate of 1.38 positive reinforcements per minute to 26 children tested individually while the teacher emitted less than one every ten minutes to these same children in a classroom instructional situation.

Thus, each child received more than 15 times as many reinforcements in a testing situation than he did in a classroom. In fact, the average reinforcement emission by a teacher to a whole classful of children was less than that for one child by a psychologist. This hardly makes the two situations comparable and extrapolation from one to the other very risky, if not absurd.

On a larger scale the Hollands & Richards (1965) study corroborates the conclusion that test information correlates poorly with real-life performance. In their particular investigation the American College Testing Program Test Battery, a conglomerate examination allegedly tapping academic potential, was administered to 7262 college freshman from whom they also obtained high school grade point averages. The test scores and GPA served as measures of academic accomplishment. Also collected were a variety of real world, nonscholastic measures of performance. For example, some of these data consisted of a record of extra-curricular activities in art, music, literature, and science. These included such accomplishments as winning a prize for scientific work, inventing a device that could be patented, or placing high in a national science contest. All the items in the various areas were said to represent behaviors requiring complex skills, persistence, and originality.

The median correlation between the measures labeled academic and nonacademic accomplishment was .04. Ninety percent of the correlations were between $\pm$ .15 and the highest correlation was only .23. Twenty-six percent of the correlations were actually negative. Thus, the overall relationship between academic potential as measured by a standardized test and real world accomplishment was virtually zero.

Given the errors we have made in assessment by relying on psychological tests (and I have in no way catalogued them all) what solutions are there? They have already been implied. First, we need to refocus our attention and concern ourselves with validity questions, especially as they pertain to situational influences on test behavior and test-score generalizability. Second, we need to move toward more idiographic approaches and away from the normative aims of testing. Norms derived from the mass evaluation of large groups tell us nothing about what evokes or perpetuates behavior in an individual. If there is one overriding advantage of an operant orientation it is its idiographic aim — the assessment of single cases with its attempt to describe the specific environmental events which evoke and consequate behavior. Third, we need to return more closely to the definition of a test. Quoting Cronbach (1970) a test is defined as "a systematic procedure for observing a person's behavior and describing it with the aid of a numerical scale or a category system (p. 26)."

While these suggestions have been, in the main, neglected by a great body of practitioners and test developers, they are not particularly new nor original. In fact, what these recommendations boil down to is a rather reactionary request that we turn the clock back about seven decades and return to the year 1905. It may be recalled that it was in that year that Binet and Simon first discoursed "Upon the necessity of establishing a scientific diagnosis of inferior states of intelligence."

Unless one periodically rereads Binet it can be forgotten how prescient he was in anticipating many of the psychometric issues being dealt with in the contemporary literature. For example, he warned against the subtle effects the examiner may have on test behavior, questioned the value of self-report data, and derided "authors who have made a specialty

of organizing new tests according to theoretical views but who have made no effort to patiently try them out in the schools (Binet and Simon, 1905, p. 884)."

More to our point, Binet distinguished among three models possible in the assessment of intelligence. They included:

> The medical method which aims to appreciate the anatomical, physiological, and pathological signs of inferior intelligence.
>
> The pedagogical method which aims to judge of the intelligence according to the sum of acquired knowledge.
>
> The psychological method, which makes direct observations and measurements of the degree of intelligence (Binet and Simon, 1905, p. 883).

Of the three, Binet chose the psychological method precisely because it was the most direct method of measuring "the state of the intelligence as it is at the present moment (Binet and Simon, 1905, p. 883)."

Binet and Simon saw their task, then, as the development of an instrument which would most directly measure, through observation, the behavior of particular concern. Such an intent hardly differs from Cronbach's definition of a test. Tests are thus rooted in a tradition which relies on the direct observation of samples of behavior and it is to this tradition that it is suggested we return. Those psychometricians and practitioners who have moved away from this tradition are the psychological alchemists who have turned a silk purse into a sow's ear.

If we are genuinely interested in gathering data about behavior that is contextualized, i.e., rooted to environmental events, relevant to the current functioning of an individual, and appropriate for assessing behavior change, than we need to discard the tests inappropriate to these

purposes and employ techniques which can accomplish these tasks. I have suggested the term psychosituational assessment[4] (Bersoff, 1971c, 1973) to encompass those techniques. In the past two years some colleagues and I have described or devised interview procedures (Bersoff & Grieger, 1971), psychotherapeutic intervention strategies (Grieger & Bersoff, 1973), instructional methods (Tillman, Bersoff & Dolly, in preparation), and teacher-child observation schedules (Bersoff, 1971b; Bersoff & Ericson, 1971) specifically designed to conform to the criterion of relevance to behavior change.[5]

The primary focus in psychosituational assessment is the analysis of behavior and the delineation of the immediate antecedent and consequent conditions which evoke, reinforce, and perpetuate that behavior. The major concern is that the individual is assessed as he interacts and is affected by the environment. Seen in this context, those procedures which constitute a psychosituational assessment may be considered as at once more valid and humane than many other kinds of assessment and intervention programs.

The procedures are more valid because in most cases they rely on the direct and continuous measurement of performance. Direct measurement requires that behavior be observed and counted as it is manifested in the person's behavioral repertoire rather than deducing a picture of behavior from test results. Direct measurement takes place at the site where behavior naturally occurs. Continuous measurement allows for ongoing monitoring of behavior across a period of time, occurring daily or for extended periods during the day. Most academic measurement, as presently practiced, takes place at infrequent intervals. Such measurement does not allow the teacher to monitor her instructional procedures so that they

can be redesigned for those children with whom they are not effective. Children may continue to fail or not comprehend material in between widely spaced assessments. Assessment that more closely approximates continuous measurement insures that the teacher can alter the material presented or restructure her instructional strategy more quickly. In this way there is less elapsed time between student failure, recognition of that failure, and subsequent intervention to prevent continued failure.

The procedures are more humane because their intent is the assessment of the individual as he interacts with the environment. The goal is not to automatically manage or fix the person, but to restructure those situations which evoke disturbance or failure. In psychological testing, where for instance, a child is evaluated individually, removed from the functioning environment, or tested in groups, the focus for blame and change is the child. It is almost impossible for it to be any other way; no other person's functioning is assessed. This phenomenon is most clear in the assessment of "personality" by means of projective tests. By assuming that people will reveal their inner, underlying, unconscious personality through ambiguous unstructured stimuli, regardless of their objective characteristics, psychologists have consistently made errors in coming to decisions about the actual functioning of the people they have assessed. Consider the administration of the Thematic Apperception Test in this regard. If one goes about the business of obtaining objective ratings as to theme and tone of the cards one finds that they are most often seen as portraying unpleasantness, depression, loneliness, or conflict. One reason is that the cards are printed in gray, somber ink. Another is that the pictures themselves evoke such themes. One card portrays a mother turning her back on her son, another has an emaciated

man dressed in Victorian style clothes wringing his hands in a graveyard, a third shows a person with his/her head on a couch with a gun on a floor nearby, while another depicts a poor boy sitting alone in the doorway of a log cabin. These cards are more typical of the overall affect pull of the stimuli than the exception (c.f., Epstein, 1966).

What happens to a child who is referred for some behavioral difficulties and is administered the TAT? Well, like the women accused of being witches in olden times, the child is not likely to come out of the assessment experience too healthily. During the sixteenth century an alleged witch's guilt was tested in an ingenious way. The examination involved dumping her in deep water restrained by ropes. If she sank, she was innocent; if she floated she was guilty--in either case she was dead. And, like the witch, the child referred for projective testing often ends up in the same position.[6] Given cards which normally evoke themes of death or sadness or aggression, he is normally going to tell stories which involve these themes. Ignoring the fact that his behavior is the result of the interaction of what he brings to the test and the nature of the test itself, the child is likely to be labeled in "essence" terminology--depressed, aggressive, etc., and it will be assumed that this is the way he is in most, if not all, situations.

In educational settings, testing removes both the child and the psychologist from the natural situation of the classroom, restricting the psychologist's contact with the child's teacher, a primary agent of behavior change. Thus, the teacher is isolated from the information-gathering process and the psychologist is prevented from gathering data concerning those events which may be evoking academic failure (antecedent conditions) and those which are consequating it (reinforcing events).

However, in a psychosituational analysis of behavior, the whole situation is under investigation. The aim is to "contextualize" behavior and discover what Fischer (in press; 1971) calls the When/When-nots of specific behavior. In the classroom, both the child's behavior and those of the teacher are carefully assessed before any changes in the instructional environment are made. The child and his teacher "co-constitute" this instructional environment, interacting to evoke behavior in the other. Any assessment procedure which isolates the target person from the significant others who participate in his behavior can be considered, at the very least, incomplete, at the most, unethical.

There are many techniques that can be used to gather the data necessary for a psychosituational analysis of behavior. Some, more than others, approach the criteria of directness and continuity. For example, Santostefano (1968), following in the tradition of MacKinnon (1958), has suggested the construction of miniature situations. Miniature situations replicate functioning environments and enable the assessor to study behavior under controlled but natural-like conditions when observation is not possible or desirable in the real situation.

Of course, the most direct measures are the ones that occur in natural settings. Of these, time and event sampling are the most popular and useful. In time sampling the observer preselects behavior to be observed and then records its occurrence within uniform and usually short time intervals. Generally, the categories of behavior to be observed are coded in advance for quick and precise judgments in the observation setting and for later scoring and interpretation. Event sampling, like time sampling, begins with a plan to study specific behavioral events. However, the events are not counted within prescribed

time intervals but are simply recorded every time they occur. Frequencies, rather than percent of time intervals, are the basic data.

In addition to assessing behavior in natural, situation-specific settings, time and event sampling have other attractions, the prime one being reliability. By selecting the behaviors to be observed, by delimiting the temporal length of observation, by recording large numbers of behavior, and by precoding, a great deal of precision and control is gained with a commensurate reduction in the number of equivocal judgments to be made.

Many of the variables that were once thought amenable only to psychometric measurement via standardized tests are now being assessed through direct measurement. A variety of observational procedures is available for the assessment of personality (Santostefano, 1962). Recently, there has been an attempt reported to evaluate intellectual functioning of elementary school children through direct observation (Lambert, Cox & Hartsough, 1970). Within a Piagetan orientation, six classes of Intellectual Functioning Events (IFE) were defined, observed, and recorded.[7] The IFE were classified as basic language skills, concepts of time and space, logical concepts, math concepts, reasoning skills, and general signs of development. The investigators found that with appropriate behavioral definitions all six areas of intellectual functioning were observable in first grade classrooms. The experiment was not totally successful, however, and the report suffers from the absence of reliability data.

Nevertheless, it is not so much the specific piece of research that is to be applauded but, rather, its intent. Intelligent behavior measured in the classroom is much more relevant than an IQ score in assessing current functioning, developing instructional strategies, and monitoring

behavior change.

There are many limitations, of course, to observational procedures. Encoding format which use percents as measures, as do most time sampling schedules, yield misleading data (Bersoff & Ericson, 1972). Time and event sampling do not honor natural behavior units and unless provisions are made for its recording, the duration of behavior is often neglected. There are also methodological drawbacks, the most important of which are observer influence and bias. At times, reliability may suffer when the behaviors to be observed are poorly defined or poorly coded or when there are too many behaviors to be observed at one time (c.f., Medley & Mitzel, 1963; Weick, 1968; Wright, 1960).

Observation has also been criticized for being time-consuming and for producing ungeneralizable data. These, I believe, are spurious criticisms. It is because psychological tests have been used to generalize to many situations that they have not been able to predict accurately to one. There is no doubt that observational techniques are narrow band measures and in this respect they suffer in comparison to such wide band instruments as intelligence tests. But, if the goal in assessment is validity in the terms that have been defined in this paper, then such alleged assets as efficiency, temporal economy, factorial purity, and broad (but inaccurate) generalizability must become secondary considerations. The assumption of trans-situational consistency of behavior was challenged almost twenty years ago by Spiker & McCandless (1954) but little overt movement, by and large, has been made to modify assessment procedures or our perceptions of people as motivated by what Wallace (1966) calls a "bewildering phantasmagoria of putative underlying mediating structures and mechanisms (p. 132)."

Within the framework for assessment that has been described direct, continuous, naturalistic observation of behavior provides many of the advantages one seeks in evaluation. There is a great deal of relevant information yielded per assessment unit. The information is precise and because it is direct there is little question that it represents a true sample of behavior as there may be when standardized testing is used. Borrowing Gagne's (1970) term, observation approaches the criterion of distortion-freeness. Assessors receive accurate, daily, almost immediate feedback as to the effectiveness of an intervention strategy, be it instructional or therapeutic. And, most important of all, the process generates authentic data concerning the current functioning of the individuals being assessed.

References

Bersoff, D. N. School psychology as "institutional psychiatry." Professional Psychology, 1971, 2, 266-270. (a).

Bersoff, D. N. Positive reinforcement observation schedule (PROS). Paper presented at the joint meeting of the Virginia Psychological Association and the Virginia Association of School Psychologists, Charlottesville, October, 1971. (b)

Bersoff, D. N. "Current functioning" myth: An overlooked fallacy in psychological assessment. Journal of Consulting and Clinical Psychology, 1971, 37, 391-393. (c)

Bersoff, D. N. Behavioral approaches to assessment and observation in the school. In J. F. Magary (Ed.), Handbook of school psychological services. St. Louis: C. V. Mosby, 1973.

Bersoff, D. N. & Ericson, C. R. A precise and valid measure of behavior and behavior change. Proceedings of the 79th Annual Convention of the American Psychological Association, 1971, 6, 555-556.

Bersoff, D. N. & Ericson, C. R. Precise and valid measure of behavior and behavior change in the classroom. Journal of School Psychology, 1972, 10, in press.

Bersoff, D. N. & Grieger, R. M. An interview model for the psychosituational assessment of children's behavior. American Journal of Orthopsychiatry, 1971, 41, 483-493.

Bijou, S. W. & Peterson, R. F. The psychological assessment of children: A functional analysis. In P. McReynolds (Ed.), Advances in psychological assessment. Palo Alto: Science and Behavior Books, 1971.

Binet, A. & Simon, T. Upon the necessity of establishing a scientific diagnosis of inferior states of intelligence. In T. Shipley (Ed.), Classics in psychology. New York: Philosophical Library, 1961.

Cronbach, L. J. Essentials of psychological testing. New York: Harper & Row, 1970.

Educational Testing Service. Let's look at first graders. Princeton, N.J.: Educational Testing Service, 1965.

Epstein, S. Some theoretical considerations on the nature of ambiguity and the use of stimulus dimensions in projective techniques. Journal of Consulting Psychology, 1966, 30, 183-192.

Fischer, C. T. Toward the structure of privacy: Implications for psychological assessment. In A. Giorgi, W. F. Fischer, R. vonEckertsburg (Eds.), Duquesne studies in phenomenological psychology. Pittsburgh: Duquesne University Press, 1971.

Fischer, C. T. Contextual approach to assessment. Community Mental Health Journal, in press.

Fiske, D. W., & Pearson, P. H. Theory and techniques of personality measurement. Annual Review of Psychology, 1970, 21, 49-86.

Fulkerson, S. C. Some implications of the new cognitive theory for projective tests. Journal of Consulting Psychology, 1965, 29, 191-197.

Gagné, R. M. Instructional variables and learning outcomes. In M. C. Wittrock & D. E. Wiley (Eds.), The evaluation of instruction. New York: Holt, Rinehart & Winston, 1970.

Grieger, R. M. & Bersoff, D. N. Counseling and psychotherapy in the school setting. In J. F. Magary (Ed.), Handbook of school psychological services. St. Louis: C. V. Mosby, 1973.

Holland, J. L. & Richards, J. M. Academic and nonacademic accomplishment: Correlated or uncorrelated? Journal of Educational Psychology, 1965, 56, 165-174.

Lambert, N. M., Cox, H. W., & Hartsough, C. S. The observability of intellectual functioning of first graders. Psychology in the Schools, 1970, 7, 74-85.

MacKinnon, D. W. An assessment study of Air Force officers. Part V: Summary and applications. WADC Technical Report 58-91 (V), Wright Air Development Center, 1958.

Medley, D. M. & Mitzel, H. E. Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Mischel, W. Personality and assessment. New York: Wiley, 1968.

Murstein, B. I. Assumptions, adaptation level, and projective techniques. In B. I. Murstein (Ed.), Handbook of projective techniques. New York: Basic Books, 1965.

New York Times. Dumbbell test. New York Times, 1972, Editorial page.

Peterson, D. R. The clinical study of social behavior. New York: Appleton-Century-Crofts, 1968.

Rigsby, M. B. Rates of reinforcement in testing and classroom settings. Unpublished masters thesis, Clemson University, 1972.

Santostefano, S. Performance testing of personality. Merrill-Palmer Quarterly, 1962, 8, 83-97.

Santostefano, S. Miniature situations and methodological problems in a parent-child interaction research. Merrill-Palmer Quarterly, 1968, 14, 285-312.

Spiker, C. C. & McCandless, B. R. The concept of intelligence and the philosophy of science. Psychological Review, 1954, 61, 255-266.

Stake, R. Comments on Professor Glaser's paper. In M. C. Wittrock & D. E. Wiley (Eds.), The evaluation of instruction. New York: Holt, Rinehart & Winston, 1970.

Szasz, T. The manufacture of madness. New York: Harper & Row, 1970.

Terman, L. & Merrill, M. Stanford-Binet Intelligence Scale: Manual for the third revision, Form L-M. Boston: Houghton Mifflin, 1960.

Tillman, M., Bersoff, D. N. & Dolly, J. A behavioral approach to classroom instruction and management. Lexington, Mass.: D. C. Heath, in preparation.

Wallace, J. An abilities conception of personality: Some implications for personality measurement. American Psychologist, 1966, 21, 132-138.

Weick, K. E. Systematic observational methods. In G. Lindsey & E. Aronson (Eds.), Handbook of social psychology. Vol. 2, Reading, Mass.: Addison-Wesley, 1968.

Wesman, A. G. Intelligent testing. American Psychologist, 1968, 23, 267-274.

Wittrock, M. C. The evaluation of instruction: Cause-and-effect relations in naturalistic data. In M. C. Wittrock & D. E. Wiley (Eds.), The evaluation of instruction. New York: Holt, Rinehart & Winston, 1970.

Wright, H. F. Observational child study. In P. H. Mussen (Ed.), Handbook of research methods in child development. New York: Wiley, 1960.

Footnotes

[1]Paper presented at the Southeastern Invitational Conference on Measurement in Education, Athens, Georgia, December 1972.

[2]For readers fortunate enough to have escaped involvement, the former names are all successful plaintiffs in court cases involving testing, the consequences of tester's evaluation, and the suing of testers.

[3]While these views exemplify a major segment of thought in instructional evaluation, intellectual honesty demands that we note they by no means represent a universal view (c.f., Stake, 1970).

[4]Recently, a synonomous term has been suggested--contextual assessment--which, while defined from a phenomenological point of view, expresses many of the ideas described above (Fischer, in press).

[5]Also, Bersoff & Moyer, Development and applications of the Positive Reinforcement Observation Schedule (PROS); and Ellett & Bersoff, New tricks for old dogs: An integrated approach to psychological assessment, both manuscripts recently submitted for publication.

[6]I am indebted to Thomas Szasz (1970) who first suggested the analogy between witch hunting and psychological testing (c.f., Bersoff, 1971a).

[7]The material was developed by (of all people) Educational Testing Service (1965) and can be found in a booklet entitled, Let's look at first graders.