DOCUMENT RESUME

ED 342 796 TM 017 958

AUTHOR Clark, Burton A.
TITLE National Fire Academy Test Technical Manual for the Chemistry of Hazardous Materials Course. Final Exam.
INSTITUTION National Fire Academy, Emmitsburg, MD.
PUB DATE Oct 90
NOTE 30p.; The test is not included in this document.
PUB TYPE Guides - Non-Classroom Use (055) -- Reports - Evaluative/Feasibility (142)

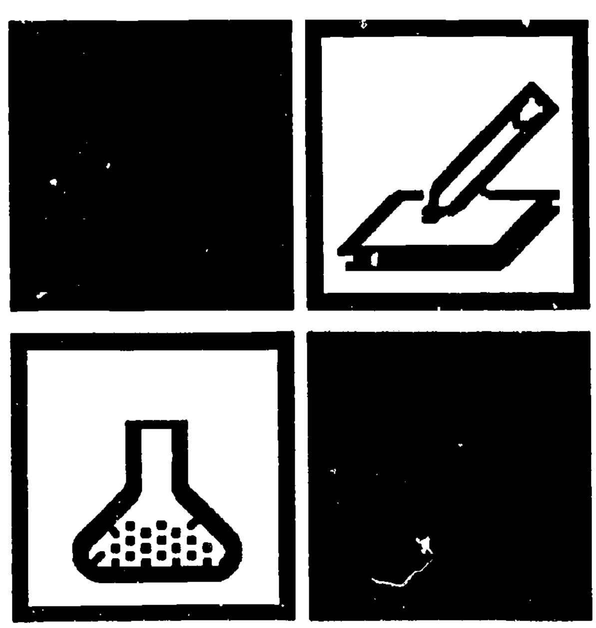
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Chemistry; *Criterion Referenced Tests; Cutting Scores; Fire Fighters; *Fire Science Education; Hazardous Materials; Item Analysis; *Job Training; Postsecondary Education; Pretests Posttests; *Scoring; Test Construction; Test Manuals; Test Reliability; *Test Use; *Test Validity
IDENTIFIERS *Final Examinations

ABSTRACT

This first test and technical manual explains the development, purpose, use, validity, and reliability of the criterion-referenced final examination for the Chemistry of Hazardous Materials course of the National Fire Academy in Emmitsburg (Maryland). Because of possible diversity in the testing knowledge and background of readers, the manual includes an overview of testing, explaining technical issues associated with validity, reliability, and optimal cut score. The examination consists of 100 questions measuring 48 objectives of the course. Six test reviewers determined content validity, and the development team and the six reviewers considered the domain validity of the instrument. Decision validity was determined by pretesting and posttesting 127 students in 5 classes. Test reliability was measured for 221 students using the threshold loss function reliability index, the internal consistency reliability, and the standard error of measurement (based on 44 posttest scores). Item analysis procedures used were: (1) sensitivity to instructional effects (127 examination scores); (2) Hambleton's item review to check bias (2 reviewers); and (3) informed feedback from 70 students. Faculty and students can be confident of the examination's validity and reliability. One graph and three tables present study data. A list of 22 references, 4 appendices describing the course, and instructions to students and reviewers are included. (SLD)

National Fire Academy Test Technical Manual for the Chemistry of Hazardous Materials Course

FINAL EXAM



By: Burton A. Clark, Ed.D.

FOREWORD

The test technical manual is designed to explain the development, purpose, use, validity, and reliability of the NFA's Chemist[illegible]azardous Materials course final exam. The manual is written to be used by NFA faculty, management, and students. In addition, other individuals interested in educational testing may find the manual useful.

It is assumed that readers of this manual may have a wide range of formal training in testing from individuals with advanced degrees in tests and measurements to individuals with no formal psychometric training. Because of this possible diversity in readership, the manual includes a basic overview of testing to allow those readers with little or no measurement training to understand the more advanced technical issues associated with the instrument's validity, reliability, and optimal cut score. The manual also presents sufficient information for the measurement specialists to evaluate the development and research procedures used in the validity and reliability studies.

This is the first test and technical manual developed by the National Fire Academy according to the American Psychological Association's Standards for Educational and Psychological Testing 1985. As such, the research procedures and published documents will serve as models for future projects.

National Fire Academy,
Emmitsburg, Maryland,
October, 1990

ACKNOWLEDGMENTS

The development of this test technical manual and the research associated with it was truly a team effort. I am grateful to the leadership of the NFA for giving me this opportunity. The quality of this work is a result of the professionalism of the development team and the individual reviewers. Without the assistance of the SCSC staff, who were responsible for word processing and editing, and the Key Data staff, who were responsible for data entry and statistical analysis, and the NETC library staff, who were responsible for literature searching, this project would not have been possible.

Finally, to all the individuals who gave me advice, counsel, and support, thank you very much.

Burton A. Clark

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

OVERVIEW OF TESTING

Introduction

E.L. Thorndike published the first book on educational testing in 1904 (Chauncey and Dobbin, 1966:11). "Educational testing has played an important role in American education for more than 70 years" (Resnick, 1982:173). The testing of large numbers of military recruits during WWI and WWII helped to make testing an acceptable part of our culture. Although testing has existed as long as learning itself, it wasn't until 1954 that the American Psychological Association (APA) published the first Standards for Educational and Psychological Testing. Subsequent versions of the APA standards were published in 1966, 1974, and 1985. The standards follow:

> Educational and psychological testing represents one of the most important contributions of behavioral science to our society. . . . the proper use of well-constructed and validated tests provides a better basis for making some important decisions about individuals and programs than would otherwise be available (American Psychological Association, 1985:1).

Tests are tools that teachers can use to help reduce the subjectivity, biases, and opinionated aspects of the educational decisionmaking process (Kubiszyn and Borich, 1990). The quality of the decisions which are made based on tests are only as good as the quality of the test in terms of the instrument's validity and reliability. But, it is important to remember ". . . that all test scores are at best estimates that are subject to greater or lesser margins of error" (Kubiszyn and Borich, 1990:2).

There are two basic classifications of test instruments: norm-referenced and criterion-referenced. The American Psychological Association gives the following definitions:

> **norm-referenced test** An instrument for which interpretation is based on the comparison of a test taker's performance to the performance of other people in a specified group. (APA, 1985:92)
>
> **criterion-referenced test** A test that allows its users to make score interpretations in relation to a functional performance level, as distinguished from those interpretations that are made in relation to the performance of others. (APA 1985:90)

The NFA's Chemistry of Hazardous Materials course final exam test instrument is a criterion-referenced test.

Validity and Reliability

Validity and reliability questions must be asked about any test instrument. A test is not valid and reliable; tests are valid and reliable for a particular population and purpose (Gay, 1981 and Garrett, 1971). Reliability is not measured, it can only be inferred; tests are judged to be adequate, marginal, or satisfactory (Guion, 1974). Garrett (1971:360) states "Reliability is concerned with the stability of test scores—does not go beyond the test itself. Validity, on the other hand, implies evaluation in terms of outside independent criteria." Test developers and test users must provide evidence of the validity and reliability of any instrument used (Guion, 1974).

Criterion-referenced tests (CRTs) are primarily studied for content validity. Content validity is determined by having the test reviewed by experts in the field. The objective of this review process is to have the experts reach consensus that the test has content validity. A group process technique or Delphi technique may be used to reach consensus (Isaac and Michael, 1981). The experts use judgment to determine if the test questions are based on the content taught in the class; this is referred to as item validity, and if the number of questions are in correct proportion to the sub-content areas of the class, this is referred to as sampling validity (Gay, 1981). According to Guion (1974:24), content validity should also include a comparison among performance domains of the instruction and the domains being measured by the test instrument.

Another important characteristic of CRTs is the decision validity of the instrument. In other words, does the test accurately identify masters, those students who pass the course, and non-masters, those students who fail the course. The procedure for measuring decision validity ". . . involves (1) setting a standard test performance, and (2) comparing the test performance of two or more criterion groups [masters vs non-masters] in relation to the specified standard" (Hambleton, 1984:220).

Classical test reliability theory, principally used for norm-referenced tests, relies on the variability of test scores. In other words, a normal distribution of scores is expected. CRT results are not expected to have great variability. For example, in competency-based instruction all or the majority of students are expected to achieve the mastery level of performance.

CRTs which are typically used to classify students into two groups, pass-fail or masters-non-masters, are subject to two types of wrong decisions or threats to the test's validity. "A test taker who actually belongs in the lower group can get a score above the passing score; a student who actually belongs in the higher group can get a score below the passing score" (Livingston and Zieky, 1982:12). These errors are referred to as false positive and false negative; Berk (1976:5) classifies these two errors as Type II and Type I respectively.

There are two principal concepts associated with the false positive and false negative threats to the reliability of CRT decisions. The first concept is referred to as "threshold loss function" which is based on the philosophy that students are classified as pass or fail based on a cut score or threshold and that false positive and false negative are equally important regardless of the error size. The second concept "squared-error loss function" is based on the philosophy that the degree of the misclassification error is important, the larger the error the less reliable the measure (Belcher, 1987 and Berk, 1980). A number of statistical procedures have been developed to calculate reliability indices for CRTs. After reviewing thirteen different statistical procedures Berk (1980) indicates that the Index of Agreement (P_o) and Kappa (K) index can be used to determine threshold loss function and K2(X,Tx) and the Φ (λ) can be used to determine the squared-error loss agreement index. In addition, Berk (1980) and Belcher (1987) both indicate that the classical internal reliability Kuder-Richardson 20 or 21 formulas have been used with CRTs.

Another reliability indic which can be applied to CRTs is the standard error of measurement (SEM). Five different statistical procedures have been developed to determine the SEM of CRTs (Berk, 1984b and Livingston

1982). Livingston's (1982:135) procedure is the only one which indicates that ". . . the important SEM is not that of the full group of test takers, but the SEM at the passing score. . ." This procedure is important for tests used to make pass/fail decisions, especially for students at or near the cut score (Livingston 1982 and Belcher, 1984).

Finally, as with norm-reference tests, the types of validity and reliability procedures applied to CRTs depends on the type of test and the purpose of the test (Gay, 1981; Berk, 1980 and Hambleton, 1984).

Item Analysis

"Besides assessing relevance and reliability, test validation studies often examine the quality of each of the test items" (Cangelosi, 1990:36). Item analysis procedures involve both quantitative and qualitative methodologies (Kubiszyn and Borich, 1990; Berk, 1984a). The item analysis procedures used for criterion-referenced tests are different from the procedures used for norm-referenced tests. The critical question to be addressed by the item analysis of a criterion-referenced test is "To what extent did the test item measure the effects of the instruction" (Gronlund, 1976:272)?

The appropriate statistical procedure is to determine the sensitivity to instructional effects for each test item. Gronlund (1976) and Berk (1984a) both agree that the pre-instruction/post-instruction method is the most appropriate procedure where feasible. To be considered effective, an item's score can be between .00 and 1.00. The higher the number the more effective, but it remains a matter of judgment as to what level will be acceptable. At a minimum the item score should be a positive number (Gronlund, 1976:274).

There are two judgmental procedures to be applied. The first examines each item for content bias based on race or gender. Hambleton's item review form to detect bias is recommended by Berk (1984a:101). The second procedure is an informed student feedback debriefing session. After students have taken the test they are asked to comment on any confusing questions, terms they did not understand, or problems they encountered (Berk, 1985; Kubiszyn and Borich, 1990).

Performance Standards

The cut score or performance standard ". . . is the Achilles heel of criterion-referenced testing" (Shepard, 1984:169). The cut score is used to determine the decision validity, threshold loss function reliability, and squared-error loss function reliability of criterion-referenced tests (Berk, 1980). This means that the validity and reliability of the test is only as good as the ability of the test to discriminate between masters and non-masters consistently. "It should be remembered that no matter how judicious the procedures are for arriving at a standard, the cut-off point still imposes a false dichotomy on a continuum of proficiency" (Shepard 1984:170). Despite these caveats, cut scores are used and the rationale for setting performance standards must be articulated by test developers (American Psychological Association, 1985).

There are two basic methodologies used to determine cut scores. The decision can be based on judgments about the test questions and/or judgments about the test takers either individually or as a group. Livingston and Zieky (1982:53) state

> There is no one method that is best for all testing situations. Your choice of a method should depend

on what kind of judgments you can get — and believe. We believe that the best kind of data to use — if you can get them — are test scores of real test takers whose performance has been meaningfully judged by qualified judges.

For judging test takers the contrasting group's method has the ". . . strongest theoretical rationale" (Livingston and Zieky, 1982:53). For criterion-referenced tests Shepard (1984:183) recommends Berk's procedure which is ". . . identical to the contrasting group's method except that criterion groups would be instructed and uninstructed groups rather than judged masters and non-masters." Berk's (1976) criterion-groups validation model utilizes the test results of instructed and uninstructed students to determine the optimal cut score. This procedure identifies the test score that results in the lowest possible false positive and false negative misclassification errors.

Objectives

"If a criterion-referenced test does not unambiguously describe just what it's measuring, it offers no advantage over norm-referenced measures" (Popham, 1984:29). Criterion-referenced tests measure student achievement of objectives. All the authors reviewed (Popham, 1984; Gronlund, 1976; Bloom, 1956; Cangelosi, 1990) agree that the criterion being tested and referenced to are the objectives of the instruction.

Objectives are based on the content of the instruction and the cognitive processes used to perform the behavior. Bloom's (1957) taxonomy of educational objectives serves as the foundation for describing the cognitive process; all of the authors reviewed (Kubiszyn and Borich, 1990; Gronlund, 1976; Cangelosi, 1990) utilize Bloom's classification system.

The taxonomy of educational objectives for the cognitive domain consists of six levels which are knowledge, comprehension, application, analysis, synthesis, and evaluation. Each domain has subsets which describe in detail the specific cognitive process the terms are meant to represent (Bloom, 1957).

Summary

The literature on educational testing is prolific; this overview represents only the information needed to have the background to understand the following sections of this technical manual.

Tests are to education like telescopes, microscopes, and stethoscopes are to astronomy, biology, and medicine. Tests are the instruments that educators use to measure phenomena. The ability of man to understand the physical, behavioral, and psychological world is directly correlated to the validity and reliability of the instruments they use.

The theories, methods, and procedures reviewed are all designed to help test developers, test users, and test takers to communicate with each other more efficiently and effectively.

DESCRIPTION: TEST AND PROCEDURES

Purpose

The purpose of the test is to determine which students pass or fail the National Fire Academy's Chemistry of Hazardous Materials course and receive a course certificate. A copy of the course description and student selection criteria are in Appendix A.

Description

The instrument is a criterion-referenced test; it consists of 100 questions (49 multiple choice, 27 matching, and 24 true/false). The test questions are written to measure the 48 objectives of the course.

The objectives for the course are contained in Appendix B. A copy of the instrument cannot be included in this manual because of the NFA's need to maintain strict security over the instrument. The instrument is available for review by appointment with the NFA Assistant Superintendent for leadership and hazardous materials.

Administration

The test is used as the final exam for the NFA's Chemistry of Hazardous Materials course for both the on-campus delivery and field delivery of the two-week course. Only NFA faculty and adjunct faculty are authorized to administer the test. A copy of the instructions to students (Appendix C) will be given to each student.

The test is administered on the ninth day of the course during the afternoon period. There is no time limit on students taking the test. The only aids to the student during the exam are the chemistry periodic table and scratch paper; no other reference materials are allowed.

Tests, answer sheets, and scratch paper are distributed to the students. Students are instructed not to write on the test, all answers are to be recorded on the answer sheet. Upon completion, students shall return all tests, answer sheets, and scratch paper to the instructor.

An NFA faculty member must be present during the testing period to proctor the exam. Cheating results in automatic failure of the course.

Scoring and Grading

The NFA faculty member will score the answer sheets using the answer key. Students receive one point for each correct answer. The student's grade is recorded on the answer sheet. A score of 70 or greater is considered passing.

Recordkeeping

NFA faculty will inform the students individually if they passed or failed. The student's score will be recorded on the grade report form. The student's official transcript only reflects a pass or fail notation.

Certificates

Certificates of completion will be awarded and presented to resident students during the graduation ceremony. Students taking the course in a field delivery will have the certificate mailed to them by the registrar's office.

Security

Strict security must be maintained on the test instruments, answer keys, and answer sheets. It is the responsibility of the NFA faculty to maintain control of the test materials. Test materials are not to be photocopied or hand copied by anyone. All test materials, in sufficient quantity, will be supplied by NFA staff. All materials are to be returned to NFA staff directly or by mail.

Distribution of Technical Manual

A copy of this test technical manual will be distributed to all NFA Chemistry of Hazardous Materials faculty. In addition, a copy of the manual is on file, for public review, at the National Emergency Training Center, Learning Resource Center, Emmitsburg, MD. The manual may be borrowed through interlibrary loan.

TEST VALIDITY

The test instrument was studied for three types of validity content, domain, and decision. Both logical and empirical research procedures were used to conduct the studies.

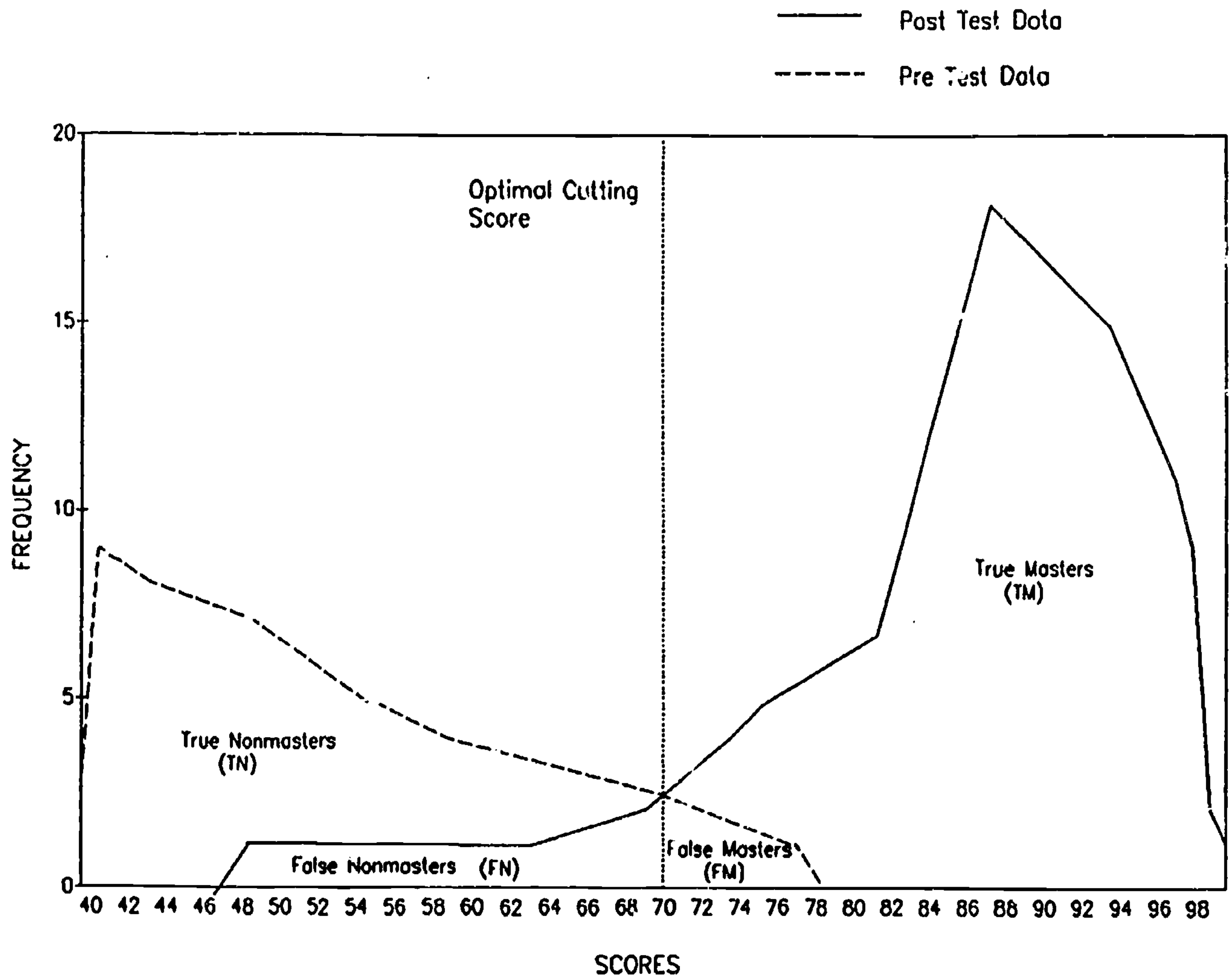
Content validity was first determined by the NFA development team which consisted of the NFA program chair for hazardous materials, the NFA hazardous materials training instructor, and the NFA program chair of management science; their biographies are contained in Appendix D.

The development team came to group consensus that all the test questions were based on the course content. These data were then reviewed by six independent reviewers. The biographies of the reviewers are in Appendix D. The six reviewers agreed that the test questions were based on the course content. The same process was used to determine that the test questions were an appropriate sample of the course content. The course content materials include the NFA Chemistry of Hazardous Materials course Instructor Guide 1985, the NFA Chemistry of Hazardous Materials course Student Manual 1985, and the text Fire Chemistry I: The Basics of H.T.M. Third Edition by Ron Edwards, 1981.

The cognitive domain classification of each course objective and test question was identified based on Bloom's (1956) Taxonomy for the Cognitive Domain. The percentage of test questions in each domain is as follows: knowledge 42 percent, comprehension 22 percent, application 18 percent, analysis 8 percent, and synthesis 10 percent. Based on the review by the development team and independent reviewers, there is an appropriate match between the objective domain and question domain.

A table of specifications was developed. The table identifies the course objectives, the domain classification of the objectives and the corresponding test question, the location of the answer, and the domain classification of the question. A copy of this table cannot be included in this manual because the NFA must maintain strict security over the test. The table of specifications is available for review by appointment with the NFA Assistant Superintendent for leadership and hazardous materials.

Decision validity was determined by pretesting and posttesting five Chemistry of Hazardous Materials classes which totaled 127 students. The students before taking the course were judged by the NFA development team to be non-masters. The number of students correctly judged to be non-masters was N=125 or 98 percent of the students failed the pretest. The same students were judged to be masters at the conclusion of the course. The number of students correctly judged to be masters on the posttest was N=121 or 95 percent of the students passed the posttest. Therefore the decision validity of the test instrument is .96 [(.95 + .98) + 2 = .96] based on a cut score (passing score) of 70 (Figure 1).



Note: TN=98%, FM=2%, FN=5%, TM=95%

FIGURE 1

DISTRIBUTION OF STUDENTS CLASSIFIED AS MASTERS AND NON-MASTERS BASED ON PRE- AND POSTTEST SCORES N=127

Based on a cut score of 70 the probability of a false negative decision error is .024 and the probability of a false positive decision error is .008. The validity coefficient is .937. (Table 1)

TABLE 1

PROBABILITY OF CORRECT DECISIONS AND MISCLASSIFICATION ERRORS, VALIDITY COEFFICIENT, AND CLASSIFICATION PROBABILITIES FOR DIFFERENT CUT SCORES

Cut Score	Probability of Correct Decision	Misclassification False Negative	Misclassification False Positive	Validity Coefficient	Classification Probabilities Sensitivity (Masters)	Classification Probabilities Specificity (Non-masters)
74	.953	.043	.004	.908	.913	.992
73	.957	.039	.004	.916	.921	.992
72	.969	.028	.004	.938	.944	.992
71	.969	.028	.004	.938	.944	.992
70*	.969	.024	.008	.937	.952	.984
69	.965	.024	.012	.929	.952	.976
68	.969	.020	.012	.937	.960	.976
67	.969	.020	.012	.937	.960	.976
66	.961	.020	.020	.921	.960	.960
65	.957	.016	.028	.914	.968	.944
64	.953	.012	.035	.907	.976	.929
63	.949	.012	.039	.899	.976	.921

*Optimal cut score for the NFA Chemistry of Hazardous Materials course final exam test instrument.
Note: These data are based on 127 students from five classes who were pretested and posttested with the same instrument.

TEST RELIABILITY

Three types of test reliability were measured using the threshold loss function reliability index (P_0), the internal consistency reliability (Kuder-Richardson 21), and the standard error of measurement.

The P_0 index was calculated for cut scores that ranged from 74 to 63, based on the pretest (non-master) results and posttest (master) results of nine classes or 221 students. The highest P_0 reliability index for non-masters .968 and masters .974 is reached at the cut score of 70. (Table 2).

TABLE 2

THRESHOLD LOSS FUNCTION RELIABILITY INDEX (P_0) ON NON-MASTERS N=127 AND MASTERS N=221 AT VARIOUS CUT SCORES

Cut Score	Non-masters P_0	Masters P_0
74	.968	.936
73	.968	.945
72	.968	.954
71	.968	.961
70	.968	.974
69	.968	.974
68	.968	.974
67	.961	.974
66	.961	.974
65	.925	.974
64	.945	.974
63	.949	.974

The internal consistency reliability based on KR-21 is .806 for the pretest data and .839 for the posttest data.

The standard error of measurement based on the posttest data and a mean of 85.1 is SEM .57. The standard error of measurement based on Livingston's procedure is SEM L=2.71 calculated on 44 posttest scores that ranged from a score of 50 to 80. This means that a student who scored 70 if repeatedly tested would score +/- 3 points 68 percent of the time and +/- 6 points 95 percent of the time.

ITEM ANALYSIS

Three item analysis procedures were used: sensitively to instructional efforts, Hambleton's item review to detect bias, and informed student feedback.

The sensitivity to instructional efforts was determined for each test question based on pretesting and posttesting five classes N=127. All the test questions had a positive score. The largest number of questions, 34 percent, scored between .40 to .59, the smallest number of test questions, 3 percent, scored .80. (Table 3) The test items were judged to be effective by the development team.

TABLE 3

PERCENTAGE OF TEST QUESTIONS AT SENSITIVITY TO INSTRUCTIONAL EFFECTS RANGES

Percentage of Test Questions	Sensitivity to Instructional Effects Range
20	.01-.19
25	.20-.39
34	.40-.59
18	.60-.79
3	.80-1.0

Note: 127 students were pretested and posttested.

Two independent reviewers analyzed the test questions for bias. Their biographies are in Appendix D. The results of the review indicate that the test questions are free of gender, cultural, racial, or ethnic language that would be offensive or misleading to the examinees.

Finally, three classes totaling 70 students reviewed the test and indicated which questions they had difficulty with. The development team reviewed the feedback from the students and concluded that no substantive changes were needed in the test questions.

SUMMARY OF PSYCHOMETRIC EVIDENCE

The purpose of the test is to determine which students pass or fail the NFA Chemistry of Hazardous Materials course and receive a course certificate. The test instrument has a high degree of validity and reliability for this purpose, based on the studies presented.

The test has item, sampling, and cognitive domain validity based on the judgments of nine experts. Based on Berk's (1976) Criterion-Groups Validation Model, the decision validity or probability of correctly identifying students that pass or fail the course is 96 percent based on the optimal cut score of 70; which results in a false negative probability of .024 and a false positive probability of .008. Based on this cut score the effectiveness of identifying non-masters from a non-master population is 98 percent and the effectiveness of identifying masters from a master population is 95 percent. These data are based on 127 Chemistry of Hazardous Materials students that were pre- and posttested.

The reliability of the pass/fail decision was determined using the P_o index of agreement. At the chosen cut score of 70 the reliability of the instrument is P_o=.97. The internal consistency reliability is KR-21=.839. The standard error of measurement is .57 based on a mean of 86, N=221. Based on Livingston's procedure, the standard error of measurement is 2.71 N=44 for scores between 50 and 80. These data are based on 221 Chemistry of Hazardous Materials students that were posttested.

In addition, the test questions have a high degree of effectiveness based on the sensitivity to instructional effects procedure. All items have a positive "S" score and 80 percent of the questions score between .20 to .80. These data are based on pre- and posttest scores of 127 NFA Chemistry of Hazardous Materials students. The instrument was found to be free of bias (gender, racial, and cultural) that may offend or confuse the students, based on the analysis of two reviewers. Finally, student evaluations (N=70) indicated little to no difficulty comprehending the questions, answers, or instructions to the test.

In conclusion, the NFA faculty that use this test instrument and the NFA students that take this exam can be confident in the test's validity and reliability to be used as the final exam for the NFA Chemistry of Hazardous Materials course.

A copy of the research report this manual is based on is available for review through interlibrary loan with the National Emergency Training Center Learning Resource Center, Emmitsburg, MD. The title of the report is Development of a Test and Establishment of a Cut Score for the National Fire Academy Chemistry of Hazardous Materials Course, Burton A. Clark, 1990.

CAUTION TO TEST USERS

This test instrument can be considered valid and reliable **only** for the intended purpose of the instrument, as explained in this manual. Various federal, national, state, and local regulations and standards refer to the NFA Chemistry of Hazardous Materials course as the training standard for emergency and nonemergency personnel. It must be clearly understood that successful completion of the course, passing the final exam, and receiving a course certificate does not certify, license, or predict the performance of an individual in an emergency or nonemergency work environment which requires knowledge, skills, and abilities associated with the chemistry of hazardous materials.

BIBLIOGRAPHY

American Psychological Association. Standards for Educational and Psychological Testing. Washington, D.C.: APA, 1985.

Belcher, Marcia. The Reliability of the CLAST. Miami, Florida: Miami-Dade Community College, 1984.

Berk, Ronald A. "A Consumer's Guide to Criterion-Referenced Test Reliability." Journal of Educational Measurement, 17:323-349. Winter, 1980.

Berk, Ronald. "Conducting the Item Analysis." In Ronald A. Berk (Ed.) A Guide to Criterion-Referenced Test Construction. (pp. 97-135) Baltimore, Maryland: The Johns Hopkins University Press, 1984a.

Berk, Ronald A. "Determination of Optional Cutting Scores in Criterion-Referenced Measures." Journal of Experimental Education, 45:4-9. 1976.

Berk, Ronald A. "Selecting the Index of Reliability." In Ronald A. Berk (Ed.) A Guide to Criterion-Referenced Test Construction. (pp. 231-265) Baltimore, Maryland: The Johns Hopkins University Press, 1984b.

Bloom, Benjamin S. Taxonomy of Educational Objectives Handbook I: Cognitive Domain. New York: David McKay Corp., Inc., 1956.

Cangelosi, James S. Designing Tests for Evaluating Student Achievement. White Plains, New York: Logman, 1990.

Chauncey, Henry and John E. Dobbin. "Testing has a History." In Clinton I. Chase and H. Glenn Ludlow (Ed.). Readings in Educational and Psychological Measurement. (pp. 3-18) Boston, Massachusetts: Houghton Mifflin Company, 1966.

Garrett, Henry E. Statistics in Psychology and Education. New York, New York: David McKay Company, Inc., 1971.

Gay, L.R. Education Research. Columbus, Ohio: Merrill Publishing, 1981.

Gronlund, Norman E. Measurement and Evaluation in Teaching. New York: MacMillan Publishing Co., Inc., 1976.

Gronlund, Norman E. Preparing Criterion-Referenced Tests for Classroom Instruction. New York: The MacMillan Co., 1973.

Guion, Robert M. Standards for Educational and Psychological Tests. Washington, D.C.: American Psychological Association, Inc., 1974.

Hambleton, Ronald K. "Validating the Test Score." In Ronald A. Berk (Ed.) A Guide to Criterion-Referenced Test Construction. (pp. 199-227) Baltimore, Maryland: The Johns Hopkins University Press, 1984.

Isaac, Stephen and William B. Mitchel. Handbook in Research and Evaluation. San Diego, California: Edits Publishers, 1981.

Kubiszyn, Tom and Gary Borich. Educational Testing and Measurement. Glenview, Illinois: Scott, Foresman and Company, 1990.

Livingston, Samuel A. "Estimation of the Conditional Standard Error of Measurement for Stratified Tests" Journal of Educational Measures 19:135-138, Summer 1982.

Livingston, Samuel A. and Michael J. Zieky. Passing Scores. Princeton, New Jersey: Educational Testing Service, 1982.

Popham, W. James. "Specifying the Domain of Content or Behavior." In Ronald A. Berk (Ed.) A Guide to Criterion-Referenced Test Construction. (pp. 29-40) Baltimore, Maryland: The Johns Hopkins University Press, 1984.

Resnick, Daniel. "History of Educational Testing." In Alexandra K. Wigdor and Wendell R. Garner (Ed.). Ability Testing: Uses, Consequences, and Controversies Part II. (pp. 173-174) Washington, D.C.: National Academy Press, 1982.

Shepard, Lorrie A. "Setting Performance Standards." In Ronald A. Berk (Ed.)." A Guide to Criterion-Referenced Test Construction. (pp. 168-196) Baltimore, Maryland: The Johns Hopkins University Press, 1984.

APPENDIX A

COURSE DESCRIPTION AND STUDENT SELECTION CRITERIA

Chemistry of Hazardous Materials (R234)

This two-week course provides the basic knowledge required to evaluate the potential hazards and behaviors of materials considered to be hazardous for one or a combination of reasons.

Directed at the underlying reasons for the chemical behavior of hazardous materials, the course is designed to improve decision-making, safety operations, and handling. The course is heavily chemistry oriented.

Student Selection Criteria: Emergency response personnel who have responsibility for analysis, management, and/or tactical response to a hazardous materials incident, or fire prevention inspection where substantial knowledge of the chemical behavior of hazardous materials is essential.

An understanding of basic chemistry is helpful in receiving optimum benefit from the course.

ACE Recommendation: In the upper division baccalaureate category, 4 semester hours in Engineering, Fire Science Chemistry, General Science, or Physical Science.

From: National Fire Academy Course Catalog, 1990-1991, p. 17.

APPENDIX B

EDUCATIONAL OBJECTIVES OF THE COURSE

OBJECTIVE 1.1 DOMAIN CLASSIFICATION 1.11

The participants will with reference to a periodic table identify symbols, names of elements, and atomic numbers.

OBJECTIVE 1.2 DOMAIN CLASSIFICATION 1.12

The participants will using a periodic table determine the logical, systematic order of elements.

OBJECTIVE 1.3 DOMAIN CLASSIFICATION 1.30

The participants will demonstrate an understanding of chemical bonding of atoms by balancing molecules containing atoms which form either Salts or Non-salts.

OBJECTIVE 2.1 DOMAIN CLASSIFICATION 1.31

The participants will after a lecture/discussion/reading and viewing Videotape 2 demonstrate their knowledge of ionic bonding by correctly naming Salts in several problems.

OBJECTIVE 2.2 DOMAIN CLASSIFICATION 1.23

The participants will given formulas or names balance compounds made from Salts.

OBJECTIVE 2.3 DOMAIN CLASSIFICATION 3.00

The participants will identify the five types of Salts and their hazards.

OBJECTIVE 3.1 DOMAIN CLASSIFICATION 2.10

The participants will apply the dash method correctly by using illustrations which depict the structure of compounds, given the name or the formula.

OBJECTIVE 3.2 DOMAIN CLASSIFICATION 2.10

The participants will distinguish a multiple bond within a compound by illustrating its structure correctly.

OBJECTIVE 3.3 DOMAIN CLASSIFICATION 2.10

The participants will correctly analyze a non-salt by determining its structure, isomers, bonds and shapes.

OBJECTIVE 3.4 DOMAIN CLASSIFICATION 4.20

The participants will identify specific non-salts and hydrocarbons given the name, formula, or structure.

OBJECTIVE 4.1 DOMAIN CLASSIFICATION 4.20

The participants will from formulas identify hydrocarbons and other organic compounds, and deduce the chemical characteristics that determine their hazardous properties.

OBJECTIVE 4.2 DOMAIN CLASSIFICATION 2.20

The participants will from formulas or structures identify the organic family to which the particular compound belongs.

OBJECTIVE 4.3 DOMAIN CLASSIFICATION 5.30

The participants will from formulas determine whether a compound has a saturated, unsaturated, or aromatic type bond.

OBJECTIVE 4.4 DOMAIN CLASSIFICATION 2.30

The participants will from the Carbon/Hydrogen ratio in the formula determine whether the compound has a straight, branched, or cyclic shape.

OBJECTIVE 6.1 DOMAIN CLASSIFICATION 3.00

The participants will recognize and assess certain physical properties of Flammable liquids and use those assessments to determine potential flammability.

OBJECTIVE 6.2 DOMAIN CLASSIFICATION 4.20

The participants will explain the relationship between boiling point and vapor pressure, and use that relationship to determine the amount of vapor produced by various liquids.

OBJECTIVE 6.3 DOMAIN CLASSIFICATION 3.00

The participants will determine boiling point from the chemical molecular characteristics of weight, polarity, and bonding for compounds within a family and between different families.

OBJECTIVE 6.4 DOMAIN CLASSIFICATION 3.00

The participants will predict changes in boiling point of certain miscible and immiscible solutions.

OBJECTIVE 7.1 DOMAIN CLASSIFICATION 3.00

The participants will use flash point, flammable range, and other parameters of burning to determine the ignition potential of various flammable liquids with varying concentrations.

OBJECTIVE 7.2 DOMAIN CLASSIFICATION 4.20

The participants will predict the quantity of vapor fuel needed to sustain combustion using flash point and the general parameters of burning.

OBJECTIVE 7.3 DOMAIN CLASSIFICATION 2.30

The participants will determine flash point of a flammable liquid using boiling point and certain other chemical and physical characteristics of the liquid.

OBJECTIVE 7.4 DOMAIN CLASSIFICATION 3.00

The participants will use flammable range to predict ignition properties and limits of various flammable liquids under different burning parameter conditions.

OBJECTIVE 8.1 DOMAIN CLASSIFICATION 5.30

The participants will use ignition temperature, flammable range, heat output and other fuel quality characteristics to predict the combustion hazard and fire behavior of various flammable liquids.

OBJECTIVE 8.2 DOMAIN CLASSIFICATION 5.30

The participants will determine relative ignition temperatures and ignition characteristics from the chemical composition of various flammable liquids.

OBJECTIVE 8.3 DOMAIN CLASSIFICATION 3.00

The participants will characterize and anticipate sustained combustion of flammable liquids based on flash point, ignition temperature, flammable range and chemical composition.

OBJECTIVE 8.4 DOMAIN CLASSIFICATION 5.30

The participants will predict the magnitude of heat output and its interaction with combustion from the chemical properties and fuel quality characteristics of various flammable liquids.

OBJECTIVE 8.5 DOMAIN CLASSIFICATION 5.30

The participants will given lists of compounds determine flash point, ignition temperature, and heat output values as a means of analyzing combustion. This analysis pertains to parameters within a family and within different families.

OBJECTIVE PL.1 DOMAIN CLASSIFICATION 1.23

The participants will identify the color of placards and relate the color to the appropriate DOT hazard category.

OBJECTIVE PL.2 DOMAIN CLASSIFICATION 4.0

The participants will interpret the placard rules for weight requirements (1,000 lb. rule).

OBJECTIVE WR.1 DOMAIN CLASSIFICATION 1.12

The participants will identify the alkali metals and alkaline earth metals from the periodic table provided.

OBJECTIVE WR.2 DOMAIN CLASSIFICATION 1.32

The participants will select the appropriate extinguishing agent for water reactives.

OBJECTIVE RAD.1 DOMAIN CLASSIFICATION 2.10

The participants will differentiate between the two major categories of ionizing radiation.

OBJECTIVE RAD.2 DOMAIN CLASSIFICATION 1.23

The participants will define isotope and identify what constitutes an isotope.

OBJECTIVE RAD.3 DOMAIN CLASSIFICATION 1.12

The participants will list the three protective measures to protect themselves from radiation.

OBJECTIVE PT.1 DOMAIN CLASSIFICATION 1.12

The participants will define the time parameter of exposure to poisons and toxics.

OBJECTIVE PT.2 DOMAIN CLASSIFICATION 1.12

The participants will identify the barriers to exposure to poisons and toxics.

OBJECTIVE OX.1 DOMAIN CLASSIFICATION 1.12

The participants will identify the hazards of oxidizers as individual chemicals and when mixed with a variety of other chemicals, and will be able to determine "worst case" scenarios.

OBJECTIVE OX.2 DOMAIN CLASSIFICATION 1.31

The participants will demonstrate a knowledge of the types of oxidizers other than those placarded and labeled by DOT.

OBJECTIVE FSD.1 DOMAIN CLASSIFICATION 1.21

The participants will identify the types of chemicals MOST likely to exhibit the properties of flammable solids and combustible dusts based on the chemical reactivity and/or physical states.

OBJECTIVE CRY.1 DOMAIN CLASSIFICATION 1.23

The participants will identify cryogenics that are not included in DOT classes.

OBJECTIVE CRY.2 DOMAIN CLASSIFICATION 1.23

The participants will identify the two major hazards of all cryogenics.

OBJECTIVE POL.1 DOMAIN CLASSIFICATION 1.12

The participants will identify the components of polymers.

OBJECTIVE FGL.1 DOMAIN CLASSIFICATION 1.12

The participants will identify the DOT class of flammable and combustible liquids.

OBJECTIVE FGL.2 DOMAIN CLASSIFICATION 4.20

The participants will determine the effects of temperature, pressure, conductivity, radiant heat on boiling point, flash point, and ignition temperature.

OBJECTIVE EX.1 DOMAIN CLASSIFICATION 3.0

The participants will identify the physical and chemical parameters required for the threat of explosive potential.

OBJECTIVE COR.1 DOMAIN CLASSIFICATION 1.12

The participants will apply the pH scale of measurement.

OBJECTIVE COR.2 DOMAIN CLASSIFICATION 1.12

The participants will list two types of decontamination.

OBJECTIVE COR.3 DOMAIN CLASSIFICATION 1.23

The participants will define concentration and strength of acid and base.

APPENDIX C

INSTRUCTIONS TO STUDENTS

1. Do not write on the test.
2. Record all answers on the answer sheet.
3. There is no time limit for taking the test.
4. You receive one point for each correct answer.
5. You must receive a 70 to pass the final and the course. Your official NFA record only indicates a pass or fail grade.
6. Answer all the questions.
7. You must return all materials, test, answer sheet, and scratch paper to the instructor.
8. You will be informed of your grade but you will not receive your test back.
9. The only reference material you may use during the exam is the periodic table.
10. Cheating will result in automatic failure.

APPENDIX D

BIOGRAPHIES OF DEVELOPERS AND REVIEWERS

NFA Development Team

Burton A. Clark — Program Chair, Management Science at the NFA for 10 years. B.S., Business Administration, Strayer College; M.A., Curriculum Instruction and Technology, Catholic University; Ed.D., Adult Education, Nova University.

Jan D. Kuczma — Assistant Superintendent, Leadership and Hazardous Materials Branch for the National Fire Academy for 10 years. B.A., Chemistry/ Biology, Lycoming College. Student Fellow, Brookings Institute for Advanced Studies and George Washington University School of Public Administration. High school chemistry teacher for 7 years. (Note: Participated in the Angoff procedure only.)

Noel P. Waters — Program Chair, Hazardous Materials at the NFA for 3 years. B.S., Fire Science and Administration, John Jay College; New York City Fire Department, 25 years, Lieutenant, Hazardous Materials Unit; high school science teacher, 17 years.

David Martin — Training Instructor in Hazardous Materials at the NFA for 3 years. B.A., Chemistry, West Virginia Wesleyan College; M.A., Science Education, West Virginia University. High school teacher for 25 years teaching chemistry, physics, and mathematics. Science Department Chairman for 10 years.

Independent Reviewers

David A. Nelson — Professor of Chemistry at the University of Wyoming, 28 years. B.S., Chemistry, Massachusetts Institute of Technology; M.S., Chemistry, University of Rhode Island; Ph.D., Chemistry, University of New Hampshire. NFA adjunct faculty, Chemistry of Hazardous Materials.

George C. Farrant — Professor of Chemistry of Catonsville Community College, Maryland, 19 years. B.A., Chemistry, Oberlin College; Ph.D., Chemistry, Case Western Reserve University. NFA adjunct faculty, Chemistry of Hazardous Materials.

Joe R. Callaway — Associate Training Specialist, Occupational and Environmental Safety Training Division, Texas A & M University, 10 years. B.A., Biology, North Texas State University; M.A., Biology, North Texas State University; Ph.D. (student) Curriculum and Instruction, Texas A & M University. NFA adjunct faculty, Chemistry of Hazardous Materials.

David M. Lesak — President and Program Manager of Hazard Management Associates, 10 years. B.S., Secondary Education (Biology and General Science), Kutztown State University. Fire Chief, Lower Macungie Township, 2 years. NFA adjunct faculty, Chemistry of Hazardous Materials.

Frank L. Fire — Marketing Manager, Americhem Inc., 27 years. B.S., Chemistry, University of Akron; M.B.A., University of Akron. NFA adjunct faculty, Chemistry of Hazardous Materials.

David L. Sealey — Senior Systems Analyst, Weeg Computer Center, University of Iowa, 12 years. B.S., Education, University of Wisconsin; M.F.A., Univeristy of Iowa; Ph.D., Instructional Design, University of Iowa. Adjunct Professor, department of education, University of Iowa. NFA adjunct faculty, Chemistry of Hazardous Materials.

Reviewers to Detect Bias

Paula McMann — Program Chair for Management Technologies at the NFA, 1 year. B.S., Information Sciences, University of Maryland. Ten years fire service experience in information management.

Noel Hart — Group Leader, Information and Technologies Program for the Emergency Management Institute, 5 years. B.A., Economics/English, New York University; Equal Opportunity Council, Federal Emergency Management Agency, National Emergency Training Center, 3 years; 25 years fire service experience.

Technical Manual Reviewers

Ronald A. Berk — Professor of Education, the Johns Hopkins University School of Nursing, 10 years. B.A., Political Science, American University; M.Ed., Administration and Supervision, University of Maryland; Ph.D., Educational Technology, Curriculum, Research, Measurement, and Statistics, University of Maryland.

William R. Cowell — Deputy Director of Professional Services, Center for Occupational and Professional Assessment, Educational Testing Service, 22 years. B.S., Mathematics and Education, Ohio State University; M.A., Mathematics, Michigan State University.