DOCUMENT RESUME

ED 324 213 SE 051 592

TITLE The North Carolina Test of Chemistry. Technical Characteristics. Forms A-D.
INSTITUTION North Carolina State Dept. of Public Instruction, Raleigh. Div. of Accountability Services/Research.
PUB DATE 90
NOTE 42p.
PUB TYPE Reports - Descriptive (141) -- Statistical Data (110)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Chemistry; *Curriculum Evaluation; Science Education; Secondary Education; *Secondary School Science; *Test Construction; Test Content; *Testing Programs; Test Norms; Test Reliability; *Test Theory; Test Validity
IDENTIFIERS *North Carolina Test of Chemistry

ABSTRACT

To facilitate the proper technical use of the test scores obtained from the administration of the tests, the curricular and psychometric characteristics of the tests are described in a series of technical manuals. This manual, the seventh in the series, contains a description of the characteristics of the North Carolina Test of Chemistry. The test was developed for use as an achievement test following the completion of the Chemistry course of study. Its design was to serve as a normative measurement of student achievement and as an objective-based measurement of curriculum coverage. The curricular, instructional, content, and concurrent validity of the test are discussed. Specific issues covered include: (1) the method for deriving test scores; (2) reliability and other statistics; (3) a curriculum assessment; (4) content of the test; and (5) test norms. The utility of the test has been determined by its statistical equivalence of core tests from year to year, its broad sampling of the curriculum across time, and its initial norms table. The goals and objectives used to assemble the test and the number of questions associated with each is appended. The percent of teachers reporting each goal/objective as basic to instruction in Chemistry as well as a list of goals/objectives that were rejected for use are also appended. (KR)

Forms A – D

Technical Characteristics of the North Carolina Test of Chemistry

NCTests

North Carolina Department of Public Instruction
Division of Accountability Services/Research
Raleigh, NC 27603-1332

Published 1990

Technical Characteristics of the

North Carolina Test of

Chemistry

Prepared by
Eleanor E. Sanford

NCTests

North Carolina Department of Public Instruction
Division of Accountability Services/Research
Raleigh, NC 27603-1332

FOREWORD

The NCDPI Division of Accountability Services/Research, in cooperation with the NCDPI Division of Curriculum and Instruction Services, has developed diagnostic achievement tests of basic skills for public school students in Grades 3, 6, and 8; survey achievement tests of Science and Social Studies for students in Grades 3, 6, and 8; and high-school course achievement tests for students taking Algebra I, Algebra II, Biology, Chemistry, English I, Geometry, Physics, and U.S. History. Physical Science and Economic, Legal, and Political Systems will be added in 1991, and other tests are being planned.[a]

To facilitate the proper technical use of the test scores obtained from the administration of the tests, the curricular and psychometric characteristics of the tests are described in a series of technical manuals. This manual, the seventh in the series, contains a description of the characteristics of the North Carolina Test of Chemistry.

[a]Readers who have an interest in the origins of the test development program are referred to the North Carolina Elementary and Secondary Reform Act of 1984, the North Carolina Basic Education Program, the North Carolina *Standard Course of Study*, and the *Teacher Handbook*.

Table of Contents

List of Tables

List of Figures

Description

The North Carolina Test of Chemistry (NCT-Chemistry) was developed for use as an achievement test following the completion of the Chemistry course of study. Its design serves a dual purpose: that of a normative measurement of student achievement and of an objective-based measurement of curriculum coverage.

The measurement of student achievement is attained by administering a basic core of 60 items to all students. The measurement of curriculum is met by an additional 40 items that vary across the four forms of the test. All four forms of the test, each form containing the same 60 core items and 40 variable items, are administered in each classroom, one form per student. Under this system, a fourth of the students in a classroom will take Form 1 of the test, a fourth will take Form 2, and so on (see Table 1).

Table 1
Organization of the North Carolina Test of Chemistry

60 Core Items			
40 Variable Items	40 Variable Items	40 Variable Items	40 Variable Items
100 Items Form 1	100 Items Form 2	100 Items Form 3	100 Items Form 4

The normative student scores are based on the 60 core items that all forms have in common. Curriculum assessment is achieved by combining the results from all four forms, which provides an assessment based on the 60 core items + 4(40 variable items), or 220 items in all.

Validity

The development of a Chemistry achievement test has two purposes. The first is to obtain scores from which inferences may be drawn concerning the degree of success a particular student, classroom, school, or school district has had in mastering the Chemistry curriculum. The second is to assess the degree to which the curriculum has been mastered by students in the aggregate. To the extent this can be done meaningfully, test scores may be said to be valid. Thus, one inference drawn from a test score may be valid, while another may not.

Theoreticians state that only inferences concerning test scores may be said to have validity. Generally, readers understand this, and this manual will employ the convenient shorthand of speaking about "test validity" rather than "inferences about achievement drawn from scores obtained from tests."

Test validity is a predominant theme in test development, from the time the idea for a test is conceived until the final test scores have been analyzed and interpreted. For convenience, the various components of test validity will be described as if they were unique, independent components rather than interrelated parts. The first component of test validity to be described will be curricular validity.

Curricular validity

If a test is to be used to measure the degree to which a course of study has been mastered, the first step is to define the curriculum. In the case of Chemistry, that was done through a cooperative effort, led by NCDPI Program Services/Division of Curriculum and Instruction Services, involving curriculum specialists, teachers, administrators, university professors, and others. The result was a list of nine goals encompassing 37 objectives (see the Appendix). Supported by expert opinion and a statewide consensus, these goals and objectives were approved by the State Board of Education in 1985 as the basis for instruction in Chemistry. Curricular validity, the first step in establishing construct validity, was established by this method.

Instructional validity

A basic course of study may not include all of the objectives taught under various circumstances in Chemistry. For example, some advanced classes may cover material that would be beyond the reach of 95% of all Chemistry students. For this reason and several others, it becomes important to know just what is being taught in the majority of Chemistry classes in the state. To determine this, all Chemistry teachers in North Carolina were surveyed in September 1987 (N = 498). The analysis of results was based on 289 responses, or 58% of all possible responses.

The Chemistry teachers examined the 37 objectives and noted whether they taught each objective every year and whether each objective was basic to instruction. 30 of the objectives were rated as basic to instruction by at least 55% of the responding teachers. After deliberation by curriculum specialists, it was determined that the 9 goals and 30 objectives formed the basic curriculum for Chemistry and the remaining seven objectives would only be used for curriculum assessment, not for students' scores. The objectives are given in the Appendix together with the proportion of teachers judging each one as basic.

Instructional validity, the second step in defining construct validity, was established by these procedures. It defines the inferences that can be drawn from the Chemistry test scores.

In summary, it was concluded that curricular and instructional validity depended jointly on the nine goals and 37 objectives under which they were collected, and that the Chemistry test should be built on that foundation.

Content validity

Content validity—the degree to which test items reflect the basic instructional program—is a quality commonly referenced in evaluating achievement tests. Content validity is built into a test from the beginning, and the procedures relating to the content validation of the North Carolina Test of Chemistry are described below.

Content validity of the item pool (Phase One). The content validity of the item pool was defined through a number of operations:

First, the item pool for the Chemistry test was created in 1987. It was specified that the pool would have 1,221 items, with 33 items per objective. The items were developed by 14 North Carolina Chemistry teachers trained in the technical aspects of item-writing. The use of classroom teachers from across the state helped to insure that instructional validity was maintained, since the items would be drawn from their classroom experiences. All total, 1,185 items were actually written for the Chemistry item pool.

Second, the item pool was edited for grammar, syntax, psychometric form, linguistic bias, and subject area content. Nine items were deleted from the Chemistry item pool at this time.

Third, the item pool was analyzed by curriculum specialists and classroom teachers to assure that the items were valid representations of the objectives for which they were written. Each item was reviewed by eight classroom teachers—one from each of the eight educational regions across the state. The criteria for evaluating each item included the following:

- conceptual—objective match, fair representation, lack of cultural bias, clear statement, single problem, one best answer, common context in foils, each foil credible
- language—appropriate for age; correct punctuation, spelling, and grammar; lack of excess words; no stem/foil clues; no negatives in foils
- format—logical order of foils; familiar presentation style, print size, and type; correct mechanics and appearance; equal length foils
- diagram—necessary, clean, relevant, unbiased

164 items were deleted from the Chemistry item pool at this time.

Fourth, the items were collected into 12 test forms for field testing. Although the forms were not the final forms of the North Carolina Test of Chemistry, they were organized in such a way that the objectives were represented equitably across all forms. Each form contained 92 to 96 items, 10 of which were common across all forms for the purpose of ability equating should that become necessary.

Fifth, test administration instructions were written, distribution procedures were organized, and administrators were trained to conduct the test administration. The test administration organization used to administer statewide tests in North Carolina was employed to do the field testing. The administration of the field test forms followed the routine eventually expected to be used when the test of record was given.

Sixth, a sample of 8,654 students was selected to take the 12 field test forms containing a total of 1,012 items. To insure broad representation, schools were selected from each of the eight North Carolina educational regions and were representative of the state based on criteria that were judged to be related at least partially to Chemistry ability levels—school performance on the 1987 NCT–Algebra II and the 1987 NCT–Biology. The 12 field tests were interleaved in all students samples, and this produced an even spread of ability across all of the tests. Consequently, each item was answered by approximately 721 students (the number of students per field test form ranged from 666 to 761).

Seventh, the field test data were analyzed using both the classical psychometric model and the one-parameter Rasch model (results were generated from the BICAL computer program). Eighteen statistics were assembled for each item, i.e., p-value, Rasch difficulty index, adjusted Rasch difficulty index, standard error of the mean, fit mean-square, item validity (point-biserial correlation), and the item characteristic curve groupings. Item bias due to gender or ethnicity was examined by computing the partial correlation between the item score and gender/ethnicity while controlling for total score.

The item statistics were submitted to computer analysis using a program designed to scan the statistics for an item and print out an appropriate psychometric notation based on the criteria that had been built into the program specifically for Chemistry. An item was classified as "too hard" if the p-value was less than .27 or as "too easy" if the p-value was greater than .93. An item was said to have "weak prediction" if the point-biserial correlation was less than .13. An item was said to have an "entrapment choice," a "marginal top group," or an "inverted ICC" if the item characteristic curve groupings displayed certain irregularities. An item was said to exhibit "gender" bias if the partial correlation with gender was more extreme than ±.145, and to exhibit "ethnic" bias if the partial correlation with ethnicity was more extreme than ±.1375.

The content of Chemistry cannot be represented by a single factor. Therefore, maximization of item-total (point-biserial) correlations was not a goal of item development. Once an item was shown to have at least modest correlation with a corrected total score (.13 or greater) and was judged to measure an objective, it was included in the item pool. While this may have reduced the potential internal reliability as measured by coefficient alpha, it increased the validity of the test by allowing for an objective factor structure that was not expected to be unitary.

This information was placed on an item record, which became the basic document to which all other records are referred. The item record contains the goal, objective, historical information, a copy of the item itself, the item field-test statistics, and the psychometric notations. Each item has a separate item record.

The psychometric notations were reviewed and decisions were made about the adequacy of the items. The decisions were then conveyed to curriculum specialists, who also reviewed the items and reached a decision about their curricular adequacy. The psychometric and curricular decisions concerning the item's adequacy for use in test development were included on the item record.

Of the 1,012 items field-tested in 1988, 324 (32%) were deleted from the item pool. This left 688 items (68%) in the Chemistry item pool for future test development.

Content validity of the item pool (Phase Two). Due to the large number of items that were deleted from the item pool in 1988, it became necessary to develop additional items for the item pool. The greatest need was for items in the higher range of p's.

First, the number of items needed per objective was determined from the existing item pool. The number needed per objective ranged from 19 to 27 for a total of 770 items. 769 new items were written by 11 previously trained North Carolina Chemistry teachers.

Second, the additional item pool was edited for grammar, syntax, psychometric form, linguistic bias, and subject area content. 51 items were deleted from the additional item pool at this time.

Third, the additional item pool was analyzed by curriculum specialists and classroom teachers to insure that the items were valid representations of the objectives for which they were written using the procedures previously described. 68 items were deleted from the additional item pool at this time.

Fourth, the items were collected into 16 sets with 40 items each set (the exception was set #16 which had 24 new items and 16 repeated items from set #1) for a total of 624 items. These sets were used as variable items during the 1990 statewide administration, and the field-test forms were interleaved with the statewide forms to insure an even spread of ability levels. The data from these field tests will be analyzed in 1990 and used in constructing the 1991 NCT-Chemistry and future tests. Of the new items in the item pool, 26 were held for later field testing because they were similar to items contained in the core being used for statewide assessment (Core 2).

Content validity of the test. After a consideration of the logistics involved, it was decided to prepare one complete test (60 core items and four sets of 40 variable items) for administration in May 1989, and to develop four additional core tests of 60 items each for use in succeeding years. The initial core and the future core tests were based on a random selection of the 30 objectives rated as basic to instruction in Chemistry, for a total of 60 items (2 per objective) randomly chosen from the approved item pool. The four variable sets consisted of four items per objective for the 30 basic objectives and six items per objective for the seven remaining objectives randomly chosen from the approved item pool. Thus, the content of the test cores and the test forms directly reflected all of the decisions that had been made earlier.

This method of item selection is a modified domain sampling model, with the various forms and cores randomly equivalent. The domain sampling model in its pure form is highly inefficient because it allows the entry of items that are grossly inappropriate for normative measurement—items that no one can answer or that everyone can answer, or items that have psychometric deficiencies of a more complex form. In the modification used here, the domain of items was limited to those items that had satisfactory psychometric and curricular characteristics. This was determined by the analyses of the item field-test data, which was used to verify the psychometric and curricular adequacy of the item pool and to direct where item revisions should be made.

After the test was assembled into forms (60-item core plus a 40-item variable set), the four forms were reviewed by one curriculum supervisor and two teachers in each of the eight educational regions. The criteria for evaluating each form of the test included the following:

- that the content of the test should reflect the goals and objectives taught
- that the items should be clearly and concisely written, and the vocabulary appropriate to the target age level
- that the content should be balanced in relation to ethnicity, sex, socioeconomic status, and geographic district of the state
- that each item should have one and only one answer that is right; however, the distractors should appear plausible for someone who has not achieved mastery of the represented objective

The ratings for the 1989 North Carolina Test of Chemistry were average to superior on all of the criteria.

Although the initial equating of the core tests depended upon the random selection of items from the item pool, the final equating was based on the statistics obtained at the time the first test of record was administered (see Table 2). This second psychometric analysis, described next, was used to eliminate random differences among the cores and thus facilitate the precision of measurement from one year to another.

Table 2
Core Development of the North Carolina Test of Chemistry

Core	Process	Mean of P-values for All Items	SD of P-values of All Items	Sum of P-values of All Items
1	Design[a]	0.592	0.151	35.50
	Administer[b]	0.625	0.158	37.47
2	Design[a]	0.589	0.160	35.34
	Field Tested[b]	0.584	0.170	35.06
	Equate[c]	0.623	0.139	37.38
3	Design[a]	0.592	0.160	35.50
	Field Tested[b]	0.565	0.174	33.88
4	Design[a]	0.591	0.157	35.45
	Field Tested[b]	0.580	0.171	34.82
5	Design[a]	0.591	0.144	35.43
	Field Tested[b]	0.584	0.152	35.04

[a]Based on the 1988 Field Test item p-values
[b]Administered in May 1989
[c]Equated to Core 1

Standardization sample. The first North Carolina Test of Chemistry consisted of four forms (Core 1/A–D), each form containing the same 60 core items and a unique set of 40 variable items. This test was administered to 33,352 North Carolina Chemistry students in May 1989. The state norm population comprises these 33,352 students.

The four additional cores (Cores 2–5) that were developed were interleaved in all student samples. A sample of 3,194 students was selected to take the four field core tests. To insure broad representation, schools were selected from each of the eight North Carolina educational regions and were representative of the state based on a criterion that was judged to be at least partially related to Chemistry ability levels—school performance on the 1988 NCT–Biology. A total of 2,370 students actually participated in the core field testing (each core was administered to an average of 593 students). This produced an even spread of ability across all four samples (as determined by the results of the statewide test administration—Core 1/Forms A–D. The results are as follows: Sample 2, mean = 36.86, $s = 8.32$; Sample 3, mean = 36.87, $s = 8.43$; Sample 4, mean = 36.73, $s = 8.04$; Sample 5, mean = 36.80, $s = 8.29$). The four field core means were not equivalent and this is due to random differences in item difficulty that generally occurs between field testing and statewide testing. These random differences can be eliminated by redeveloping the tests slightly, and this procedure will be described in the Reliability and Other Statistics section.

The lack of agreement between the mean core test scores (Cores 2–5) and the state norm mean (Core 1) indicated that the field test samples were not representative of the North Carolina Chemistry student population. Upon further examination it was determined that this unusual finding occurred because not all schools selected to participate in the field testing had actually participated. The resulting score difference based on work with other North Carolina Tests and the differences in the Core 1 means for the matched samples compared with the Core 1 statewide mean was approximately 1.5 items (see Table 3 and the discussion above). After adjusting for the relative ability of the students in the field core samples, the agreement of the mean core test scores supported the view that the initial equating process was successful.

Concurrent validity of the test

When the 1989 North Carolina Test of Chemistry was administered, Chemistry teachers were asked to indicate the expected final letter grade for each student in their classes. Figure 1 displays a comparison of letter grades in Chemistry and the mean NCT-Chemistry core score corresponding to each letter grade for the overall student population. The figure corresponds closely to expectation and adds to the evidence concerning the validity of the test.

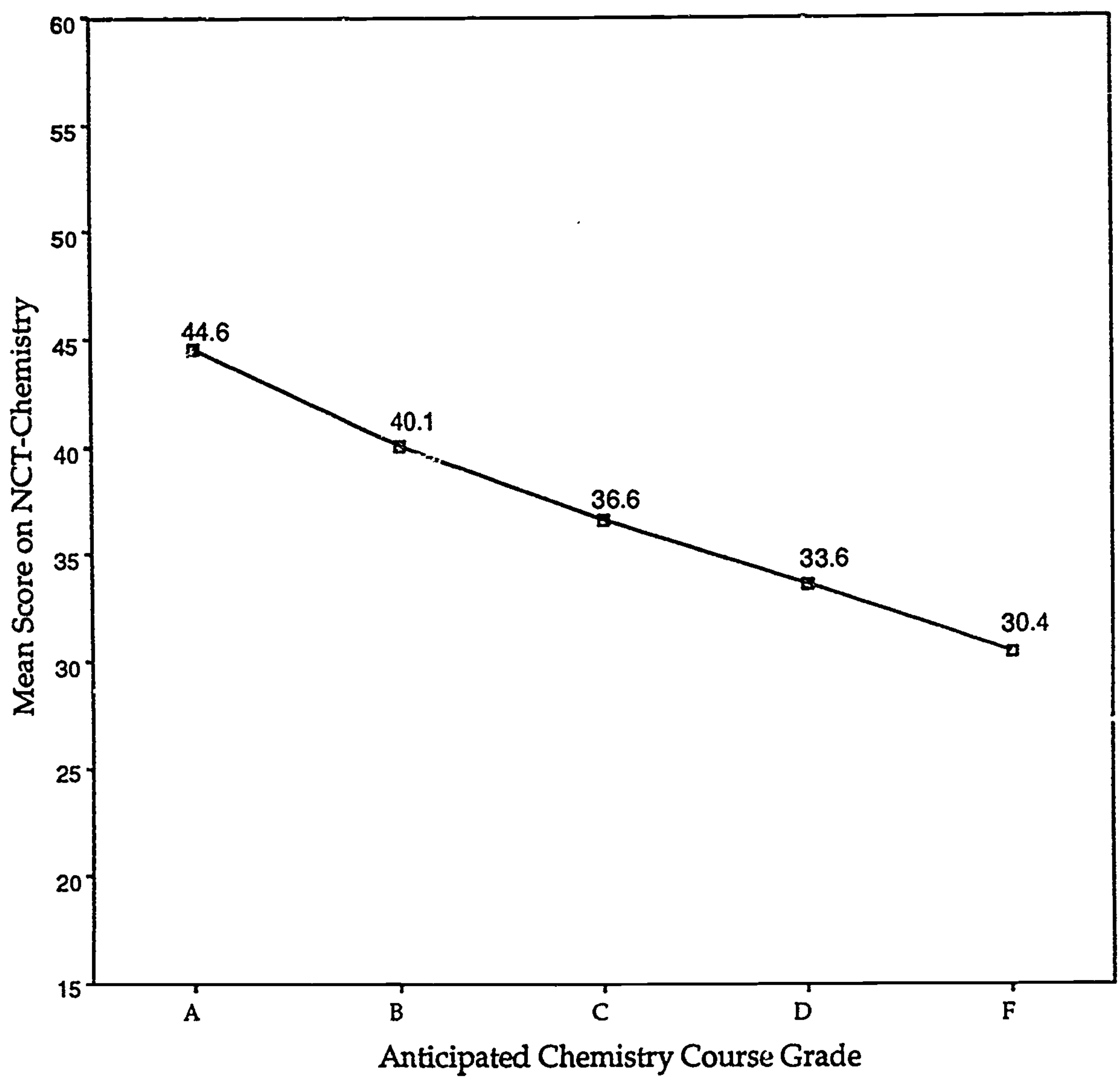

Figure 1. Comparison of letter grades teachers expected students to receive and scores subsequently earned on the 60-item 1989 North Carolina Test of Chemistry (N = 32,924).

Method for Deriving Test Scores

Item information was available to support the classical scoring model and the Rasch scoring model. The classical scoring model gives a unitary weight to each item; a correct choice adds one to the total score; an incorrect choice adds zero. The one-parameter Rasch model also uses unitary weighting. (The two- and three-parameter item response models give more credit for answering some items correctly and less credit for answering other items correctly. These models assume that each item has a fundamental, unchanging difficulty level.)

The classical scoring model was utilized to score the North Carolina Test of Chemistry because it is fundamentally sound, simple to use, and easy to interpret. Each student's total core score consists of the sum of right answers to the 60 core items.

Reliability and Other Statistics

The descriptive statistics, the standard errors of measurement, the alternate form reliability estimates (correlation between Core 1 and the other four cores field-tested in May 1989), and the alpha reliability coefficients for the first statewide administration of the North Carolina Test of Chemistry in May 1989 (Core 1/Forms A–D) and the cores field-tested in May 1989 (Cores 2-5) are given in Table 3.

Table 3
Descriptive Statistics of the North Carolina Test of Chemistry

Core	N	Mean	SD	Median	se_{meas}	Reliability Alternate Form	Reliability Coefficient Alpha
1/A-D	33,352	37.47	8.37	38	3.45	.75, .77, .72, .78	.83[b]
2	589	35.07	8.41	35			
	580[a]	35.10	8.37		3.45	.75	.83
3	597	33.82	8.74	34			
	578[a]	33.86	8.66		3.35	.77	.85
4	600	34.70	7.95	35			
	586[a]	34.90	7.81		3.40	.72	.81
5	584	35.03	8.42	36			
	569[a]	35.13	8.39		3.46	.78	.83

[a]Matched samples of students that took statewide core 1 and one of the field cores (2-5).
[b]Average of reliability coefficients obtained from the four matched samples of students (Core 1 with Cores 2-5). Core 2, Coefficient Alpha = .82; Core 3, Coefficient Alpha = .84; Core 4, Coefficient Alpha = .82; and Core 5, Coefficient Alpha = .83.

The alternate form reliability estimates have a mean value of .76; and the alpha reliability estimates have a mean of .83. While these reliability estimates are lower than those found on other tests (NCT–Algebra I, NCT–Biology, and NCT–US History), they are acceptable. The lower estimates can be attributed to the restricted range of ability of the students taking the North Carolina Test of Chemistry. Chemistry is not a required course for high school graduation and, therefore, students taking Chemistry are a select group of students when compared to the general population of high school students taking other North Carolina Tests. The 1989 core scores are symmetrically distributed about a mean of 37.47, or 62% correct (see Figure 2).

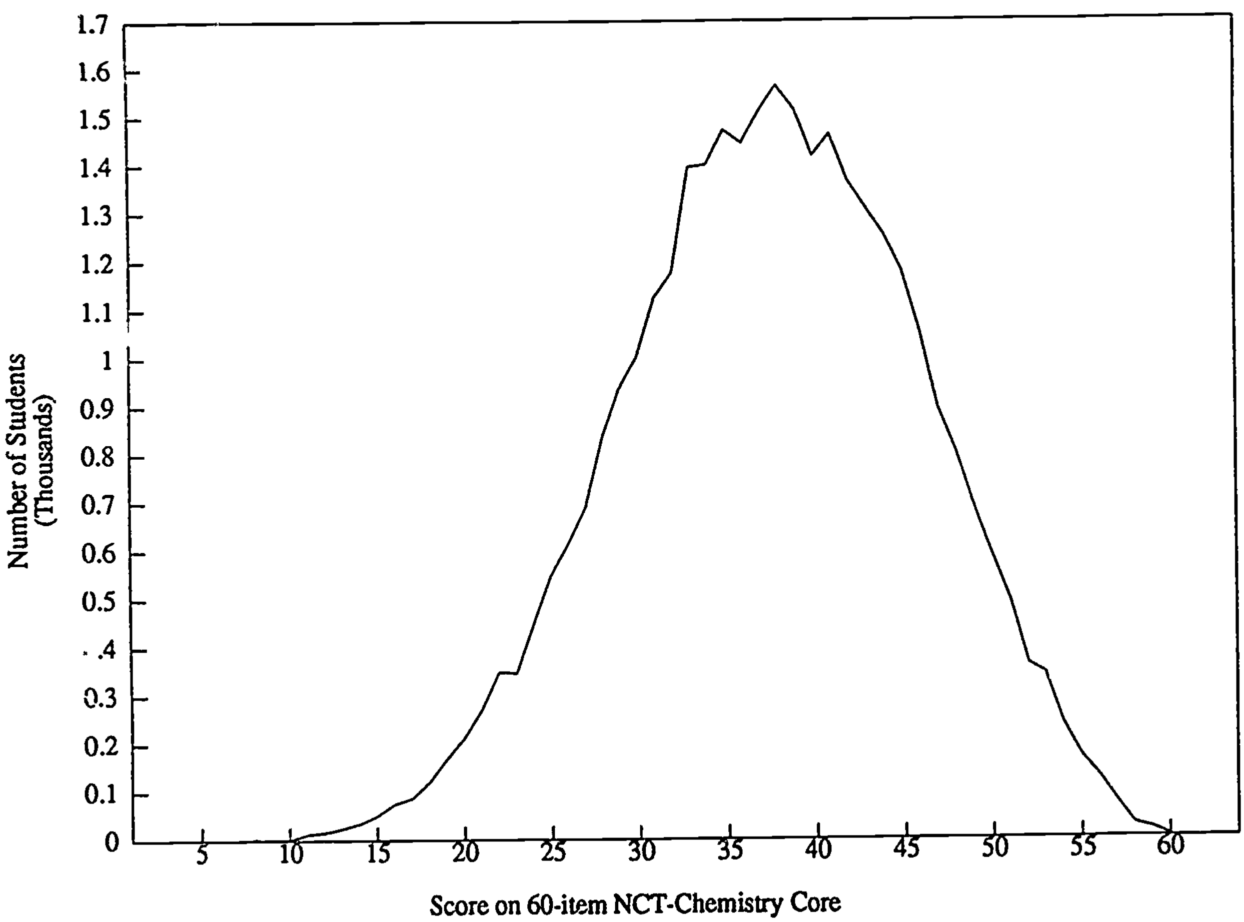

Figure 2. Frequency distribution of scores on the 60-item 1989 North Carolina Test of Chemistry—Core 1/Forms A–D (N = 33,352.

For practical purposes, the proper measure of reliability is the alternate form reliability. The calculation of this statistic requires that two or more equivalent forms be developed. The older alternate form reliability procedure required the development of one form, which was then "cloned" to obtain a second, alternate form of the test. A judicious selection of alternate items was recommended to prevent direct memory transfer from the administration of one test to its alternate form. But the possibility remained that errors of selection in the first form would be duplicated in the second form. A newer procedure requires that the tests be truly equivalent—that is, that two or more tests be developed in exactly the same way, but independent of one another. This permits the reliability coefficient to reflect any random errors in section made in the development of either of the test forms.

The alternate forms developed for the North Carolina Test of Chemistry reflect this newer procedure. That is, each test form is developed from the domain of items in exactly the same manner. Any failure of the alternate form reliability to be 1.00 reflects:

- trait instability not following from maturation or instruction
- instrument instability resulting from fallible test development procedures
- administrative instability reflecting different testing occasions

The square of the alternate form reliability coefficient accurately reflects the maximum proportion of variance one can legitimately expect to predict from the administration of the North Carolina Test of Chemistry ($r^2 = .83^2 = .69$) when test scores are compared across time or with other measures of student abilities or personality traits that have similar reliabilities. In brief, the alternate form reliability coefficient is the statistic to use when correcting for attenuation.

Of special significance to the comparison of students scores across time is the equivalence of the four future core tests (Cores 2–5) to the first core test of record (Core 1 administered May 1989). An equipercentile analysis was made of the relationship of the four future core tests to Core 1 (core test of record). Matched samples of student scores were obtained for each of the four future cores that consisted of student's scores on the statewide core and one of the four future cores (580 students were administered Core 1 and Core 2; 578 students were administered Core 1 and Core 3; 586 students were administered Core 1 and Core 4; and 569 students were administered Core 1 and Core 5—see Tables 2 and 3). To make the equipercentile comparison, the mean of a block of scores on Core 1 within groups of five successive percentile points was taken to compare with the mean of a block of scores on Core 2 within the same five percentile points. This procedure yielded twenty reasonably reliable points of comparison. This procedure was repeated for the matched samples from Cores 3–5. Each of the semidecile scores was adjusted for differences in relative ability of the four field core samples by adding 1.5 points. The results of the equipercentile comparisons of Core 1 with Cores 2–5 are presented in Figures 3, 5, 6, and 7.

In Figures 3, 5, 6, and 7, the differences of the data points from perfect agreement (a slope of 1.00) are small. These differences could be adjusted statistically by providing a separate set of norms for each test. A simple and efficient alternative is to redevelop the core tests slightly so that even the small differences disappear. With this technique, a single norms table can be used for all five core tests (20 test forms). To accomplish this transformation, the test developer had statistical information available on the 60 items in Core 1 plus the 160 variable items—a total of 220 items for which comparable psychometric data were available across all four future core tests.

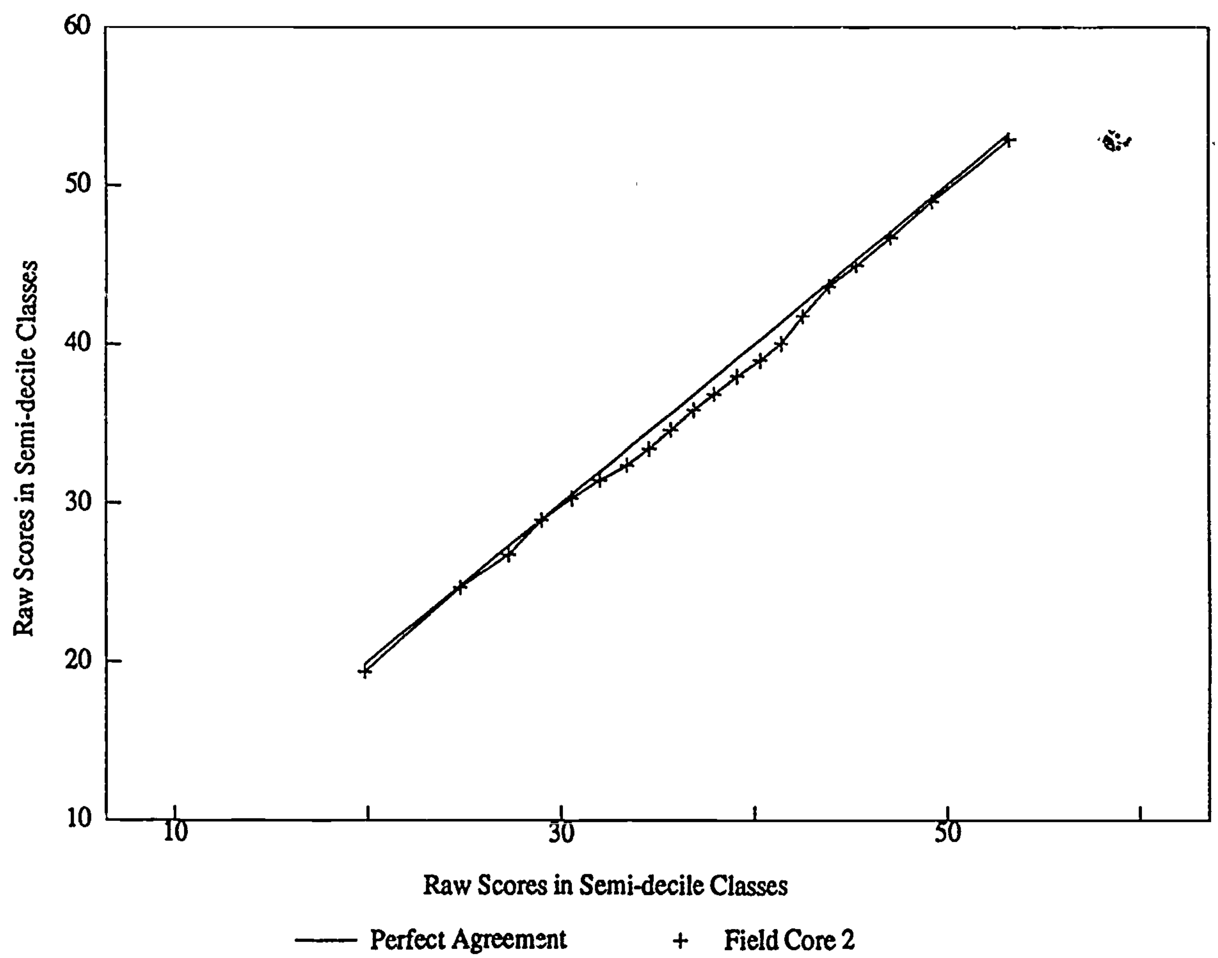

Figure 3. Equipercentile comparison of the 1989 North Carolina Test of Chemistry—Core 1/Forms A–D and Field Core 2 (adjusted for unrepresentativeness of the sample).

The results of the equating for Core 2, employed in 1990, are presented in Table 2 and Figure 4. The required changes were minimal (the substitution of 4 items).

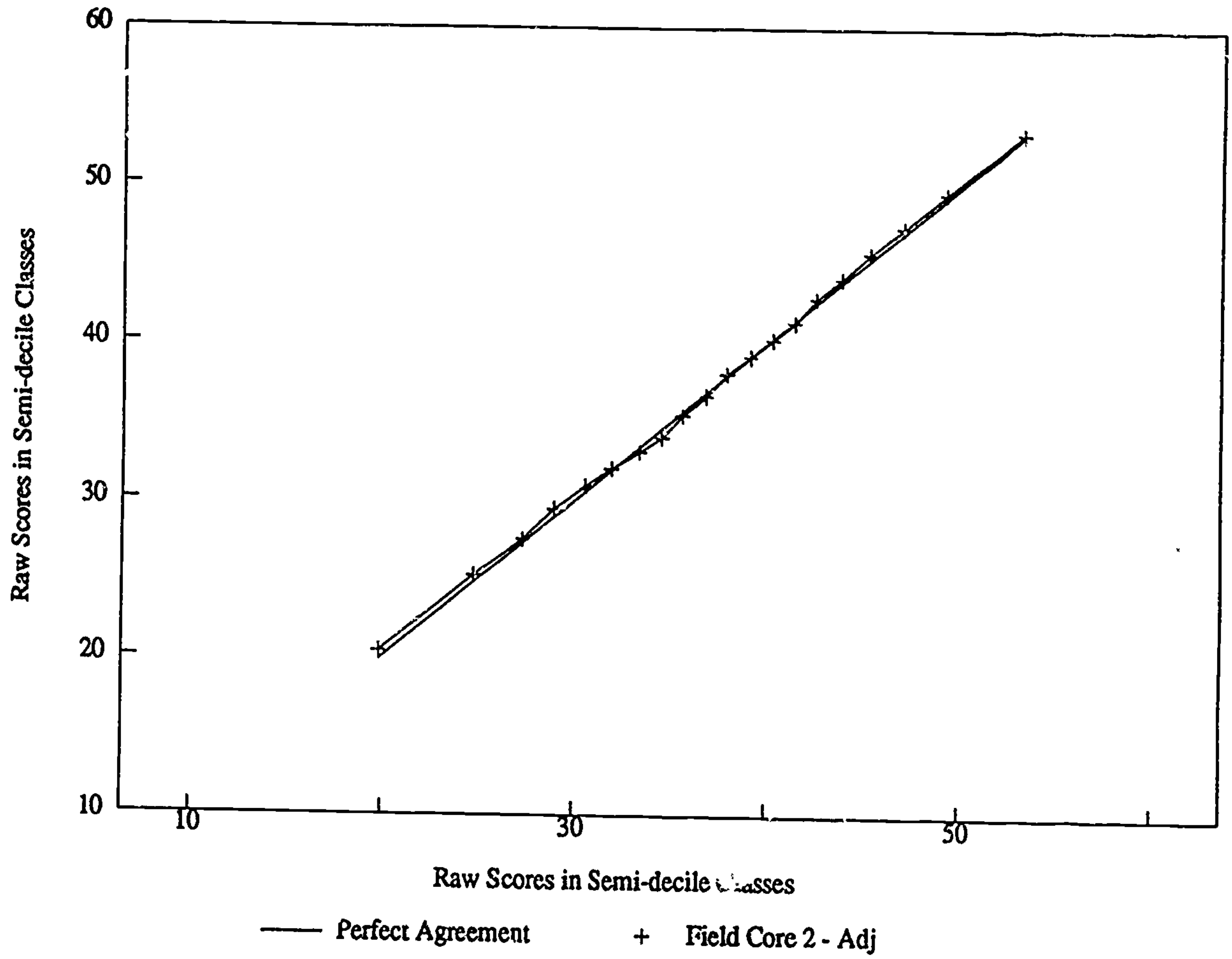

Figure 4. Equipercentile comparison of the 1989 North Carolina Test of Chemistry—Core 1/Forms A–D and Field Core 2 Equated (adjusted for unrepresentativeness of the sample).

Similar adjustments based on the 1989 administration of Core 1 and the field test administration of Cores 3–5 will be made to the third, fourth, and fifth core tests as needed in the future. The adjustments to the core tests will assure continuity of the norms table for future years while providing new test items each year. The new test items prevent the loss of confidentiality, and therefore validity, that occurs with the continued use of the same items. Student scores have a common reference point from 1989 onward, barring changes in the definition of the basic instructional program.

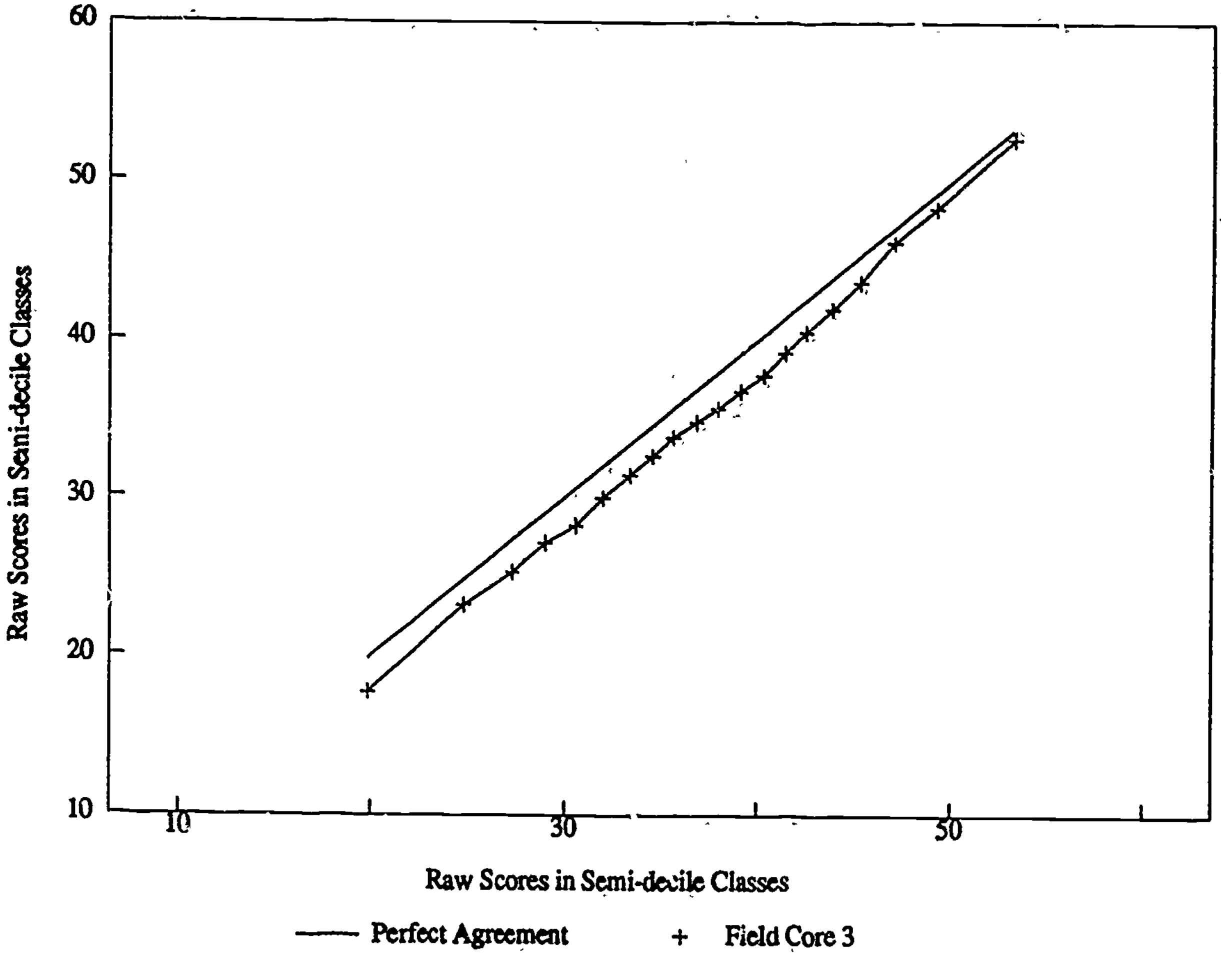

Figure 5. Equipercentile comparison of the 1989 North Carolina Test of Chemistry—Core 1/Forms A–D and Field Core 3 (adjusted for unrepresentativeness of the sample).

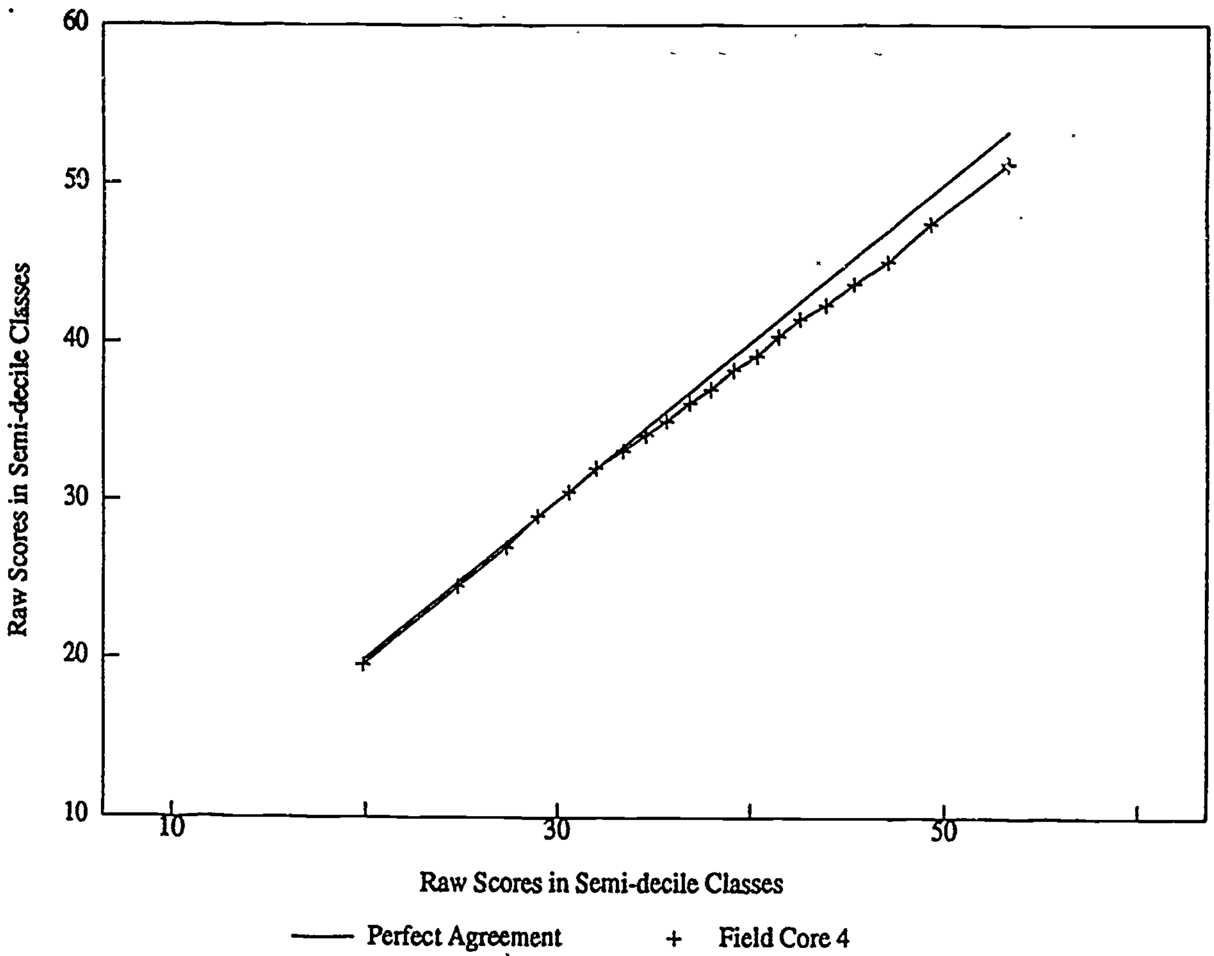

Figure 6. Equipercentile comparison of the 1989 North Carolina Test of Chemistry—Core 1/Forms A–D and Field Core 4 (adjusted for unrepresentativeness of the sample).

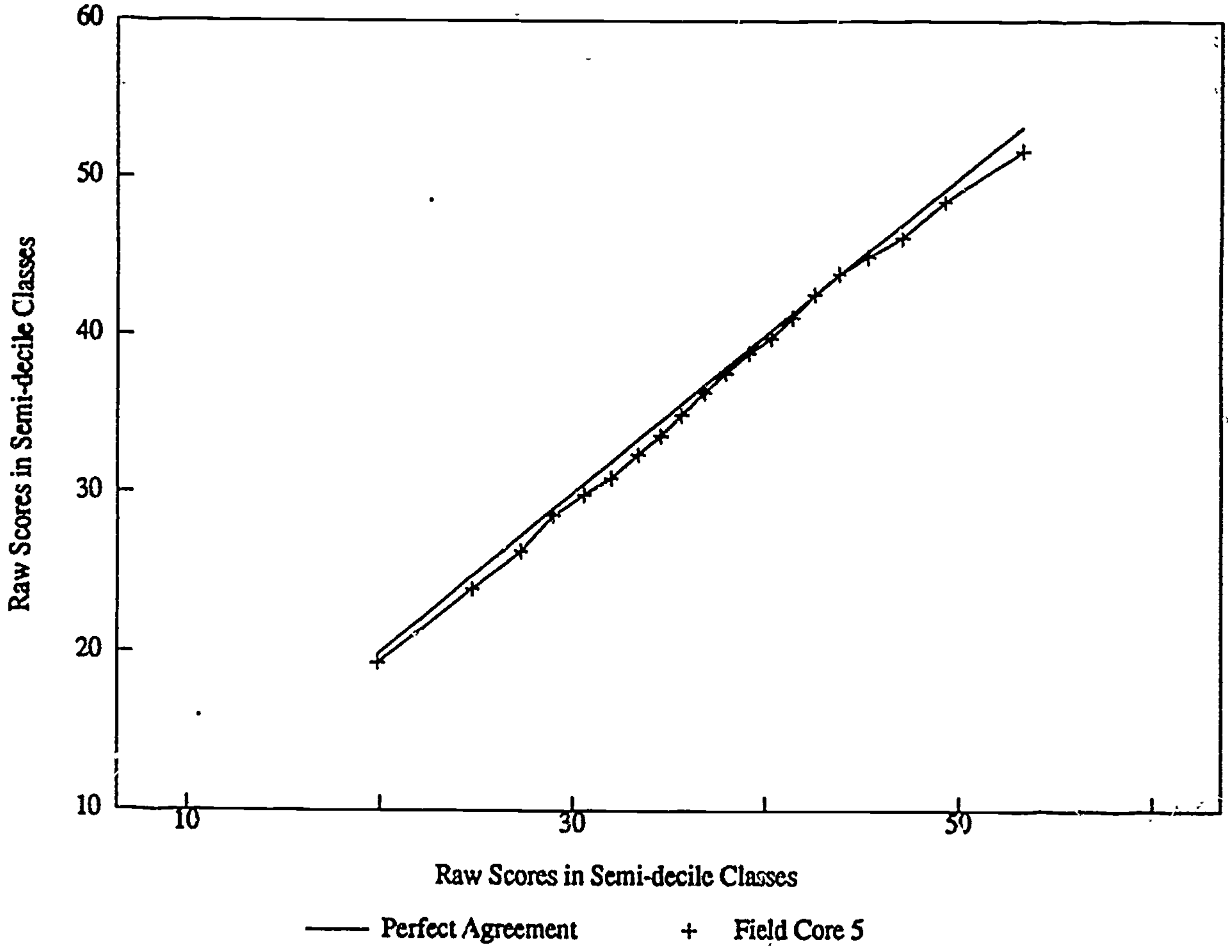

Figure 7. Equipercentile comparison of the 1989 North Carolina Test of Chemistry—Core 1/Forms A–D and Field Core 5 (adjusted for unrepresentativeness of the sample).

Curricular Assessment

Thus far in this manual the 60-item core test has been discussed as if it made up the entire test. In actuality, the four test forms of the 1989 North Carolina Test of Chemistry consisted of a 60-item core and 40 additional items that varied across the four forms (see Table 1). The variable items were not intended to contribute to individual student scores, but rather to curriculum assessment. Each variable item was answered by one-fourth of the students.

At the classroom level, 220 items are answered during each test administration by an average of five students. This procedure provides a database of six items per objective across five students. From this database of information, evidence of how various portions of the curriculum are being mastered in the classroom may be drawn. At the school, school district, and state level, the 220 items are answered by a much larger number of students: over 8,000 students per item. This assures a more stable measurement, but does not include a larger number of objectives or items. The accumulation of item and objective information depends upon measurement across successive years.

The measurement afforded by the 160 variable items is critical to assessing curriculum mastery at the classroom, school, school district, and state levels. Each year of test administration adds to the database and gives a more detailed and comprehensive picture of curriculum success.

Content of The Test

The North Carolina Test of Chemistry is objective-referenced; that is, its reference is to a domain of objectives. This domain is mapped over a domain of items, where the items reflect the objectives, equal in kind and number except for random fluctuations. The Chemistry tests were designed to achieve an even assessment across all objectives; in short, each objective was to be represented by the same number of items. This design is consistent with the concept of a domain of objectives mapped over by a domain of items. Although the objectives have unit weighting, the goals are weighted by the number of objectives assigned to them. From empirical analyses, this is a traditional aspect of curriculum development: the more important a goal is considered to be, the greater number of objectives that will be developed for it. Thus, an underlying system of weights exists for the curricular goals.

The Appendix lists each goal and objective and the numerical item representation for each objective as it appears on the 1989 North Carolina Test of Chemistry (Forms A–D). In addition, the proportion of teachers rating each objective as basic to instruction in the Chemistry curriculum is listed.

Tables 4-1 through 8 list the difficulty level for all items on the 1989 North Carolina Test Of Chemistry (Core 1/Forms A–D) and for all items on the four future core tests (Cores 2–5) in terms of p-values (proportion of all students answering the item correctly).

Table 4-1
Item Difficulty by Item Number for the NCT–Chemistry—Core 1

Item #	P-value	Item #	P-value	Item #	P-value
1	.59	21	.79	41	.55
2	.62	22	.71	42	.59
3	.68	23	.72	43	.85
4	.70	24	.74	44	.48
5	.52	25	.61	45	.81
6	.71	26	.61	46	.55
7	.77	27	.65	47	.54
8	.68	28	.63	48	.39
9	.86	29	.57	49	.38
10	.61	30	.75	50	.55
11	.86	31	.57	51	.66
12	.68	32	.36	52	.32
13	.57	33	.34	53	.51
14	.61	34	.39	54	.32
15	.56	35	.47	55	.59
16	.62	36	.42	56	.35
17	.54	37	.46	57	.79
18	.50	38	.32	58	.46
19	.75	39	.31	59	.85
20	.67	40	.67	60	.77

Table 4-2
Item Difficulty by Item Number for the 1989 NCT–Chemistry—Variables A–D

Item #	P-value	Item #	P-value	Item #	P-value	Item #	P-value
A1	.90	B1	.92	C1	.95	D1	.95
A2	.46	B2	.70	C2	.33	D2	.38
A3	.69	B3	.39	C3	.46	D3	.73
A4	.71	B4	.47	C4	.80	D4	.95
A5	.43	B5	.94	C5	.56	D5	.77
A6	.79	B6	.59	C6	.60	D6	.74
A7	.54	B7	.80	C7	.65	D7	.64
A8	.52	B8	.73	C8	.43	D8	.49
A9	.63	B9	.82	C9	.83	D9	.58
A10	.69	B10	.43	C10	.80	D10	.45
A11	.59	B11	.55	C11	.91	D11	.48
A12	.51	B12	.86	C12	.74	D12	.39
A13	.71	B13	.70	C13	.71	D13	.66
A14	.74	B14	.36	C14	.57	D14	.44
A15	.73	B15	.48	C15	.64	D15	.63
A16	.47	B16	.80	C16	.48	D16	.74
A17	.67	B17	.58	C17	.57	D17	.37
A18	.57	B18	.32	C18	.35	D18	.67
A19	.67	B19	.37	C19	.50	D19	.72
A20	.69	B20	.52	C20	.30	D20	.30

Table 4-2 (continued)
Item Difficulty by Item Number for the 1989 NCT–Chemistry—Variables A–D

Item #	P-value	Item #	P-value	Item #	P-value	Item #	P-value
A21	.49	B21	.35	C21	.59	D21	.43
A22	.39	B22	.40	C22	.45	D22	.51
A23	.48	B23	.42	C23	.48	D23	.75
A24	.42	B24	.42	C24	.49	D24	.52
A25	.47	B25	.38	C25	.33	D25	.45
A26	.49	B26	.34	C26	.43	D26	.31
A27	.41	B27	.41	C27	.43	D27	.41
A28	.40	B28	.65	C28	.31	D28	.46
A29	.56	B29	.42	C29	.44	D29	.32
A30	.61	B30	.84	C30	.51	D30	.74
A31	.46	B31	.64	C31	.42	D31	.47
A32	.41	B32	.91	C32	.54	D32	.52
A33	.55	B33	.62	C33	.79	D33	.51
A34	.49	B34	.43	C34	.66	D34	.42
A35	.40	B35	.47	C35	.38	D35	.38
A36	.40	B36	.34	C36	.40	D36	.38
A37	.42	B37	.39	C37	.45	D37	.44
A38	.47	B38	.37	C38	.64	D38	.46
A39	.46	B39	.40	C39	.61	D39	.87
A40	.65	B40	.62	C40	.63	D40	.71

Table 5
Item Difficulty by Item Number for the NCT-Chemistry—Core 2

Item #	P-value	Item #	P-value	Item #	P-value
1	.49	21	.44	41	.35
2	.59	22	.45	42	.72
3	.83	23	.86	43	.40
4	.53	24	.63	44	.36
5	.90	25	.82	45	.66
6	.62	26	.46	46	.70
7	.52	27	.60	47	.39
8	.64	28	.85	48	.50
9	.71	29	.77	49	.32
10	.80	30	.58	50	.55
11	.70	31	.71	51	.42
12	.73	32	.46	52	.49
13	.84	33	.36	53	.41
14	.40	34	.53	54	.34
15	.74	35	.56	55	.59
16	.67	36	.44	56	.53
17	.56	37	.60	57	.82
18	.46	38	.41	58	.51
19	.36	39	.79	59	.85
20	.77	40	.67	60	.58

Table 6
Item Difficulty by Item Number for the NCT–Chemistry—Core 3

Item #	P-value	Item #	P-value	Item #	P-value
1	.45	21	.52	41	.64
2	.66	22	.84	42	.40
3	.44	23	.84	43	[illegible]
4	.81	24	.67	44	.63
5	.53	25	.61	45	.52
6	.56	26	.43	46	.78
7	.57	27	.83	47	.72
8	.57	28	.59	48	.50
9	.44	29	.72	49	.38
10	.64	30	.52	50	.55
11	.79	31	.41	51	.32
12	.72	32	.84	52	.38
13	.84	33	.59	53	.51
14	.53	34	.31	54	.33
15	.66	35	.42	55	.43
16	.53	36	.45	56	.74
17	.77	37	.42	57	.88
18	.53	38	.56	58	.88
19	.68	39	.33	59	.73
20	.66	40	.57	60	.86

Table 7
Item Difficulty by Item Number for the NCT–Chemistry—Core 4

Item #	P-value	Item #	P-value	Item #	P-value
1	.45	21	.70	41	.54
2	.61	22	.71	42	.59
3	.41	23	.75	43	.62
4	.69	24	.87	44	.66
5	.45	25	.40	45	.57
6	.78	26	.74	46	.59
7	.51	27	.77	47	.36
8	.81	28	.56	48	.54
9	.69	29	.44	49	.50
10	.90	30	.40	50	.55
11	.86	31	.59	51	.45
12	.70	32	.40	52	.35
13	.73	33	.62	53	.48
14	.55	34	.38	54	.32
15	.60	35	.41	55	.36
16	.72	36	.56	56	.42
17	.87	37	.65	57	.81
18	.53	38	.30	58	.44
19	.67	39	.84	59	.74
20	.54	40	.67	60	.73

Table 8
Item Difficulty by Item Number for the NCT–Chemistry—Core 5

Item #	P-value	Item #	P-value	Item #	P-value
1	.44	21	.74	41	.36
2	.63	22	.40	42	.45
3	.76	23	.87	43	.55
4	.71	24	.73	44	.35
5	.51	25	.62	45	.39
6	.53	26	.74	46	.91
7	.63	27	.67	47	.75
8	.70	28	.85	48	.59
9	.51	29	.55	49	.47
10	.88	30	.41	50	.55
11	.78	31	.47	51	.39
12	.64	32	.80	52	.68
13	.54	33	.40	53	.48
14	.62	34	.49	54	.63
15	.67	35	.64	55	.50
16	.80	36	.49	56	.37
17	.48	37	.41	57	.59
18	.53	38	.41	58	.69
19	.56	39	.58	59	.64
20	.58	40	.76	60	.56

Test Norms

Students who answer all 60 of the Chemistry core items correctly could be assumed to be excellent Chemistry students. If everyone answered all of the items correctly, however, a different interpretation would have to be placed on the scores. At some point, scores must have a reference group grounded in the experience of all students. In some respect, at least, everything is good or bad by comparison. Norms tables provide that reference. Given a norms table, a student's score can be compared with other students' scores.

Norms tables commonly have two points of reference: a scale of percentiles and a scale of standard scores. The former permits the location of a score within percentile ranks; thus a student is said to have exceeded the performance of 80% of the students in the norm group (in this case, Chemistry students taking the North Carolina Test of Chemistry in May 1989). The latter, standard scores, permits the location of a score within normally-distributed standard scores. This reference is appropriate if the student abilities are believed to be normally distributed. In a normal distribution, raw scores are given greater and greater weight as they diverge from the mean in either direction.

The choice of a metric for the standard score is arbitrary. To avoid inappropriate and confusing comparisons with some of the more common metrics, such as those employed in IQ scores or NCE scores, a metric having a mean of 50 and a standard deviation of 10 was chosen. Most curriculum research studies involving the summation of scores will find the standard score to be the statistic of choice.

The norms table for student scores on the North Carolina Test of Chemistry is given in Table 9. These scores set a baseline of comparison for present and future achievement in Chemistry. Thus, a student score in 1990, 1991, and future years can be referenced to the scores of all 1989 Chemistry students in North Carolina.

In summary, the utility of a test is its statistical equivalence of core tests from year to year, its broad sampling of the curriculum across time, and its initial norms table.

Table 9
Norms for Student Scores on the North Carolina Test of Chemistry

Raw Score	Percentile	Standard Score[a]
60	99	76.8
59	99	75.6
58	99	74.4
57	99	73.2
56	99	72.0
55	99	70.8
54	98	69.6
53	97	68.5
52	96	67.3
51	95	66.1
50	94	64.9
49	92	63.7
48	89	62.5
47	87	61.3
46	84	60.1
45	81	58.9
44	77	57.7
43	73	56.5
42	69	55.4
41	65	54.2
40	60	53.0
39	56	51.8
38	51	50.6
37	47	49.4
36	42	48.2
35	38	47.0
34	34	45.8
33	30	44.6
32	26	43.5
31	22	42.3
30	19	41.1
29	16	39.9
28	13	38.7
27	11	37.5
26	9	36.3
25	8	35.1
24	6	33.9
23	5	32.7
22	4	31.5
21	3	30.4
20	2	29.2
19	2	28.0
Less Than 19	1	

[a]Adjusted to a mean of 50 and a standard deviation 10.0.

Appendix

Test Content—Item Representation by Goal and Objective

Goal/Obj	Description	No. Items 1989	% Teachers Reporting as Basic[a]
Goal 1	The learner will have an understanding of the history, scope, basic concepts, and techniques related to the study of chemistry.		
1.1	Know and apply accepted methods, processes, and procedures for conducting scientific study.	6	97.18
1.2	Know the properties of matter and energy.	7	97.54
1.3	Know the concept of conservation of matter and energy.	6	99.30
Goal 2	The learner will understand systems of classification of matter, nuclear, physical, and chemical changes.		
2.1	Know various descriptive classifications of matter.	7	97.16
2.2	Know the basic chemical concepts of atoms and molecules.	6	99.65
2.3	Know the concepts of elements, compounds, and mixtures.	6	99.30
2.4	Know various types of nuclear changes.	6	44.60
2.5	Know about physical changes such as phase changes and the characteristics of these changes.	6	97.89
2.6	Have knowledge of the nature and evidence of chemical changes.	6	98.59
Goal 3	The learner will have an understanding of descriptive chemistry and periodic properties of elements.		
3.1	Know descriptive terminology pertaining to atomic models and configurations of electrons.	6	95.74
3.2	Know the [illegible]in and nature of the periodic properties or the elements, and the utility of the periodic table.	7	95.74

[a]Percentage of North Carolina Chemistry teachers rating the objective as basic to instruction in Chemistry.

Goal/Obj	Description	No. Items 1989	% Teachers Reporting as Basic[a]
Goal 4	The learner will understand concepts and techniques of measurement and computation as they relate to chemistry.		
4.1	Know how to measure accurately length, area, volume, mass, weight, temperature, and time, and record the measurement as precisely as the measuring devices permit.	7	97.52
4.2	Know how to use scientific notation.	6	97.51
4.3	Have a knowledge of mathematical operations involving manipulation of units and unit conversions.	6	97.87
Goal 5	The learner will have an understanding of stoichiometry and kinetic molecular theory.		
5.1	Know how to construct and use chemical formulas and equations.	7	99.65
5.2	Know how to use the mole concept.	6	96.47
5.3	Know how to make calculations involving stoichiometry given a periodic table and a calculator.	6	95.41
5.4	Know how to make calculations for the prediction of the behavior of gases.	6	84.04
Goal 6	The learner will understand fundamental principles related to chemical reactions, kinetics, and thermodynamics.		
6.1	Know the concept of oxidation-reduction.	5	63.96
6.2	Have knowledge of basic principles in electrochemistry.	7	29.43
6.3	Have knowledge of various energy effects in chemical reactions.	6	57.65
6.4	Know factors that affect the rate of a reaction.	5	74.20
6.5	Know the concept of dynamic equilibrium.	6	55.87

[a]Percentage of North Carolina Chemistry teachers rating the objective as basic to instruction in Chemistry.

Goal/Obj	Description	No. Items 1989	% Teachers Reporting as Basic[a]
Goal 7	The learner will have an understanding of the properties of electrolyte solutions.		
7.1	Know the importance of acids, bases, and salts in industry and in the home.	3	81.21
7.2	Know the names and formulas of selected acids, bases, and salts.	7	96.47
7.3	Know physical characteristics and chemical properties of solutions of acids, bases, and salts.	6	90.11
7.4	Know selected expressions of concentration of solutions.	3	84.10
7.5	Have a knowledge of the phenomenon of ionization.	5	77.94
7.6	Have a knowledge of acid-base equilibria and pH.	6	64.29
7.7	Have a knowledge of solubility equilibria.	7	37.99
Goal 8	The learner will have an understanding of the principles, reactions, and related compounds studied in organic chemistry.		
8.1	Have knowledge of chemical properties, physical forms, and atomic structure of carbon.	6	67.62
8.2	Have knowledge of hybridization and its relationship to bonding and molecular geometry.	5	33.09
8.3	Have knowledge of hydrocarbons.	6	49.82
8.4	Have knowledge of major hydrocarbon substitution products.	6	28.42
8.5	Know that activities of living things involve chemical reactions.	6	51.07
Goal 9	The learner will have an understanding of the relevance of current topics in chemistry.		
9.1	Have knowledge of the relevance of current topics in chemistry.	6	58.57
9.2	Be aware of careers available in chemistry.	7	58.87

[a]Percentage of North Carolina Chemistry teachers rating the objective as basic to instruction in Chemistry.

Appendix 16

END

U.S. Dept. of Education

Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 21,1991