RASCH ITEM BANKING

BENJAMIN D. WRIGHT
SUSAN R. BELL

RESEARCH MEMORANDUM NO. 32
MESA PSYCHOMETRIC LABORATORY
DEPARTMENT OF EDUCATION
UNIVERSITY OF CHICAGO

JULY, 1981

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure page

Chapter I

THE IDEA OF ITEM BANKING

1.1 THE HISTORY OF ITEM POOLS

Asking questions has always been a fundamental part of education. The earliest teachers depended on questions to stimulate dialogue, to bring out what the student did not understand, and to determine what should be done next to promote learning. There is a sense in which the questions and answers in the educating dialogue are the essence of the curriculum.

As education became universal and group instruction replaced tutorial, the role and format of questioning changed. As textbooks became specified at the district level and the curriculum at the state level, the opinion of the individual teacher about what was worth knowing became expressed primarily in the content of his classroom tests. The items which the teacher included in his own tests were the operational definition of his curriculum. They specified what his students should learn and remember. If it wasn't "on the test", it might be interesting or useful but it wasn't part of this teacher's curriculum.

Since the teacher was not able to ask every question of every student personally, it was necessary to use more formal written "tests" so that each student would have the opportunity to respond to every question. This impersonal questioning placed a heavier load on the quality of the questions because the opportunity to clarify a question or redirect a response was gone. Written tests lost the vitality of a true dialogue between teacher and student.

As long as a relatively small number of items were selected by the teacher and administered to a few students in whom the teacher was interested, a thorough analysis of each student's responses to the items was natural, convenient and informative. From the items answered correctly and incorrectly, the teacher learned what the student did or did not know. The right and wrong answers described where the student was in the curriculum and suggested what should be done next.

The teacher, through his experience with his items, knew which items were easy and which hard, and from his association with his students, had a good idea of what each student knew. This permitted informed judgments about the believability of each student's responses. If a student missed "easy" items or passed "hard" ones, the teacher could temper his opinion of the test score. He could discount the unbe-

lievable responses when he formed his assessment of the student's standing. If the surprises related to a particular topic, he could provide remediation for surprising failures or advanced work for surprising successes.

Regardless of how useful classroom tests are to the teacher, to do them well is hard work. To minimize this effort, many teachers have discovered they can accumulate 'item pools' saving their best items for reuse from year to year. This permits more time and energy to be devoted to instruction and to the improvement of their items. It also allows the teacher to make more objective comparisons between current and previous classes.

As item pools proved their utility, their organization became increasingly complex. At its crudest, a pool is no more than a collection of items related to a common theme. But within this broad classification, items may also be arranged by grade level, instructional unit, goal, and objective. As the degree of specification increases, the task of composing a test for a specific occasion becomes clearer and the success of the test more certain.

As tests became the basis for promotions, awards and accountability, they changed their meaning. They became "contests" rather than conversations. To be "fair", a con-

test requires that every student be treated in the same way regardless of where they are in the curriculum. It is more important that everyone have the same opportunity to win than that the questioning provide useful information for educational decisions. As emphasis shifted to the contest, it became more important to develop questions that selected winners than to understand the education of individual students.

The difficulty of developing quality items and the emphasis on tests as contests led to a shift away from locally developed pools to nationally standardized tests. Because there was a large market for quality tests, testing agencies and publishers could devote substantial resources to the development of "good" items and to large scale pilot administrations. The opportunity to sell the same set of items to thousands of students made a large investment in each item economical. The heart of this system is the standardized test form composed of a fixed set of items. Test constructors also found that item pools were convenient resources for the efficient and economical construction of new forms.

Local school districts found nationally standardized tests attractive because they provided an external and seemingly objective standard for comparing performances among

schools and cities. In addition, local educators no longer needed to take full responsibility for the quality of their tests. They could take refuge in the reputations of outside experts.

But in spite of their attractiveness, standardized tests could not replace locally developed tests completely. In order to serve the widest possible market, they reduced the content of the curriculum to a skeleton acceptable everywhere but comprehensive nowhere, focused on a "common" core but insensitive to differences among curricula. Schools that emphasized what standard tests asked appeared more effective on standard tests. The consequence has been a long lamented and much criticized narrowing and thinning of the curriculum.

In order to preserve the "fair contest" aspect of standardized tests, it has been necessary to keep their items as secret as possible. Otherwise teachers who wished their students to do well in the competition, would be tempted to teach the answers to the questions they expected rather than the curriculum as they saw it. And publishers whose standardization depended on a fixed set of items, would have to undertake a new round of expensive standardization studies every time items became public. All this interfered with the possibility of using specific results for detailed diagnosis of individual students.

While seeming to permit national comparisons, the "measures" of standardized tests are basically normative. They are explained by statements like "this student performed better than 60% of the students at his grade level who took the test." But knowing this standing gives no insight into what the student can do. What is needed is some way to specify what it is that the student can do better or worse and some indication of what should be done next to improve his performance rather than his standing.

The psychometric advances of the past twenty years enable the control of testing to be returned to the local level while retaining the useful aspects of standardized tests. Comparability between performances of students or between performances of the same student at different times can be achieved through the identification of the underlying variable common to the items in a pool. This is quite different from the comparability achieved through the nomination of a remote and often quite artificial reference sample performance on a fixed set of items.

When items are related to a common variable, the individual item can become transient. It can represent a position on the variable that could be represented equally well by any number of other items. This makes it possible to describe a student's position with respect to the basic varia-

ble regardless of which items are administered or who else has been tested. The summary information about the student can describe his competence in terms of what it is he has or has not mastered.

Because items can be administered and scored locally and because the need for secrecy disappears, it is once again possible to report and analyze the individual interaction between student and item. The psychometric calibration of the items facilitates this analysis because it enables a quantification of the consistency of each response. This relieves the teacher of the clerical work associated with detailed analyses of test performance and focuses his attention on the particular idiosyncratic responses which are most pertinent to the education of his students.

Underlying this process is the idea that there is one basic issue or topic which all items have in common (eg., reading). The arrangement of the items with respect to this underlying topic gives the variable its definition in terms of range and direction. Items which are identified as "first grade, unit one, objective sixteen", for example, define a definite portion of the curriculum. Observing how item content changes over the range of the pool defines what progress through the curriculum means in terms of competence.

Building item pools is well within the capability of local school districts. Their use should eliminate most of the current criticism of testing. Because psychometric calibration is fundamental to the full utilization of an item pool, we reserve the term "item bank" (Choppin, 1968, 1978) to distinguish a pool based on calibrated items from a pool that is still only a collection.

1.2 A DEFINITION OF ITEM BANK

An item bank is an operational definition of the curriculum. The set of items implies a complete specification of what is important. The idea that the items share is more general than any one item. The items are repeated examples of the basic idea. But because they involve many extraneous factors including the mechanics of administering the test and observing and scoring the response, the items are also imperfect. Understanding the common variable requires identifying what it is that the items share in spite of their inevitable imperfections.

The items in a bank, like those in a pool, are organized to represent the structure of the curriculum. However, this organization is not limited to the subjective judgments of educational experts. Banked items are arranged empirically by the responses of students. This objective definition of the curriculum shows how the students find it rather than how the experts intend it.

Each item represents a smallest element of the curriculum at a specified point on the common variable underlying the bank. Each item illustrates the knowledge, skill, competency or behavior that defines the variable at that point. There can be no more explicit statement about the implications of that point in the curriculum, except through the additional examples of additional items located at the same point.

The calibration that is attached to each item puts the definition of the curriculum on a continuum. The items with low calibrations represent easy, low level tasks, defining one end of this continuum. The items with high calibrations represent difficult, complex tasks define the other end. The progression through the items in the order of their calibrations describes the path that most students take through the curriculum.

Item calibrations are obtained through the application of a probabilistic model for what should happen when a student attempts an item. This allows for give and take between what is expected and what is observed. Students will not all follow exactly the same path and perhaps none will agree completely with what the experts thought. However, assessing the extent of agreement among students and between students and experts is essential for establishing the validity of the items as representatives of the curriculum.

If there is no agreement among the students about which items are hard and easy, there is no common basis for describing progress. It becomes impossible to say that this student is more able than that one or that this student's position on the variable indicates what his next logical task is.

If the empirical ordering surprises the experts, then this is evidence that they do not understand what their items measure. Then, even were it possible to place every student precisely on the variable and to say that this one is more able than that one, we would not know what it is the students are more able at doing.

Fortunately, students do tend to agree with one another and with the experts in the way they use most test items. The process of building an item bank can be exceedingly constructive and educational for the educators involved. Selecting and arranging items promotes a level of communication about what the curriculum contains that does not occur when isolated tests are written by teachers on their own or purchased en masse from publishers. The probabilistic model used to analyze tests and build item banks is not a mysterious black box but a useful frame of reference for describing the curriculum and for identifying deviations between expectation and observation.

Item banks offer many attractive features to educators, such as:

1. Tests can be tailored to the situation in which they are to be used.
2. Items can be chosen to suit the particular level and motivation of the individual student.
3. Individual performance can be monitored for individual item fairness.
4. Scores can be transformed into general measures with explicit competency implications.
5. Progress can be measured longitudinally over a wide range of development.
6. The curriculum can be explicitly described and publicly discussed.

The methodology presented in this report is not magic. Persistence and good judgment are required to implement it. However, when carefully applied, it can expedite an important step toward fair and useful testing.

1.3 ORGANIZATION OF THE REPORT

This report is organized into two parts. The first, Chapters 2 through 4, develops the theory and mechanics of Rasch item banking. The second, Chapters 5 through 8, illustrates the construction of item banks with four typical data sets.

Chapter 2 introduces the basic ideas of Rasch measurement and discusses the importance of specific objectivity. Chapter 3 develops a useful approach to estimation and fit ana-

lysis. Chapter 4 reviews various methods for designing and linking forms.

Chapter 5 explains the computer programs developed to implement item banking and demonstrates them with an example based on the Knox Cube nonverbal intelligence test (Wright and Stone, 1979). Chapters 6 through 8 discuss three item banks that have been constructed by these procedures. These chapters illustrate further the interpretation of the computer output and the process by which educational variables are developed.

These analyses and interpretations are intended to be suggestive and not necessarily definitive or exhaustive. They are not the final word on these data sets. They illustrate useful ways of looking at the results which should help stimulate discussion and hopefully insight by the content specialists who planned the bank and who plan on using it.

Chapter 9 summarizes the work that has been done and suggests directions for additional development.

Chapter II

VARIABLES AND MEASUREMENT

2.1 CONSTRUCTING THE VARIABLE

Most educators are accustomed to using the number of correct answers to test items as the best measure of performance. Although this approach has a long and useful history, it is "test-bound", and this limits its utility. First, all students for which we wish to compare performances must take exactly the same set of items. Second, all items must be given to the same sample if we wish to compare item difficulties. Third, test scores are not linearly related to ability, because of the compression of their scale caused by floor and ceiling effects. To realize their full potential scores must be freed from their test and sample dependencies and stretched at their upper and lower ends.

This is the approach we take. We use the traditional "number correct" to produce "sample-freed item calibrations" and "test-freed person measurements" which are linear in the variable they express. Item difficulties are corrected for the abilities of the persons used to calibrate them. This means that results can be statistically equivalent regard-

less of the distribution of ability in the calibrating sample and hence, need not depend on whose responses are used to calibrate items.

Person abilities are corrected for the difficulties of the items used to obtain them. This means that all students do not have to take the same set of items to obtain results which are directly comparable. If tests are to have uniform meaning, irrespective of the sample of persons who takes them, and if abilities are to be compared whatever the selection of items used, then sample-freed item calibration and test-freed person measurement are necessary conditions.

If the observation of behavior is to be the basis for "measuring" achievement, clarification of that behavior is needed. In order to formulate variables for the assessment of performance, the behavior underlying these variables must be explored in depth. There are four steps needed to make our ideas about ability and the behaviors which reflect it explicit. These steps can lead us to the definition of a measurement scale which has the fundamental properties of sample-freed item calibration and test-freed person measurement.

The <u>first</u> step in defining a variable is to specify the intended content or domain on which measurements are to be

made. "Content validity is dependent on the existence of a clear definition of the pedagogic subject matter."[1] This step is necessary to establish and maintain a consistent definition of the variable. Without careful consideration at this point, there will be no clear guidelines for selecting items and the quality of the bank will always be in question.

The second step is to give substance to the statement of the variable specified by developing test items which elicit behaviors from the persons to be measured that could be indications of their positions on the variable. These items must provide a plausible operational definition of the variable.

A useful representation of this operational definition is to position the items along a line in their expected order of difficulty. This forces careful consideration of what the variable is intended to mean. It also makes possible an assessment of scale validity by enabling a comparison of the "judgmental" ordering with the empirical ordering determined by the administration of the items.

---

[1]Smetherham, D. Banking school knowledge. British Journal of Educational Studies, 1979, 27, 1, 57-68.

Some items will be ordered by the logic of their content. For example, simple multiplication will always be judged more difficult than simple addition because multiplication includes addition. Other items will be ordered by convention or experience. For example, multiplication of fractions is usually judged more difficult than multiplication of decimals, although they might be learned in either order.

We will put off, for a moment, how the final scale is determined, but the empirical ordering for the items in a particular test taken by a particular sample, will correspond to the number of persons who answered the item correctly. If we place the easiest items on the left of the line, then as we move along the line to the right the number of correct answers will decrease as the items become more difficult. The ordering, but not the spacing, will be identical to the arrangement based on number correct.

The relative positions of persons on the line are determined by the number of items they answer correctly. For persons to be positioned along the same line as items, person ability must be expressed in the same units as item difficulty. To illustrate this, Figure 2.1 shows where two persons might fall on a variable defined by seven items.

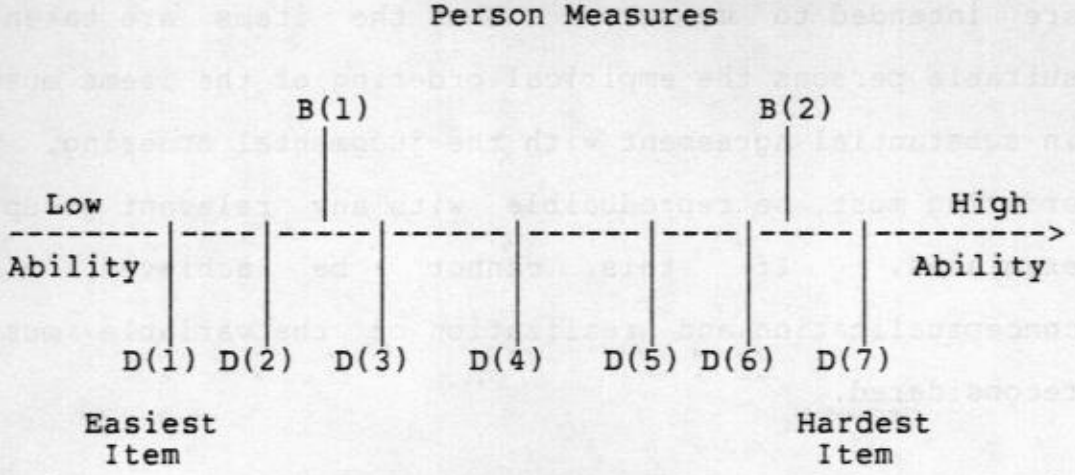


Figure 2.1: The Line of a Variable

The arrow indicates increasing ability, with low ability to the left and high ability to the right. The D's are the calibrations of the items. They are the operational definition of the variable being measured and give the line its meaning. Easy items define the left, or low end of the line and hard items define the right, or high end of the line. The B's mark the persons' locations on the line, and thus represent ability levels defined by items. In this case, person measure B(1) locates this person to the right of the two easiest items and to the left of the five hardest ones, and B(2) positions this person between the two hardest items and well beyond the person with measure B(1).

The third step is to verify that the results are consistent with the original idea of the variable the items

are intended to measure. When the items are taken by suitable persons the empirical ordering of the items must be in substantial agreement with the judgmental ordering. This ordering must be reproducible with any relevant group of examinees. If this cannot be achieved, the conceptualization and realization of the variable must be reconsidered.

The fourth step is to verify that each person's pattern of responses is consistent. A person is expected to succeed on items that are to the left of his position on the variable and hence should be easy for him and to fail on items to the right. Since actual responses are influenced by many factors, the expected pattern cannot be obtained with certainty. Before accepting a person's ability estimate, his response pattern must be examined for its statistical consistency with expectation. When excessive inconsistency is observed, a decision must be made whether to retest the person or to qualify our judgement about his probable measure in some reasonable way.

These four steps are essential to a clear, concise, and objective definition of a variable. The method for building item banks to be described and illustrated succeeds in this in ways which traditional test construction techniques fail to achieve.

2.2 RASCH MEASUREMENT

> The measurement and evaluation of our pupils is a responsibility we must assume with competence and objectivity.[2]

The psychometric method of this study comes from the work of Georg Rasch (1952,1958,1960,1966,1977). It is based on a simple, stochastic model for describing what happens when a person takes an item. This model is unique among latent trait models because of the property Rasch called "specific objectivity." This concept is also described by the dual terms "test-freed" person measurement and "sample-freed" item calibration (Wright, 1968, 1977).

Rasch's simple mathematical model describes the probable outcome of an encounter between person n and item i as governed by two parameters - the "ability" of the person, B, and the "difficulty" of the item, D. The interaction of these two parameters can be represented by their difference (B-D). This difference is used to describe the likelihood of person n with ability B(n) succeeding on item i with difficulty D(i). The difference [B(n)-D(i)] has a range from minus infinity to plus infinity. To bring it between zero and one, it is applied as the exponent of the base e and expressed as the ratio:

---

[2]Taylor, J. S. Improve your classroom testing skills. Clearing House, 1977, 50, 9, 381-385.

$$P(ni) = \exp[B(n) - D(i)]/\{1 + \exp[B(n) - D(i)]\} \qquad [1]$$

This expression describes the ogive shown in Figure 2.2. When the ability of person n is greater than the difficulty of item i, that is, when the difference [B(n)-D(i)] is positive, the probability of success for person n on item i is greater than a half. As person ability increases, this difference increases, and the probability of success becomes closer to one. On the other hand, when item i is harder than person n is able, the difference is negative and the probability of success for person n on item i is less than a half. As item difficuly increases, person n's probability of success approaches zero.

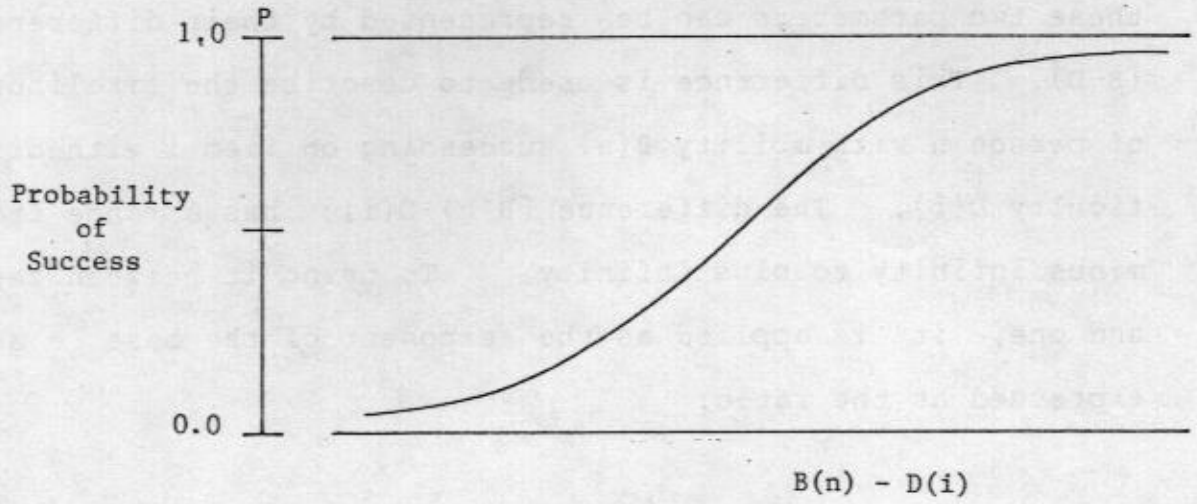


Figure 2.2: The Response Curve

If X=1 is used to represent a correct answer and X=0 to represent an incorrect one, then the formulation of Equation 1 leads to the model probability:

$$P(ni) = \exp\{X[B(n) - D(i)]\}/\{1 + \exp[B(n) - D(i)]\} \qquad [2]$$

This is the Rasch model for dichotomous scoring.

There are three aspects of specific objectivity which distinguish Rasch models from other latent trait models. These properties, order, separability, and sufficiency, are closely related and can be thought of as the substantive, mathematical, and statistical meanings of specific objectivity. In order to appreciate the implications of specific objectivity, it is useful to consider these properties individually.

2.2.1 Order

> It is not very useful to know that Johnny is superior to 84% of his peers unless we know what it is he can do better than they, and just how well he can do it. To be meaningful, test scores must be related to test content as well as to scores of other examinees.[3]

The order property of specific objectivity permits clear comparisons not only between persons and between items, but also between a person and an item. This stochastic ordering

---

[3]Ebel, R. L. Content standard scores. Educational and Psychological Measurement, 1962, 22, 15-25.

can be stated in two propositions:

1. For any item, a more able person always has a greater chance of responding correctly than a less able person;

2. For any person, an easier item always has a greater chance of being answered correctly than a harder item.

These propositions, which define a stochastic Guttman scale, imply that every item and every person can be ordered uniquely along a single line. This ordering of items and persons is necessary to support unambiguous comparisons such as, "Item A is more difficult than Item B" and "Person A is more able than Person B." It is also necessary to support general statements concerning what a person can do in terms of actual items and how well he can do it. These statements of unidimensional relationships are fundamental for measurement.

2.2.2 Separability

The simple logistic Rasch model can be used to seek a substantively meaningful ordering of items and persons, because of a mathematical property: parameter separation. The simple manner in which the parameters enter the model, linearly and without interactions, permits us to develop the

likelihood equations so that the relation between data and the person ability parameter is entirely contained in one factor and the relation between data and the item difficulty parameter entirely contained in another. This allows us to derive conditional estimation equations for either set of parameters such that the equations for estimating item difficulties do not involve the person ability parameters and vice versa.

This parameter separation can be demonstrated by comparing the responses of two persons to any item. The probablity that person n answers item i correctly is given by

$$P(ni) = \exp[B(n) - D(i)]/\{1 + \exp[B(n) - D(i)]\} \quad [3]$$

Hence, his odds for success are

$$P(ni)/[1 - P(ni)] = \exp[B(n) - D(i)] \quad [4]$$

Taking logarithms gives

$$\log\{P(ni)/[1 - P(ni)]\} = B(n) - D(i) \quad [5]$$

Similarly, for person m and item i

$$\log\{P(mi)/[1 - P(mi)]\} = B(m) - D(i) \quad [6]$$

A comparison of person n and person m can then be made by subtracting expression [6] from expression [5]:

$$B(n) - B(m) = \log\{P(ni)[1 - P(mi)]/P(mi)[1 - P(ni)]\} \quad [7]$$

which does not involve the item parameter D(i) at all.

This comparison of person n and person m depends on the use of an appropriate item, but not on any particular item. Both P(ni) and P(mi) are dependent on how difficult item i is, but the separability of parameters allows us to combine them in expression [7] so that D(i) cancels. The possibility of estimation equations for B(n) which are free from the effects of D(i) is referred to as "test-freed" person measurement. An analagous comparison can be made for two item difficulties, to show "sample-freed" item calibration.

2.2.3 Sufficiency

The third property, sufficiency, is a statistical property which follows from parameter separation. The demonstration of the preceeding section depended on a single common item for both persons. In order to use more than one item, we must be able to summarize all of the information about the items in a few observable values. When a model permits such values to be constructed they are called "sufficient statistics."

The sufficient statistics for the simple logistic Rasch model are the number of correct answers for each person and the number of persons who responded correctly to each item. It is a consequence of the Rasch model that these familiar raw scores contain all of the modelled information about the abilities of the persons and the difficulties of the items.

Raw scores are the statistics long used in traditional practice to summarize performances of persons and items. The Rasch model not only justifies this practice, but shows how to use raw scores to solve the many problems which beset traditional psychometrics. This connection with traditional practice and experience provides a validation and consensus lacking in other latent trait models.

Although they are statistically sufficient, raw scores as they stand are not adequate as item calibrations or person measures. They are not linearly related to ability or difficulty and they are test and sample dependent. But because they are sufficient, it is possible to transform raw scores into linear, objective indicators of the person and item positions on a variable. The differences between scores and measures are linearity and generality.

Linearity is achieved through the logistic transformation which expands the score metric at the extremes. An increase

of one logistic unit (logit) represents the same increase of ability and difficulty in any region of the scale. In contrast, one "unit" in a raw score metric can mean a large or small difference in ability depending on where in the test the difference occurs.

Generality is achieved through the specific objectivity, which allows ability measures to be test-freed and difficulty calibrations to be sample-freed. A measure must imply more than just the count of correct answers on a particular set of items. It must provide general quantitative information about how the person can be expected to perform in any related situation.

An item's raw score is specific to the sample used, but its difficulty calibration can be sample-freed. A person's raw score is specific to the test he took, but his ability measure can be general on the variable defined by calibrated items.

2.3 CONCLUSIONS

The special properties of the simple logistic Rasch model give it a unique status among measurement models. They enable the construction of variables which are consistent with common sense and traditional practice. These variables are defined in terms of a unique stochastic ordering of persons

and items along a single line. Objective comparisons can be made between any two elements on the line and the implications of a position can be described in terms of whatever elements are close by. The remainder of this report describes how to construct and use item banks to define educational variables.

Chapter III

CONSTRUCTING AN ITEM BANK

3.1 ESTIMATION

3.1.1 Calibrating Items

The calibration phase of item banking consists of computing "sample-freed" item difficulties from the sample-dependent sufficient statistics. These difficulty estimates then enable the construction of exams for any application and the estimation of "test-freed" measures for every person.

To show that one set of parameters can be freed from the effects of the other, the model probability of Equation 2 (Chapter 2) is used to compute the probability of the data

$$P[((X))] = \prod_{n}^{N} \prod_{i}^{L} \left\{ \frac{\exp\{X(ni)[B(n) - D(i)]\}}{1 + \exp[B(n) - D(i)]} \right\} \qquad [8]$$

where ((X)) is the data matrix of responses X, over all persons n=1,N and over all items i=1,L.

When the continued product operators are moved into the numerator and denominator separately,

$$\prod_{n}^{N} \prod_{i}^{L} \exp\{X(ni)[B(n)-D(i)]\} = \exp\{ \sum_{n}^{N} \sum_{i}^{L} X(ni)[B(n)-D(i)]\}$$

$$\sum_{n}^{N}\sum_{i}^{L} X(ni)B(n) = \sum_{n}^{N} r(n)B(n)$$

$$\sum_{n}^{N}\sum_{i}^{L} X(ni)D(i) = \sum_{i}^{L} s(i)D(i)$$

and Equation 8 becomes

$$P[((X))] = \frac{\exp\left[\sum_{n}^{N} r(n)B(n) - \sum_{i}^{L} s(i)D(i)\right]}{\prod_{n}^{N}\prod_{i}^{L}\{1 + \exp[B(n) - D(i)]\}} \qquad [9]$$

It can be seen from Equation 9 that only the person scores r(n) and the item scores s(i) are needed in order to estimate B(n) and D(i).

Factoring the numerator of Equation 9 into two parts, such that

$$P[((X))] = \frac{\{\exp\left[\sum_{n}^{N} r(n)B(n)\right]\}/\{\exp\left[\sum_{i}^{L} s(i)D(i)\right]\}}{\prod_{n}^{N}\prod_{i}^{L}\{1 + \exp[B(n) - D(i)]\}} \qquad [10]$$

makes the separation of

$$\sum_{n}^{N} r(n)B(n)$$

and

$$\sum_{i}^{L} s(i)D(i)$$

clear. Either set of parameters can be conditioned out of Equation 10 when estimating the other set.

There are several ways estimates can be computed for these parameters, two will be described in this section. The first is the unconditional maximum likelihood procedure, UCON, developed by Wright and Panchapakesan (1966, 1969). The second is Cohen's approximation method, PROX (Wright and Douglas, 1977; Wright, 1977).

3.1.1.1 Unconditional Maximum Likelihood Estimation (UCON) The UCON estimation procedure is "unconditional" because the item difficulties and person abilities are estimated simultaneously rather than after "conditioning" out one set of parameters while estimating the other. It is a maximum likelihood procedure because it uses as the estimates of the parameters those values which maximize the likelihood of the data.

The estimation equations are derived by differentiating the likelihood expression in Equation 10 with respect to each parameter, setting the derivatives equal to zero and

solving for the unknown parameters. Because the equations cannot be solved explicitly, a numerical method, such as the Newton-Raphson procedure, must be used. An algorithm for the unconditional procedure is described below.

The estimation equation for item difficulty is

$$s(i) = \sum_{n}^{N} P(ni) \qquad [11]$$

where s(i) is the total number of successes by all persons on item i, and

$$P(ni) = \exp[B(n) - D(i)]/\{1 + \exp[B(n) - D(i)]\}$$

is the probability of success by person n on item i.

The estimation equation for person ability is

$$r(n) = \sum_{i}^{L} P(ni) \qquad [12]$$

where r(n) is the total number of successes on all items by person n.

Because raw score is the only information used to describe a person's performances, ability estimates will be the same for everyone with a given score, regardless of which

items were correct. All persons with identical raw scores will receive identical estimates of ability. For estimation purposes, this allows persons to be indexed by their score, so that the number of persons at each raw score become the sufficient statistics for persons. Equation 11 now becomes

$$s(i) = \sum_{r=1}^{L-1} N(r)p(ri) \qquad [13]$$

where L is the largest possible score, N(r) is the number of persons with score r and

$$p(ri) = \exp[b(r) - d(i)]/\{1 + \exp[b(r) - d(i)]\}$$

where b(r) is the ability estimate for any person with a score r and d(i) is the difficulty estimate of item i.

Equation 12 then becomes

$$r = \sum_{i=1}^{L} p(ri) \qquad [14]$$

Once item scores, s(i), and counts of the number of persons at each score, N(r), have been obtained, all items either no one gets right, s(i)=0, or no one gets wrong, s(i)=N, and all persons with either no correct responses, r=0, or no incorrect responses, r=L, must be deleted. This

is because finite estimates do not exist for these situations. This editing process is recycled until all zero and perfect scores are eliminated.

An initial set of [d(i)] is required to begin estimation. These can be

$$d(i) = \log\left[\frac{N - s(i)}{s(i)}\right] \qquad i=1,L \qquad [15]$$

where d(i) is the maximum likelihood estimate for D(i) if all B(r)=0.

An initial set of [b(r)] can be

$$b(r) = \log\left[\frac{r}{L - r}\right] \qquad r=1,L-1 \qquad [16]$$

where b(r) is the maximum likelihood estimate of B(r) if all D(i)=0.

The item difficulties [d(i)] are centered at zero by subtracting $d. = \Sigma d(i)/L$ from each item so that d(i) becomes redefined as

$$d(i) = d(i) - d. \qquad [17]$$

The Newton-Raphson method is then applied to improve each d(i) according to

$$d(i) = d(i) - \frac{s(i) - \sum_r N(r)p(ri)}{\sum_r N(r)p(ri)[1 - p(ri)]} \qquad [18]$$

where the current set of [b(r)] is given by the previous cycle. Equation 18 is iterated until convergence is achieved.

The improved set of [d(i)] is recentered at d.=0 and the Newton-Raphson method applied again to improve each b(r) according to

$$b(r) = b(r) + \frac{r - \sum_i p(ri)}{\sum_i p(ri)\{1 - p(ri)\}} \qquad [19]$$

The last three steps are repeated until stable values for the [d(i)] are obtained.

The slight bias in this unconditional maximum likelihood procedure can be averaged out of the centered d(i) by multiplying them by (L-1)/L. The final set of [b(r)] is calculated using the unbiased [d(i)] and the resulting [b(r)] are rescaled by the same factor.

Asymptotic estimates of the standard errors of item difficulty estimates are

$$SEC(i) = \{ \sum_{r} N(r)p(ri)[1 - p(ri)]\}^{**}(-1/2) \qquad [20]$$

and of person ability estimates are

$$SEM(r) = \{ \sum_{i}^{L} p(ri)[1 - p(ri)]\}^{**}(-1/2) \qquad [21]$$

3.1.1.2 Cohen's Normal Approximation (PROX)

PROX is a simple economical alternative for estimating model parameters which assumes that the distribution of person abilities [B(n)] and item difficulties [D(i)] can be described adequately by their first two moments. This method can be used when the calibrating sample does not have an asymmetrical distribution of ability and when it is not too far off target for the test.

Initial estimates of difficulties [d(i)] and abilities [b(r)] are defined by Equations [15] and [16].

The item set is then centered at zero by subtracting $d. = \Sigma d(i)/L$ (Equation [17]). But now, the difficulty and ability estimates are computed by

$$d(i) = Y*d(i) \qquad i=1,L \qquad [22]$$

and

$$b(r) = X*b(r) \qquad r=1,L-1 \qquad [23]$$

where

$$Y = [(1 + U)/(1 - U*V)]**1/2 \qquad [24]$$

$$X = [(1 + V)/(1 - U*V)]**1/2 \qquad [25]$$

$$U = \sum_{r}^{L-1} N(r)[b(r) - b.]^2/[2.89(N-1)] \qquad [26]$$

$$V = \sum_{i}^{L} [d(i)]^2/[2.89(L-1)] \qquad [27]$$

The value $2.89 = 1.7^2$ arises from the relationship between the normal and logistic cumulative distributions.

PROX assumes that the model parameters can be approximated by an explicit function of total score and a single scaling factor. The item difficulties are corrected for the mean and standard deviation of sample ability and the person abilities are corrected for the mean and standard deviation of test item difficulty. The standard error of calibration is

$$SEC(i) = Y[N/\{s(i)[N - s(i)]\}]**1/2 \qquad [28]$$

and the standard error of measurement is

$$SEM(r) = X\{L/[r(1 - r)]\}^{**}1/2 \qquad [29]$$

3.1.2 Calibrating Forms

Once difficulties have been estimated for all items within the forms in which they were used, the next step is to determine the relative difficulty of each form. These "form difficulties" will permit us to place all items in all forms on a common origin. A weighted average of differences in difficulty, called the observed shift, is calculated for the items linking forms k and j by

$$t(kj) = \frac{\sum_{i}^{n} [d(ik) - d(ij)]/w(ikj)}{\sum_{i}^{n} [1/w(ikj)]} \qquad [30]$$

where d(ik) and d(ij) are the estimated difficulties of the linking item in forms k and j, n is the number of items in the link, and where $w(ikj) = [se(ik)^2 + se(ij)^2]$ is a weight based on the calibration standard errors se(ik) and se(ij). The value t(kj) is the estimated difference in the origins of the two forms.

The standard error associated with this observed shift is

$$se(kj) = \frac{\{\sum_{i}^{n} [w(ikj)]/n\}^{**1/2}}{n} \qquad [31]$$

An observed shift is calculated for every pair of forms which are linked by common items.

The translation constant, which is the position of the form relative to all other forms, is computed as the average of the observed shifts for that form,

$$T(k) = \frac{\sum_{j}^{M} t(kj)}{M} \qquad [32]$$

where M is the number of forms to which form k is linked. The translation standard error is

$$se(k) = \frac{\{\sum_{j}^{M} [se(ikj)]^2\}^{**1/2}}{M} \qquad [33]$$

Equations 30 through 33 assume that every form is linked to every other form. But it is not necessary that the matrix of observed shifts be complete in order to estimate the translation constants. The skew symmetry of the shift matrix, in which t(kj) = -t(jk), makes estimation manageable even when many cells are empty.

To estimate translation constants when there are missing links between forms:

Start the empty cells at

$$t(kj) = 0 \qquad [34]$$

and calculate the translation constants

$$T(k) = \frac{\sum_{j}^{M} t(kj)}{M} \qquad [35]$$

where M is the number of forms.

Now, reestimate the empty cells

$$t(kj) = T(j) - T(k) \qquad [36]$$

and repeat steps 35 and 36 until the T(k) stabilize. The sum of the translation constants over all forms will be zero by this process. This sets the bank origin at the center of the forms. The T(k) give the form difficulties relative to this origin.

3.2 ANALYSIS OF FIT

The estimation of the model parameters, item difficulties, observed shifts, and form difficulties, is based on all available data, and the expectation that the data can approximate a simple logistic Rasch model. If this approximation is reasonable, then the estimates of the item difficulties are sample-freed. The practical criterion for this is that subdivisions of the data, such as ability group, grade level, or sex, are observed to produce statistically equivalent estimates.

Since data may not approximate any model, analyses of fit must be done to ascertain whether or not a useful degree of objectivity can be achieved with the data at hand. When dealing with the calibration of a single form, a rather loose approximation to objectivity is often adequate for all practical purposes. However, when constructing an item bank with a large number of items covering a wide range of difficulty, a stricter consistency with objectivity is usually required. In order to support "test-freed" person measurement, the bank must be constructed so that the item difficulties are relatively invariant over the kinds of persons to be measured.

3.2.1 Item Within Form Fit Analysis

The first step in checking whether or not the estimates of item difficulty are approximately sample-free is done when calibrating each form. If the estimates are free of the abilities of the calibrating sample, subdivisions of the sample, by ability group, sex, race, etc. will give statistically equivalent item difficulty estimates.

One way to check this is by dividing the sample into subgroups based on raw score (the sufficient statistic for ability), and comparing the observed number succeeding on each item in each ability subgroup with the number predicted for that subgroup. If overall estimates of the parameters are adequate for describing the data for group g, then the observed number correct in group g will be near the model estimated expectation

$$R(gi) = \sum_{r<g} N(r)p(ri) \qquad [37]$$

and estimated variance

$$S(gi)^2 = \sum_{r<g} N(r)p(ri)[1 - p(ri)] \qquad [38]$$

where N(r) is the number of persons who achieved the score of r, and p(ri) is the estimated probability of success on a score of r for item i. If there is agreement between

observed and expected, then the subgroups concur on the estimated difficulty, and the confidence to be placed in this estimate is reflected in its standard error. Similar analyses can be done for subgroups defined in any other manner.

A second, more general fit statistic to check within form item fit evaluates the agreement between the variable defined by the item and the variable defined by all the other items over the sample. This statistic is formed from a weighted mean square, in which each standardized squared residual, $z(ni)^2 = (x-p)^2/[p(1-p)]$, is weighted by its information on the item, $w(ni) = p(1-p)$:

$$V(i) = \frac{\sum_{n}^{N} [z(ni)^2 w(ni)]}{\sum_{n}^{N} w(ni)} \qquad [39]$$

When items are identified as misfitting, reasonable explanations for their nonconformity can usually be found. Typically, items misfit because of mechanical failures, such as miskeying, misprinting, no right answer and more than one right answer. When the problem with a misfitting item is not mechanical, the explanation can often be traced to requirements for special knowledge, such as knowing that

multiplication by zero is different than multiplication by any other number.

When misfitting items with special problems are found, correcting the problem, such as fixing the key, or eliminating the item with flaws that cannot be fixed, is a reasonable course of action which is defensible and reproducible. When no clear explanation can be found for a misfitting item however, it might be set aside for further consideration or, if there are curricular reasons for doing so, left in the bank with a "cautionary" note. A few such items can be absorbed by a bank without spoiling its quality of measurement.

3.2.2 Item Within Link Fit Analysis

The fit statistics just described are used to evaluate the suitability of the item difficulty estimates. Once these estimates are deemed satisfactory they can be used to link the forms that have common items. Because the items which comprise the links serve a special function in the construction of the bank, they undergo additional investigation.

First, we verify that the linking items performed adequately within their own forms. This is done by checking the general item within form fit statistic computed for each item in its own form,

$$\text{Within form fit} = \frac{\sum_{i}^{n} \left\{ \frac{[V(ik)^2 + V(ij)^2]}{2} \right\}}{n} \qquad [40]$$

where V(ik) is the fit of item i in form k, and n is the number of items in the link. This fit statistic is computed for every link. It will be near one if all items in the link had satisfactory fit within their own forms. This checks explicitly whether persons who took the same set of items have agreed on the item difficulties.

3.2.3 Item Between Link Fit Analysis

The next level of link fit analysis is used to check the extent to which the items in the link agree on the relative difficulties of the two forms involved. This fit statistic can be calculated from the ratio of the observed variance to the modelled variance. The observed variance is

$$S^2 = \frac{\sum_{i}^{n} \left\{ \frac{[d(ik) - d(ij)]^2}{w(ikj)} \right\}}{\sum_{i}^{n} [1/w(ikj)]} - \frac{\sum_{i}^{n} \left\{ \frac{d(ik) - d(ij)}{w(ikj)} \right\}^2}{\sum_{i}^{n} [1/w(ikj)]^2} \qquad [41]$$

where the d(ik) have been translated to the bank, such that

$$d(ik) = \hat{d}(ik) + T(k)$$

and where $w(ikj) = [se(ik)^2 + se(ij)^2]$ as before.

The modelled variance is

$$V = \frac{\sum_{i}^{n} [w(ikj)]}{n} \qquad [42]$$

As this fit statistic is based on the differences in difficulty of the items in the link, misfit can be attributed to items that operated differently in the two forms, and as a result have different observed difficulties. For example, different items may have inadvertantly been given the same identification number or an item that was intended to be the same in both forms may have been misprinted or reworded.

Item format can also influence misfit. Arithmetic problems written horizontally are more difficult than the same problems written vertically for some age groups. This variation in format may induce misfit in a link that includes this type of item. Although the items appear the same, they are, in fact, different as far as persons at different ability levels.

3.2.4 Link Within Bank Fit Analysis

It is also useful to assess the extent of agreement among link shifts with respect to relative form difficulty. Each entry in the matrix of observed link shifts should be predictable from the margins. This produces a matrix of

expected link shifts. To evaluate whether a link fits the bank, a residual can be calculated by

$$y(kj) = t(kj) - [T(j) - T(k)] \qquad [43]$$

where t(kj) is the observed shift between forms k and j and T(k) and T(j) are the translation constants (or relative difficulties) of forms k and j.

These link residuals can be standardized by dividing them by the standard errors, se(kj), of their corresponding t(kj) and multiplying them by [M/(M-2)]**1/2 to form a standardized link residual. When the link fits the bank these standardized link residuals approximate a standard normal distribution.

3.2.5 Form Within Bank Fit Analysis

The residuals for each of the observed links can also be used to evaluate the extent to which a form fits to the bank. The standardized residuals for each of the observed links can be used to compute a fit statistic for each form.

$$V(k) = \frac{\sum_{j}^{M} [y(kj)/se(kj)]^2}{M - 1} \qquad [44]$$

The fit of a link or form into the bank is related to how well the linking items fit within their own forms. Careful investigation of doubtful items is always instructive and invariably leads to new insights about the nature of the variable. Sometimes the misfit of links within the bank is associated with particular forms. This can occur, for example, when a form is administered to a group for which it is inappropriate.

The procedures presented in this chapter are extensions of standard estimation and fit analyses. No matter how careful we do these steps, the collection of data is the first and most important step. The following chapter presents a new methodology for collecting data, and an example of how some of the procedures in this chapter and the next can be done by hand.

Chapter IV

METHODOLOGY FOR ITEM BANKING

4.1 DESIGN

The motivation for building an item bank is to obtain convenient and useful measures of individuals. In order to make repeated yet independent measures over a range of development, items are needed which cover a wider range of difficulty than any one person can handle at a single sitting. To be useful for measurement these items must be calibrated together onto the common variable they are supposed to define. Once a range of items has been satisfactorily calibrated, new tests designed to be fair and useful are easily constructed. Any new combination of items taken from the bank connected through their common item calibrations is automatically equated to any other tests which might be formed from the same bank.

4.1.1 Sources of Data

The usual data used for item analysis and banking is collected through extracurricular testing. This is thought apt because it allows the use of published tests which have been subjected to pilot testing and editing. The items in these tests, however, are frequently not directly relevant to the

local curriculum. Because the testing is separate from the instruction, teachers and students can be quite uninterested in the testing activity, raising further doubts about the relevance of the results.

More meaningful data would be available if testing and banking were based on locally developed items specifically relevant to the local curriculum and administered as part of the ongoing curriculum. The banking procedures we describe can be used with data collected in the ordinary course of instruction through classroom tests administered to assess student progress in the natural course of instruction.

4.1.2 Connecting Tests

All that is necessary to construct an item bank from classroom data is to have successive tests contain some items from previous tests. These linking items provide the reference points for estimating the relationships between tests.

A key feature of this kind of testing is the network of shared items. The network design must take account of the number of items to be calibrated, the number of items a student can be expected to take, and the difficulty range that must be covered. The design must provide sufficient information to place the test forms on a common origin. In principle, this requires only that each new form share some

items with a previous form. Maintaining control over the process however, is facilitated by adding redundancy to the design.

Any design, no matter how elaborate, is composed of basic elements. The fundamental building block for creating a bank is a LINK of common items shared between a pair of tests, as shown in Figure 4.1.

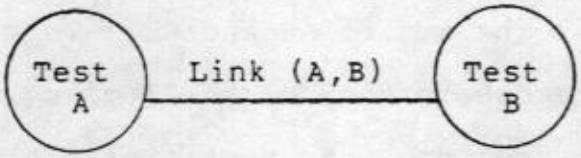


Figure 4.1: A Link Between Two Tests

Each circle represents a test form that is short enough for one person to take at one time and narrow enough in difficulty range for a suitably chosen sample of persons. The line connecting the circles represents a subset of linking items shared by the two tests. All banking designs are built out of these links.

A CHAIN, as shown in Figure 4.2, is the simplest series of links. At first, it might appear to be an efficient arrangement for covering a broad range of difficulty, but a chain is deficient for controlling the quality of the bank.

Figure 4.2: A Chain of Two Links and Three Tests

A LOOP, Figure 4.3, of three links arises when items from test A are also included in test C. It does not cover as wide a range of difficulty or cover as many items as a chain, but it provides two ways to estimate the AB link. The AB link can be estimated directly and it can be estimated as the sum of AC and CB. This redundancy allows a check on the consistency of the loop's links, thus making it a stronger structure than a pair of links.

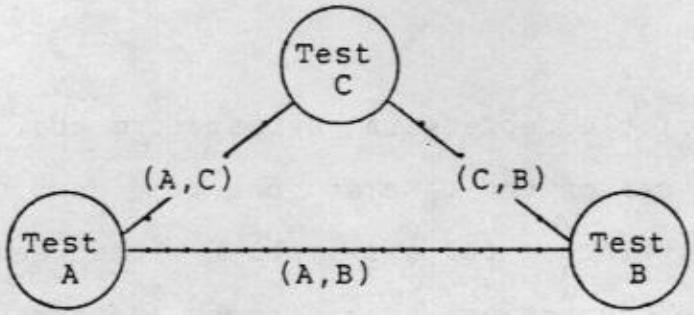


Figure 4.3: A Loop of Three Links and Three Tests

A NET of connected loops, Figure 4.4, can support the common calibration of a wide range of items. This example uses seven tests, twelve links, and five levels of difficulty, but the design can be extended in either direction. The number of items that could be calibrated

with this net depends on the number that can be included in one form and the number that are included in each link.

The maximum test lengths that can be used, of course, depend on subject matter, item type, and the age of the persons to be tested. The number of items in each link is determined by the number of test forms used, the length of the forms, the difficulty range that must be covered, and the number of items that need to be calibrated. These relationships are given by:

$N = M*L - m*n$

where M = the number of forms

L = the number of items per form

m = the number of links per form

n = the number of items per link

and N = the total number of items

A special kind of net in which each form is linked to as many other forms as possible is a WEB. A web is the best general structure for bank building because it maximizes the number of links among test forms, thus controlling the quality of the bank most effectively. Figure 4.5 illustrates the situation where every form is linked to every other form.

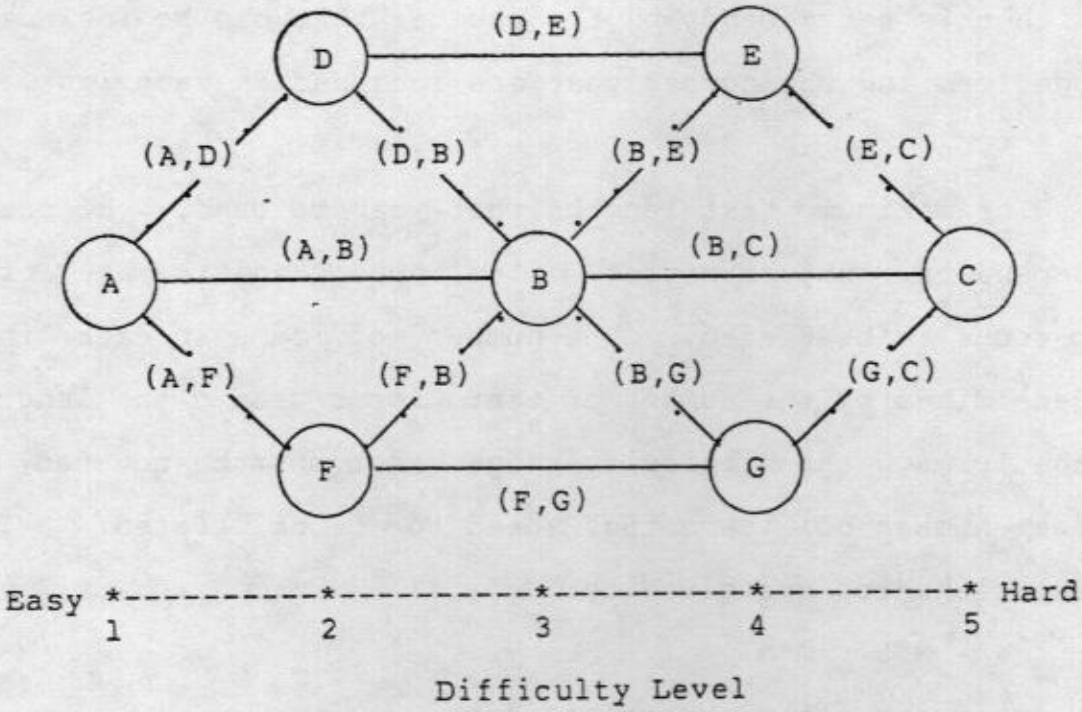


Figure 4.4: A Net of Twelve Links and Seven Tests

If the total number of items to go into a bank needs to be increased, but the number of items in each form must remain the same, an incomplete web can be used. This type of web, Figure 4.6, increases the number of forms while keeping the number of items in each form constant.

In order for these webs to be used, the items must be sufficiently similar in difficulty that some items in Form A can be included in Form N. These webs will not work when growth over several years needs to be measured. Longitudinal studies are better served by webs which cover a wide range of difficulty. Figure 4.7 shows how such a web

FORMS	A	B	C	D	E	F	G	H	I	J	K
A		1	2	3	4	5	6	7	8	9	10
B			11	12	13	14	15	16	17	18	19
C				20	21	22	23	24	25	26	27
D					28	29	30	31	32	33	34
E						35	36	37	38	39	40
F							41	42	43	44	45
G								46	47	48	49
H									50	51	52
I										53	54
J											55
K											

Figure 4.5: A Complete Web of 11 Forms Connected by 55 Links

FORMS	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A		1	2	3	4	5				6	7	8	9	10
B			11	12	13	14	15				16	17	18	19
C				20	21	22	23	24				25	26	27
D					28	29	30	31	32				33	34
E						35	36	37	38	39				40
F							41	42	43	44	45			
G								46	47	48	49	50		
H									51	52	53	54	55	
I										56	57	58	59	60
J											61	62	63	64
K												65	66	67
L													68	69
M														70
N														

Figure 4.6: An Incomplete Web of 14 Forms Connected by 70 Links

can be constructed. In this web, only forms near each other in difficulty level are linked together directly. This allows the set of test forms to cover a wide range of difficulty. The design in Figure 4.7 is an extension of the net shown in Figure 4.4.

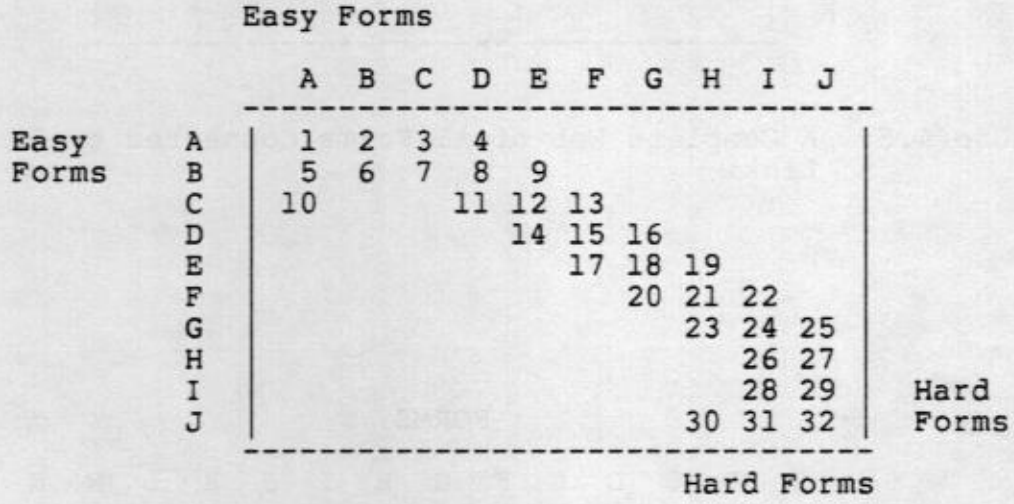


Figure 4.7: A Sequential Web of 10 Forms Connected by 32 Links

The design of webs must take four factors into account:

1. the total number of items to go into the bank;

2. the maximum number of items acceptable in one form;

3. the intended difficulty range of the bank; and,

4. the acceptable difficulty range of a form.

The total number of items that can be banked by one set of forms depends on the type of web used, the number of forms, the number of links per form, and the number of items in each link.

In Figures 4.5, 4.6, and 4.7, the numbers entered in the cells are a serial identification of the links connecting the pair of forms which define the position of that cell. If there is one item per link, the number of links in a web is the same as the number of items banked. When more than one item is allocated to each link, the size of the bank is increased by that multiplier.

For any web, the number of items per form, L, is n*m, where n is the number of items per link and m is the number of links per form. For a complete web, m=(M-1), since each form is connected to every form but itself. The number of items banked in a complete web is given by:

$$N = (M*L)/2$$

In Figure 4.5, the 11 forms produce (11*1*10)/2=55 items, when there is only one item per link. If the size of each link in Figure 4.5 were 4, then there would be (11*4*10)/2=220 items in the bank.

The number of items in an incomplete web is calculated with the same formula, but allowing L to be less restricted than n*(M-1). In Figure 4.6, if n=1, N=(14*1*10)/2=70 items in the bank. The number of items per form remained the same as in the complete web, but the size of the bank increased from 55 to 70 items because the number of forms increased from 11 to 14. If the number of items in each link were 4, the number of items would increase from 220 in the complete web to 280 in the incomplete web.

The formula to calculate the number of items in a sequential web is:

$$N = (M*L/2) + C$$

where C=L/4 if L/2 is even, otherwise C=(L+2)/4. In Figure 4.7, if n=1, N=[(10*6)/2]+2=32 items in the bank.

4.2 EXAMPLE

To illustrate how items are used to link and calibrate forms, a small problem will be done by hand. The data are responses to 41 Knox Cube Test items. These data are also used in Chapter 6 to illustrate the item banking computer programs.

Once a suitable set of items is written, the next step in the development of an item bank is to allocate the items among test forms. The 41 Knox Cube Test items were spread

among five forms so that every form was linked to every other form. This structure maximizes the number of connections. The web for this problem is shown in Figure 4.8. The 'x' in each cell indicates a link between two forms.

		Forms				
		KCTA	KCTB	KCTC	KCTD	KCTE
	KCTA		x	x	x	x
	KCTB			x	x	x
Forms	KCTC				x	x
	KCTD					x
	KCTE					

Figure 4.8: A Complete Web for 5 KCT Forms

As Figure 4.8 shows the upper triangular part of the matrix is filled by 10 links, the maximum number possible between five forms.

It is not necessary that the number of items making up each link be the same and in the KCT example they are not. Figure 4.9 shows the complete web again with numbers entered in the cells to indicate how many items each link contains.

After the forms are administered, the item difficulties are computed using one of the estimation procedures, UCON or PROX, described in Chapter 3.[4] The next step is the

---

[4]For an example of PROX done by hand see: Wright, B. D. and Stone, M. H. Best Test Design. Chicago: MESA Press, 1979.

		Forms			
		KCTA	KCTB	KCTC	KCTD	KCTE
	KCTA		14	13	14	11
	KCTB			20	26	18
Forms	KCTC				28	16
	KCTD					25
	KCTE					

Figure 4.9: Number of Items/Link in KCT

calibration of the forms.

The observed shift is computed between each pair of forms which are connected by one or more items. Table 4.1 shows the steps for computing the observed shift from form KCTD to form KCTE. Forms KCTD and KCTE are linked by the 25 items in this table. The first five columns in Table 4.1 list for each linking item its identification number and its difficulty and standard error in each form, estimated by the UCON procedure. The last four columns show the steps needed to compute the observed shift and its standard error: the differences in the difficulties, $D(d)-D(e)$; its variance, $w = se^2(d)+se^2(e)$; the inverse variance weight, $\{1/[se^2(d)+se^2(e)]\}$; and the weighted element, $\{[D(d)-D(e)]/[se^2(d)+se^2(e)]\}$. The observed shift and its standard error are computed at the bottom of Table 4.1, using Equations 32 and 33 from Chapter 3.

TABLE 4.1

SHIFT By Hand

Item Num	Difficulty KCTD	Difficulty KCTE	s.e. KCTD	s.e. KCTE	Dd-De	w	1/w	Dd-De/w
1	-8.02	-8.15	1.36	1.24	0.13	3.39	0.30	0.04
3	-7.14	-7.40	1.15	0.81	0.26	1.98	0.51	0.13
4	-3.96	-4.80	0.39	0.35	0.84	0.28	3.64	3.06
5	-5.40	-6.30	0.55	0.59	0.90	0.65	1.54	1.38
6	-3.31	-3.03	0.34	0.22	-0.28	0.16	6.10	-1.71
7	-2.23	-2.87	0.28	0.20	0.64	0.12	8.45	5.41
10	-3.09	-2.95	0.37	0.20	-0.14	0.18	5.65	-0.79
11	-1.88	-2.28	0.26	0.18	0.40	0.10	10.00	4.00
12	-1.23	-1.64	0.24	0.17	0.41	0.09	11.56	4.74
14	-1.39	-0.84	0.27	0.15	-0.55	0.10	10.48	-5.77
20	1.20	0.68	0.22	0.16	0.52	0.07	13.51	7.03
21	2.00	0.66	0.23	0.16	1.34	0.08	12.74	17.07
24	2.11	0.96	0.24	0.15	1.15	0.08	12.48	14.36
25	0.23	-0.61	0.22	0.15	0.84	0.07	14.10	11.85
26	-1.07	-1.87	0.26	0.17	0.80	0.10	10.36	8.29
28	2.75	1.83	0.28	0.17	0.92	0.11	9.32	8.57
29	2.95	2.19	0.26	0.18	0.76	0.10	10.00	7.60
30	4.73	3.30	0.38	0.24	1.43	0.20	4.95	7.08
31	3.02	2.07	0.27	0.18	0.95	0.11	9.50	9.02
35	4.23	2.95	0.37	0.22	1.28	0.19	5.40	6.91
36	5.42	4.72	0.46	0.38	0.70	0.36	2.81	1.97
37	5.42	4.15	0.47	0.31	1.27	0.32	3.16	4.01
38	4.59	4.46	0.37	0.37	0.13	0.27	3.65	0.48
39	5.91	4.87	0.54	0.41	1.04	0.46	2.18	2.26
40	5.64	4.87	0.54	0.41	0.77	0.46	2.18	1.68
						10.03	174.57	118.67

where $w = se^2(d) + se^2(e)$

Observed shift = 118.67/174.57 = 0.68

Shift standard error = [(10.03/25)**1/2)/25] = .03

Table 4.2 shows the matrix of observed shifts for the KCT example. The observed shift for form KCTA to KCTD is -0.21; from form KCTB to KCTD, 0.12; from form KCTC to KCTD, 0.77;

from form KCTD to itself, 0.0; and from form KCTE to KCTD, -0.68. These observed shifts produce a translation constant for form KCTD of 0.00, with a standard error of 0.02.

TABLE 4.2

Matrix of Observed Shifts

Forms	KCTA	KCTB	KCTC	KCTD	KCTE	Sum	Mean	-T
KCTA	0.0	0.01	-0.23	-0.21	0.44	0.01	0.002	0.00
KCTB	-0.01	0.0	-0.37	0.12	0.49	0.23	0.046	0.04
KCTC	0.23	0.37	0.0	0.77	1.03	2.40	0.480	0.48
KCTD	0.21	-0.12	-0.77	0.0	0.68	0.00	0.000	0.00
KCTE	-0.44	-0.49	-1.03	-0.68	0.0	-2.64	-0.528	-0.53
Sum	-0.01	-0.23	-2.40	-0.00	2.64	Gr. Mn. 0.000		
T	-0.00	-0.04	-0.48	-0.00	0.53			

In Table 4.2, the numbers in the column labelled "Mean" are the averages of the observed shifts and the numbers in the column labelled "T" are the translation constants actually computed by the SHIFT program. The slight difference between these columns is due to rounding error.

Now that we have an initial frame of reference, we must analyze each item's various difficulties for consistency, so that items with unreasonably large bank differences can be identified and deleted from the linking structure. This protects the calibration of forms from the impact of items

which are discovered to behave differently in different forms.

Table 4.3 reviews the 25 items which link forms KCTD and KCTE. The difficulty of each item has been modified by adding the form difficulties, 0.00 to KCTD items and 0.53 to KCTE items. The residual differences between these "bank" difficulties are listed in the last column. All items with residuals greater than 0.70 were eliminated leaving the nineteen non-asterisked items in the edited link between forms KCTD and KCTE. This edited link produced a new observed shift of 0.69 (in contrast to the previous 0.68) with standard error 0.04.

The next step is to add the translation constant of each form to every item within that form, so that all items become located on the one common bank scale.

The last step is to average each item's various bank difficulties (i.e., with the translation constants added), to produce one bank difficulty for each item. Although an item may be eliminated from a link because of a large residual, it still may be useful to average its various values to produce one bank difficulty for this item. If, however, it can be determined that the problem is due to one form, it may be more reasonable to exclude the flawed form from any averaging.

TABLE 4.3

List of Linking Items with Residuals

Item Num	Difficulty KCTD	KCTE	Residual	
1	-8.02	-7.62	-0.40	
3	-7.14	-6.87	-0.27	
4	-3.96	-4.27	0.31	
5	-5.40	-5.77	0.37	
6	-3.31	-2.50	0.81	*
7	-2.23	-2.34	0.11	
10	-3.09	-2.42	-0.67	
11	-1.88	-1.75	-0.13	
12	-1.23	-1.11	-0.12	
14	-1.39	-0.31	-1.08	*
20	1.20	1.21	-0.01	
21	2.00	1.19	0.81	*
24	2.11	1.49	0.62	
25	0.23	-0.08	0.31	
26	-1.07	-1.34	0.27	
28	2.75	2.36	0.39	
29	2.95	2.72	0.23	
30	4.73	3.83	0.90	*
31	3.02	2.60	0.42	
35	4.23	3.48	0.75	*
36	5.42	5.25	0.17	
37	5.42	4.68	0.74	*
38	4.59	4.99	-0.40	
39	5.91	5.40	0.51	
40	5.64	5.40	0.24	

where Translation Constant for KCTD = 0.00
Translation Constant for KCTE = 0.53

Chapter V

PROGRAM DESCRIPTION

Form design, item placement, calibration, and fit analyses can all be done by hand, but in practice it is really necessary to automate the process to make it more convenient. This chapter describes a series of computer programs designed to manage the details of item banking. Descriptions of the control cards needed to operate the programs are given in Appendix A.

The data used for illustrating the programs' application come from various administrations of 41 items of the Knox Cube Test (KCT), a subtest of the Arthur Point Scale (Arthur, 1947). This test requires visual attention and short-term memory to accomplish a simple tapping task. The test uses five one-inch cubes, four of which are attached two inches apart on a board. The fifth is used to tap a series on the other four. The four fixed cubes will be called "1," "2," "3," and "4," from left to right, so that each tapping task can be uniquely specified.

The items in these forms range from tasks requiring two taps, 1-4 and 2-3, to one task requiring eight taps,

1-3-2-4-2-3-1-2. A "reversal" refers to a two-tap subsequence that must be performed from right to left, and a "jump" refers to a two-tap subsequence that skips one or more cubes. For example, in the eight-tap sequence 1-3-2-4-2-3-1-2, subsequences [3-2], [4-2], and [3-1] are reversals and subsequences [1-3], [2-4], [4-2], and [3-1] are jumps.

The 41 items used in this example were spread among five forms, such that every form had some items in common with every other form. Table 5.1 shows this complete web with the numbers in each cell indicating the number of items in each of the 10 links. For example, form KCTB was linked to form KCTC by 20 items. The number of items in each form and the number of persons who took each form are shown at the bottom of Table 5.1.

5.1 FORM

5.1.1 Description

The construction of an item bank begins with collecting a pool of items with a common theme. When these items have been written and assembled, computer program, FORM, can be used to distribute them among test forms in a web which maximizes the statistical strength of the linking structure, while meeting the practical requirements of the anticipated testing situation. Input specifications to FORM include the

FORMS ANALYSIS FOR KCT5 BANK - FIVE FORMS - 40 ITEMS
TOT. NO. ITEMS= 40 ITEMS/FORM= 16 ITEMS/LINK= 4
TOT. NO. LINKS= 10 LINKS/FORM= 4 NO. FORMS= 5

ITEM NUMBER	ITEM NUMBER	ITEM NAME	LINK NUMBER	OBJC NUMBER	LINK DIFF	CORR ALT	1ST FORM	2ND FORM
1	1	2101	1	0	3	1	2	1
2	2	2102	1	0	3	1	2	1
3	3	2201	1	0	3	1	2	1
4	4	3102	1	0	3	1	2	1
5	5	3104	2	0	4	1	3	1
6	6	3201	2	0	4	1	3	1
7	7	3301	2	0	4	1	3	1
8	8	4101	2	0	4	1	3	1
9	9	4102	3	0	5	1	4	1
10	10	4103	3	0	5	1	4	1
11	11	4104	3	0	5	1	4	1
12	12	4202	3	0	5	1	4	1
13	13	4203	4	0	5	1	3	2
14	14	4301	4	0	5	1	3	2
15	15	4401	4	0	5	1	3	2
16	16	4402	4	0	5	1	3	2
17	17	5101	5	0	6	1	5	1
18	18	5102	5	0	6	1	5	1
19	19	5103	5	0	6	1	5	1
20	20	5104	5	0	6	1	5	1
21	21	5105	6	0	6	1	4	2
22	22	5108	6	0	6	1	4	2
23	23	5109	6	0	6	1	4	2
24	24	5110	6	0	6	1	4	2
25	25	5201	7	0	7	1	5	2
26	26	5202	7	0	7	1	5	2
27	27	6101	7	0	7	1	5	2
28	28	6102	7	0	7	1	5	2
29	29	6103	8	0	7	1	4	3
30	30	6104	8	0	7	1	4	3
31	31	6108	8	0	7	1	4	3
32	32	6109	8	0	7	1	4	3
33	33	6110	9	0	8	1	5	3
34	34	6111	9	0	8	1	5	3
35	35	6401	9	0	8	1	5	3
36	36	7104	9	0	8	1	5	3
37	37	7105	10	0	9	1	5	4
38	38	7201	10	0	9	1	5	4
39	39	7301	10	0	9	1	5	4
40	40	7402	10	0	9	1	5	4

TOT. NO. ITEMS= 40 ITEMS/FORM= 16 ITEMS/LINK= 4
TOT. NO. LINKS= 10 LINKS/FORM= 4 NO. FORMS= 5

FORMS ANALYSIS FOR KCT5 BANK - FIVE FORMS - 40 ITEMS
TOT. NO. ITEMS= 40 ITEMS/FORM= 16 ITEMS/LINK= 4
TOT. NO. LINKS= 10 LINKS/FORM= 4 NO. FORMS= 5

PROGRAM GENERATED FORM NUMBER	PROGRAM GENERATED ITEM NUMBER	ITEM NUMBER	ITEM NAME	OBJECTIVE NUMBER	CORRECT ALTER	OTHER FORM	LINK NUMBER	LINK DIFF	FORM DIFF	FORM WIDTH
1	1	1	2101	0	1	2	1	3		
1	2	2	2102	0	1	2	1	3		
1	3	3	2201	0	1	2	1	3		
1	4	4	3102	0	1	2	1	3		
1	5	5	3104	0	1	3	2	4		
1	6	6	3201	0	1	3	2	4		
1	7	7	3301	0	1	3	2	4		
1	8	8	4101	0	1	3	2	4		
1	9	9	4102	0	1	4	3	5		
1	10	10	4103	0	1	4	3	5		
1	11	11	4104	0	1	4	3	5		
1	12	12	4202	0	1	4	3	5		
1	17	17	5101	0	1	5	5	6		
1	18	18	5102	0	1	5	5	6		
1	19	19	5103	0	1	5	5	6		
1	20	20	5104	0	1	5	5	6		
									4.5	4
2	1	1	2101	0	1	1	1	3		
2	2	2	2102	0	1	1	1	3		
2	3	3	2201	0	1	1	1	3		
2	4	4	3102	0	1	1	1	3		
2	13	13	4203	0	1	3	4	5		
2	14	14	4301	0	1	3	4	5		
2	15	15	4401	0	1	3	4	5		
2	16	16	4402	0	1	3	4	5		
2	21	21	5105	0	1	4	6	6		
2	22	22	5108	0	1	4	6	6		
2	23	23	5109	0	1	4	6	6		
2	24	24	5110	0	1	4	6	6		
2	25	25	5201	0	1	5	7	7		
2	26	26	5202	0	1	5	7	7		
2	27	27	6101	0	1	5	7	7		
2	28	28	6102	0	1	5	7	7		
									5.3	5
3	5	5	3104	0	1	1	2	4		
3	6	6	3201	0	1	1	2	4		
3	7	7	3301	0	1	1	2	4		
3	8	8	4101	0	1	1	2	4		
3	13	13	4203	0	1	2	4	5		
3	14	14	4301	0	1	2	4	5		
3	15	15	4401	0	1	2	4	5		
3	16	16	4402	0	1	2	4	5		
3	29	29	6103	0	1	4	8	7		
3	30	30	6104	0	1	4	8	7		
3	31	31	6108	0	1	4	8	7		
3	32	32	6109	0	1	4	8	7		
3	33	33	6110	0	1	5	9	8		
3	34	34	6111	0	1	5	9	8		
3	35	35	6401	0	1	5	9	8		
3	36	36	7104	0	1	5	9	8		
									6.0	5
4	9	9	4102	0	1	1	3	5		
4	10	10	4103	0	1	1	3	5		
4	11	11	4104	0	1	1	3	5		
4	12	12	4202	0	1	1	3	5		
4	21	21	5105	0	1	2	6	6		
4	22	22	5108	0	1	2	6	6		
4	23	23	5109	0	1	2	6	6		
4	24	24	5110	0	1	2	6	6		
4	29	29	6103	0	1	3	8	7		
4	30	30	6104	0	1	3	8	7		
4	31	31	6108	0	1	3	8	7		
4	32	32	6109	0	1	3	8	7		
4	37	37	7105	0	1	5	10	9		
4	38	38	7201	0	1	5	10	9		
4	39	39	7301	0	1	5	10	9		
4	40	40	7402	0	1	5	10	9		
									6.8	5
5	17	17	5101	0	1	1	5	6		
5	18	18	5102	0	1	1	5	6		
5	19	19	5103	0	1	1	5	6		
5	20	20	5104	0	1	1	5	6		
5	25	25	5201	0	1	2	7	7		
5	26	26	5202	0	1	2	7	7		
5	27	27	6101	0	1	2	7	7		
5	28	28	6102	0	1	2	7	7		
5	33	33	6110	0	1	3	9	8		
5	34	34	6111	0	1	3	9	8		
5	35	35	6401	0	1	3	9	8		
5	36	36	7104	0	1	3	9	8		
5	37	37	7105	0	1	4	10	9		
5	38	38	7201	0	1	4	10	9		
5	39	39	7301	0	1	4	10	9		
5	40	40	7402	0	1	4	10	9		
									7.5	4

TOT. NO. ITEMS= 40 ITEMS/FORM= 16 ITEMS/LINK= 4
TOT. NO. LINKS= 10 LINKS/FORM= 4 NO. FORMS= 5

TABLE 5.1

A Web for the Knox Cube Test Example

	Form Name				
	KCTA	KCTB	KCTC	KCTD	KCTE
KCTA		14	13	14	11
KCTB			20	26	18
KCTC				28	16
KCTD					25
KCTE					
No. of Items	18	28	38	38	26
No. of Persons	35	102	145	160	313

total number of items to be calibrated, the number of items desired in each form, the number of items desired in each link, and the pattern of form difficulties, parallel or sequential. FORM determines the number of links per form, total number of links, and total number of forms necessary to complete an optimal web.

5.1.2 Example of Output

The FORM example shown here and its resulting web do not coincide with the KCT example used in Chapter 4 or in the illustrations of the other computer programs. This resulted because the constructors of the KCT forms designed them by hand, rather than using FORM. The output presented in this section is one which might have resulted if they had used FORM. It uses a parallel pattern of form difficulties with

a total of 40 items, 16 items per form, and 4 items per link. This produced a web of 5 forms and 10 links, with 4 links per form.

The output from FORM is arranged to serve two purposes. The first part of the FORM output orders the items by identification number and shows each item's name, link number, objective number, link difficulty, correct response, and other forms to which it is linked. This output is used to check that each item is satisfactorily placed and to initiate a file of item specifications from which the banking system works.

The second part of the output groups items by form and lists them in order of their within form position. The same information as in the item list is provided. This output determines which items are to be placed in which forms and facilitates the verification of each form's content coherence.

5.2 FORCAL

5.2.1 Description

When the forms have been designed and constructed, they are administered to samples of suitable persons. The responses to these forms are collected and stored so that there is a record for each person which includes the person's identifi-

cation, the form taken, their item response string, and whatever demographic information is useful. This file of person records, sorted by form, becomes the input for the calibration of forms.

Computer program, FORCAL, takes the item file produced by the FORM program and the person file prepared from the testing, calibrates the items within each form, and performs the initial analysis of the fit of the items within each form. The program also supplies the within form measurement of persons, that is, person ability estimates, their standard errors, and person fit statistics similar to those calculated for items. Although all items on all forms are calibrated at this stage, both the difficulties and abilities are relative to the local origin defined by each form. Comparisons across forms cannot be made until the link analysis has been completed.

5.2.2 Example of Output

The FORCAL output begins with two summary tables for each form. The table on the left shows the number of persons with zero or perfect scores, the number of examinees with scores below the minimum and above the maximum, and the total number of persons in the calibration. This table is compiled before the editing of items and persons for zero or perfect scores. For example, of the 35 examinees who took

the first form, KCTA, no one had zero or perfect raw scores and no one made a score below the minimum or above the maximum. This left 35 persons in the calibration before editing. When the editing process was completed, three items were deleted for zero item scores, one was deleted for a perfect item score, and one person was deleted for a zero raw score. This left 34 persons to calibrate 14 items.

The table on the right of the summary page identifies potentially misfitting items at a glance by listing items with a between fit-t larger than 2.0, total fit-t greater than 1.5 or less than -1.5, or discrimination index less than 0.5. For each of these items, the item name, difficulty, error impact, between and total fit-t, weighted mean square, and discrimination index are listed. The mean and standard deviation for all items are also shown for these statistics. In form KCTA, Items #10 and #28 are selected for this table, because #10 had a between fit-t of 2.66 and #28 had a total fit-t of 1.51. After these two summary tables, the number of items and the number of persons used to calibrate these items are shown along with the mean ability and standard deviation for the "measurable" persons.

The count of persons used to calibrate the items does not include any persons edited out due to zero or perfect raw scores or a raw score outside the chosen minimum/maximum

cation, the form taken, their item response string, and whatever demographic information is useful. This file of person records, sorted by form, becomes the input for the calibration of forms.

Computer program, FORCAL, takes the item file produced by the FORM program and the person file prepared from the testing, calibrates the items within each form, and performs the initial analysis of the fit of the items within each form. The program also supplies the within form measurement of persons, that is, person ability estimates, their standard errors, and person fit statistics similar to those calculated for items. Although all items on all forms are calibrated at this stage, both the difficulties and abilities are relative to the local origin defined by each form. Comparisons across forms cannot be made until the link analysis has been completed.

5.2.2 Example of Output

The FORCAL output begins with two summary tables for each form. The table on the left shows the number of persons with zero or perfect scores, the number of examinees with scores below the minimum and above the maximum, and the total number of persons in the calibration. This table is compiled before the editing of items and persons for zero or perfect scores. For example, of the 35 examinees who took

KNOX CUBE TEST ITEM BANK - FIVE FORMS - KCTA, KCTB, KCTC, KCTD, KCTE

FORM 001 18KCTA

NUMBER OF ZERO SCORES	0
NUMBER LESS THAN MINIMUM	0
NUMBER IN CALIBRATION	35
NUMBER ABOVE MAXIMUM	0
NUMBER OF PERFECT SCORES	0
TOTAL NUMBER OF EXAMINEES	35

ITEM NAME	ITEM DIFF	ERROR IMPACT	FIT T-TESTS BETWN	FIT T-TESTS TOTAL	WGHT MNSQ	DISC INDX
10	-3.65	0.17	2.66	1.17	1.53	0.67
28	3.21	0.26	0.32	1.51	1.77	0.88
MEAN	-0.00		-0.29	0.43	1.14	1.09
S.D.	3.33		1.05	0.70	0.29	0.18

14 ITEMS CALIBRATED ON 34 PERSONS
34 MEASURABLE PERSONS WITH MEAN ABILITY = -0.18 AND STD. DEV. = 1.58

FORM 002 28KCTB

NUMBER OF ZERO SCORES	0
NUMBER LESS THAN MINIMUM	0
NUMBER IN CALIBRATION	102
NUMBER ABOVE MAXIMUM	0
NUMBER OF PERFECT SCORES	0
TOTAL NUMBER OF EXAMINEES	102

ITEM NAME	ITEM DIFF	ERROR IMPACT	FIT T-TESTS BETWN	FIT T-TESTS TOTAL	WGHT MNSQ	DISC INDX
4	-4.43	0.01	3.11	0.35	1.07	0.98
9	-2.40	0.01	2.70	0.44	1.08	1.02
11	-2.40	0.04	2.35	0.69	1.14	1.01
6	-2.92	0.14	0.87	1.51	1.37	0.89
19	1.62	0.13	0.75	1.96	1.34	0.84
MEAN	0.00		-0.16	0.17	1.07	1.05
S.D.	3.86		1.72	1.11	0.28	0.12

26 ITEMS CALIBRATED ON 102 PERSONS
102 MEASURABLE PERSONS WITH MEAN ABILITY = -0.16 AND STD. DEV. = 2.62

FORM 003 38KCTC

NUMBER OF ZERO SCORES	0
NUMBER LESS THAN MINIMUM	0
NUMBER IN CALIBRATION	145
NUMBER ABOVE MAXIMUM	0
NUMBER OF PERFECT SCORES	0
TOTAL NUMBER OF EXAMINEES	145

ITEM NAME	ITEM DIFF	ERROR IMPACT	FIT T-TESTS BETWN	FIT T-TESTS TOTAL	WGHT MNSQ	DISC INDX
17	0.47	0.0	3.97	-5.06	0.45	1.47
18	0.53	0.0	3.57	-4.97	0.45	1.47
26	0.96	0.0	2.48	-3.86	0.54	1.34
15	-1.14	0.0	3.00	-0.12	0.98	1.08
5	-8.48	0.24	6.43	0.86	1.60	0.70
10	-3.04	0.06	4.56	1.36	1.18	0.85
9	-2.75	0.06	6.31	1.49	1.18	0.72
4	-6.62	0.28	2.74	1.50	1.70	0.79
13	-0.96	0.09	6.34	2.16	1.24	0.65
14	-0.72	0.10	3.51	2.20	1.26	0.78
11	-1.00	0.11	3.03	2.49	1.28	0.65
12	-1.51	0.12	3.00	2.81	1.30	0.62
16	-3.04	0.20	9.60	3.36	1.50	0.37
MEAN	-0.00		1.97	-0.13	0.98	1.01
S.D.	3.44		2.84	2.06	0.33	0.26

29 ITEMS CALIBRATED ON 145 PERSONS
145 MEASURABLE PERSONS WITH MEAN ABILITY = -0.98 AND STD. DEV. = 2.06

FORM 004 38KCTD

NUMBER OF ZERO SCORES	1
NUMBER LESS THAN MINIMUM	0
NUMBER IN CALIBRATION	158
NUMBER ABOVE MAXIMUM	1
NUMBER OF PERFECT SCORES	0
TOTAL NUMBER OF EXAMINEES	160

ITEM NAME	ITEM DIFF	ERROR IMPACT	FIT T-TESTS BETWN	FIT T-TESTS TOTAL	WGHT MNSQ	DISC INDX
32	0.87	0.0	2.10	-2.37	0.76	1.27
16	-3.31	0.05	4.15	0.67	1.14	0.91
3	-7.14	0.35	5.67	1.43	1.89	0.85
14	-1.39	0.09	2.88	1.66	1.22	0.76
26	-1.07	0.10	1.84	2.01	1.26	0.81
15	-0.76	0.18	4.03	3.33	1.43	0.57
19	1.69	0.19	3.28	3.50	1.46	0.56
MEAN	-0.00		0.01	0.06	1.03	1.03
S.D.	3.68		2.04	1.36	0.26	0.16

38 ITEMS CALIBRATED ON 158 PERSONS
159 MEASURABLE PERSONS WITH MEAN ABILITY = 0.32 AND STD. DEV. = 2.43

FORM 005 26KCTE

NUMBER OF ZERO SCORES	2
NUMBER LESS THAN MINIMUM	0
NUMBER IN CALIBRATION	309
NUMBER ABOVE MAXIMUM	2
NUMBER OF PERFECT SCORES	0
TOTAL NUMBER OF EXAMINEES	313

ITEM NAME	ITEM DIFF	ERROR IMPACT	FIT T-TESTS BETWN	FIT T-TESTS TOTAL	WGHT MNSQ	DISC INDX
10	-2.95	0.0	2.11	0.03	1.00	0.98
3	-7.40	0.04	16.02	0.40	1.12	0.63
26	-1.87	0.0	2.62	0.43	1.04	0.90
1	-8.15	0.16	22.21	0.69	1.40	0.21
5	-6.30	0.14	9.47	0.93	1.37	0.89
4	-4.80	0.09	3.71	1.08	1.25	0.98
21	0.66	0.06	1.00	2.24	1.16	0.90
MEAN	0.00		2.23	0.28	1.05	1.00
S.D.	3.92		5.49	0.75	0.14	0.21

26 ITEMS CALIBRATED ON 309 PERSONS
311 MEASURABLE PERSONS WITH MEAN ABILITY = -0.22 AND STD. DEV. = 1.91

FORM 005 26KCTE

1	1	2101	1	9	1
2	3	2201	2	9	1
3	5	3104	3	9	1
4	4	3102	4	9	1
5	6	3201	5	9	1
6	7	3301	6	9	1
7	10	4103	7	9	1
8	11	4104	8	9	1
9	12	4202	9	9	1
10	14	4301	10	9	1
11	26	5202	11	9	1
12	25	5201	12	9	1
13	20	5104	13	9	1
14	21	5105	14	9	1
15	24	5110	15	9	1
16	28	6102	16	9	1
17	29	6103	17	9	1
18	31	6108	18	9	1
19	35	6401	19	9	1
20	30	6104	20	9	1
21	38	7201	21	9	1
22	37	7105	22	9	1
23	40	7402	23	9	1
24	39	7301	23	9	1
25	36	7104	25	9	1
26	41	8103	26	9	1

NUMBER OF ACCEPTED ITEMS 26
RECORD FOR FIRST EXAMINEE
001094 005 11110101001100000000000000 080
ITEM SCORES FOR FIRST EXAMINEE
11110101001100000000000000
TOTAL SCORE = 8

NUMBER OF ZERO SCORES	2
NUMBER LESS THAN MINIMUM	0
NUMBER IN CALIBRATION	309
NUMBER ABOVE MAXIMUM	2
NUMBER OF PERFECT SCORES	0
TOTAL NUMBER OF EXAMINEES	313

REJECTED ITEMS

ITEM NUMBER	ITEM NAME	ANSWERED CORRECTLY
NONE		

SUBJECTS DELETED =	0
SUBJECTS REMAINING =	309
ITEMS DELETED =	0
POSSIBLE SCORE =	26
MINIMUM SCORE =	1
MAXIMUM SCORE =	24

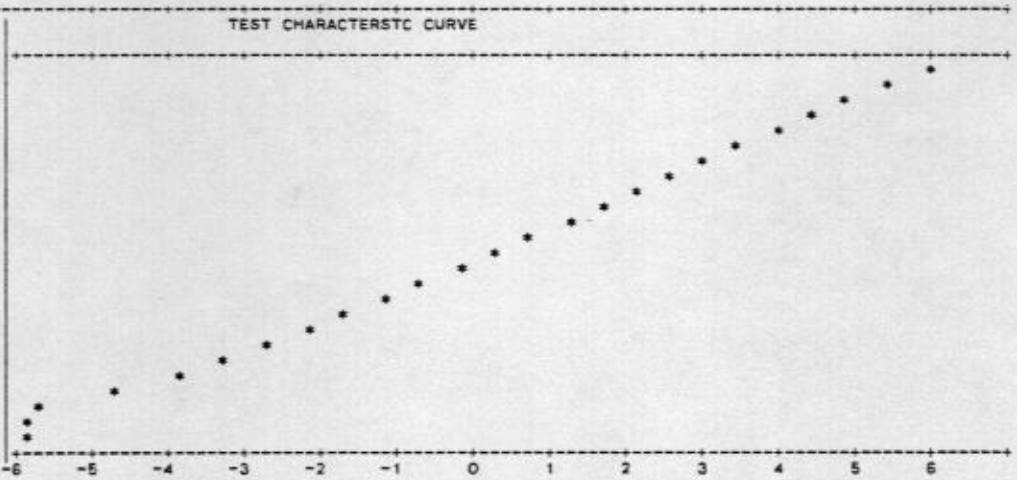
KNOX CUBE TEST ITEM BANK - FIVE FORMS - KCTA, KCTB, KCTC, KCTD, KCTE
FORM 005 26KCTE
PROCEDURE USED UCON
DIFFICULTY SCALE FACTOR 1.436 ABILITY SCALE FACTOR 2.55
NUMBER OF ITERATIONS = 5

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-8.155	1.048	-0.055	-8.527	-7.513
2	3	-7.398	0.767	-0.051	-7.527	-6.788
3	5	-6.298	0.507	-0.042	-6.197	-5.762
4	4	-4.798	0.317	-0.031	-4.572	-4.383
5	6	-3.032	0.208	-0.018	-2.878	-2.780
6	7	-2.870	0.201	-0.017	-2.731	-2.635
7	10	-2.950	0.204	-0.017	-2.803	-2.706
8	11	-2.280	0.179	-0.012	-2.199	-2.102
9	12	-1.639	0.162	-0.008	-1.622	-1.524
10	14	-0.836	0.148	-0.003	-0.894	-0.799
11	26	-1.874	0.168	-0.010	-1.834	-1.736
12	25	-0.606	0.146	-0.001	-0.683	-0.592
13	20	0.678	0.149	0.007	0.511	0.569
14	21	0.656	0.148	0.006	0.491	0.549
15	24	0.960	0.152	0.008	0.780	0.826
16	28	1.829	0.171	0.014	1.627	1.622
17	29	2.192	0.183	0.016	1.994	1.958
18	31	2.065	0.179	0.015	1.865	1.840
19	35	2.953	0.217	0.020	2.799	2.676
20	30	3.304	0.238	0.022	3.189	3.013
21	38	4.460	0.346	0.026	4.578	4.152
22	37	4.148	0.309	0.025	4.187	3.840
23	40	4.868	0.404	0.027	5.105	4.563
24	39	4.868	0.404	0.027	5.105	4.563
25	36	4.717	0.381	0.027	4.908	4.411
26	41	5.038	0.432	0.028	5.331	4.737

ROOT MEAN SQUARE = 0.025
26 ITEMS CALIBRATED ON 309 PERSONS
311 MEASURABLE PERSONS WITH MEAN ABILITY = -0.22 AND STD. DEV. = 1.91

KNOX CUBE TEST ITEM BANK - FIVE FORMS - KCTA, KCTB, KCTC, KCTD, KCTE
FORM 005 26KCTE
COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS
25	2	6.38	1.08
24	0	5.55	0.82
23	1	4.97	0.74
22	0	4.47	0.70
21	2	4.02	0.68
20	5	3.57	0.68
19	10	3.13	0.68
18	6	2.68	0.68
17	9	2.23	0.68
16	20	1.78	0.69
15	22	1.32	0.69
14	25	0.85	0.70
13	33	0.37	0.71
12	32	-0.12	0.72
11	46	-0.62	0.72
10	18	-1.11	0.72
9	20	-1.61	0.72
8	20	-2.10	0.72
7	9	-2.61	0.74
6	6	-3.16	0.78
5	14	-3.80	0.86
4	7	-4.61	0.97
3	2	-5.64	1.08
2	2	-6.81	1.12
1	0	-8.12	1.25



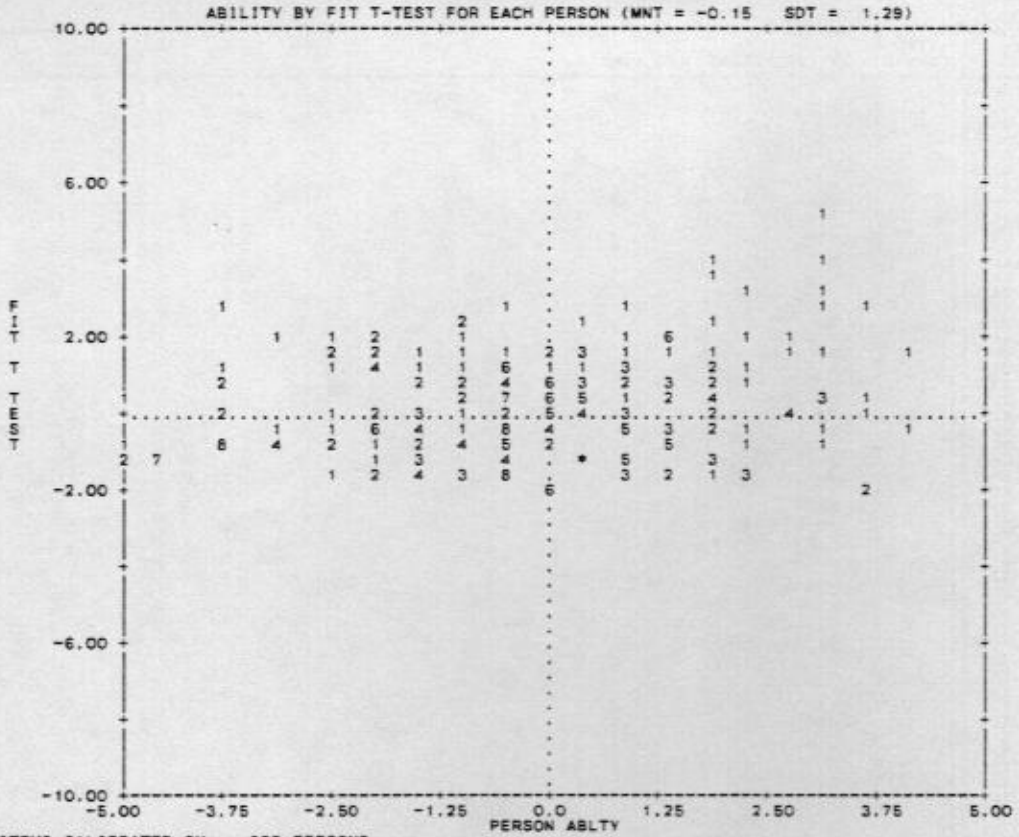
PERSON SEPARABILITY INDEX 0.87 (EQUIVALENT TO KR20)
26 ITEMS CALIBRATED ON 309 PERSONS
311 MEASURABLE PERSONS WITH MEAN ABILITY = -0.22 AND STD. DEV. = 1.91

KNOX CUBE TEST ITEM BANK - FIVE FORMS - KCTA, KCTB, KCTC, KCTD, KCTE
FORM 005 26KCTE
MAP OF VARIABLE

PERSON STATS	COUNT	RAW SCORE	MEASURE MIDPOINT(S.E.)	ITEM COUNTS	TYPICAL ITEMS (BY NAME)	
	2	25	6.30(1.08)			
			6.10(1.08)			
			5.90(1.08)			
			5.70(1.08)			
+3SD		24	5.50(0.82)			
			5.30(0.82)			
			5.10(0.82)	1	41	
	1	23	4.90(0.74)	2	40	39
			4.70(0.74)	1	36	
		22	4.50(0.70)	1	38	
			4.30(0.70)			
	2	21	4.10(0.68)	1	37	
			3.90(0.68)			
+2SD			3.70(0.68)			
	5	20	3.50(0.68)			
			3.30(0.68)	1	30	
	10	19	3.10(0.68)			
			2.90(0.68)	1	35	
	6	18	2.70(0.68)			
			2.50(0.68)			
	9	17	2.30(0.68)			
			2.10(0.68)	2	29	31
			1.90(0.68)	1	28	
+1SD	20	16	1.70(0.69)			
			1.50(0.69)			
	22	15	1.30(0.69)			
			1.10(0.69)			
	25	14	0.90(0.70)	1	24	
			0.70(0.70)	2	20	21
			0.50(0.70)			
	33	13	0.30(0.71)			
			0.10(0.71)			
	32	12	-0.10(0.72)			
MEAN			-0.30(0.72)			
			-0.50(0.72)			
	46	11	-0.70(0.72)	1	25	
			-0.90(0.72)	1	14	
	18	10	-1.10(0.72)			
			-1.30(0.72)			
			-1.50(0.72)			
	20	9	-1.70(0.72)	1	12	
			-1.90(0.72)	1	26	
-1SD	20	8	-2.10(0.72)			
			-2.30(0.72)	1	11	
			-2.50(0.72)			
	9	7	-2.70(0.74)			
			-2.90(0.74)	2	7	10
	6	6	-3.10(0.78)	1	6	
			-3.30(0.78)			
			-3.50(0.78)			
			-3.70(0.78)			
	14	5	-3.90(0.86)			
-2SD			-4.10(0.86)			
			-4.30(0.86)			
			-4.50(0.86)			
	7	4	-4.70(0.97)	1	4	
			-4.90(0.97)			
			-5.10(0.97)			
			-5.30(0.97)			
			-5.50(0.97)			
	2	3	-5.70(1.08)			
-3SD			-5.90(1.08)			
			-6.10(1.08)			
			-6.30(1.08)	1	5	
			-6.50(1.08)			
			-6.70(1.08)			
	2	2	-6.90(1.12)			
			-7.10(1.12)			
			-7.30(1.12)	1	3	
			-7.50(1.12)			
			-7.70(1.12)			
-4SD			-7.90(1.12)			
		1	-8.10(1.25)	1	1	
			-8.30(1.25)			

26 ITEMS CALIBRATED ON 309 PERSONS
311 MEASURABLE PERSONS WITH MEAN ABILITY = -0.22 AND STD. DEV. = 1.91

KNOX CUBE TEST ITEM BANK - FIVE FORMS - KCTA, KCTB, KCTC, KCTD, KCTE
FORM 005 26KCTE



26 ITEMS CALIBRATED ON 309 PERSONS
311 MEASURABLE PERSONS WITH MEAN ABILITY = -0.22 AND STD. DEV. = 1.91

SEQ NUM	ITEM NAME	ITEM CHARACTERISTIC CURVE 1ST GROUP	2ND GROUP	3RD GROUP	4TH GROUP	5TH GROUP	6TH GROUP	DEPARTURE FROM EXPECTED ICC 1ST GROUP	2ND GROUP	3RD GROUP	4TH GROUP	5TH GROUP	6TH GROUP	ITEM FIT STATISTICS * ERR IMPAC	FIT T-TESTS BETWN	TOTAL	WTD MNSQ	MNSQ SD	DISC INDX	POINT BISER
1	1	1.00	1.00	1.00	1.00	1.00	0.98	0.01	0.00	0.00	0.00	0.00	-0.02	0.16	22.21	0.69	1.40	0.99	0.21	-0.10
2	3	0.98	1.00	1.00	1.00	1.00	0.98	0.01	0.00	0.00	0.00	0.00	-0.02	0.04	16.02	0.40	1.12	0.66	0.63	0.03
3	5	0.93	1.00	1.00	1.00	1.00	0.98	0.00	0.01	0.00	0.00	0.00	-0.02	0.14	9.47	0.93	1.37	0.41	0.89	0.13
4	4	0.80	0.97	0.98	1.00	1.00	0.98	-0.00	0.00	-0.01	0.01	0.00	-0.02	0.09	3.71	1.08	1.25	0.23	0.98	0.29
5	6	0.48	0.89	0.89	0.97	0.98	0.98	-0.01	0.04	-0.04	0.01	-0.01	-0.01	0.04	-0.03	0.92	1.12	0.13	0.99	0.48
6	7	0.38	0.92	0.93	0.97	0.94	1.00	-0.00	0.09	0.02	0.01	-0.05	0.01	0.0	1.61	-0.01	0.99	0.13	1.07	0.52
7	10	0.45	0.89	0.93	0.95	0.98	0.96	-0.03	0.05	0.01	-0.01	-0.01	-0.03	0.0	2.11	0.03	1.00	0.13	0.98	0.49
8	11	0.37	0.55	0.87	0.95	0.98	1.00	0.02	-0.18	0.02	0.03	0.01	0.01	0.0	0.97	0.01	1.00	0.11	1.05	0.58
9	12	0.23	0.58	0.74	0.88	0.91	0.96	0.00	-0.00	-0.01	0.02	-0.03	-0.02	0.01	-1.10	0.79	1.07	0.09	0.96	0.56
10	14	0.08	0.42	0.65	0.72	0.83	0.92	-0.04	0.04	0.09	-0.01	-0.04	-0.04	0.01	0.27	0.93	1.07	0.07	0.95	0.58
11	26	0.25	0.71	0.80	0.91	0.91	0.92	-0.02	0.07	0.01	0.02	-0.04	-0.06	0.0	2.62	0.43	1.04	0.09	0.90	0.55
12	25	0.08	0.21	0.59	0.71	0.89	0.89	-0.02	-0.11	0.08	0.03	0.05	-0.06	0.0	1.31	0.28	1.02	0.07	1.01	0.59
13	20	0.03	0.13	0.20	0.38	0.62	0.79	0.00	0.02	-0.02	0.02	0.03	-0.05	0.03	-1.47	1.37	1.10	0.07	0.94	0.54
14	21	0.02	0.08	0.28	0.43	0.62	0.74	-0.01	-0.04	0.07	0.06	0.02	-0.10	0.06	1.00	2.24	1.16	0.07	0.90	0.52
15	24	0.02	0.03	0.11	0.35	0.60	0.77	-0.01	-0.06	-0.06	0.06	0.08	-0.02	0.0	0.29	-0.22	0.98	0.07	1.07	0.58
16	28	0.0	0.03	0.02	0.15	0.36	0.66	-0.01	-0.01	-0.05	0.01	0.06	0.03	0.0	-0.32	-0.00	1.00	0.10	1.10	0.53
17	29	0.0	0.0	0.0	0.03	0.32	0.66	-0.01	-0.03	-0.05	-0.08	0.09	0.11	0.0	1.97	-1.03	0.89	0.11	1.21	0.56
18	31	0.0	0.03	0.0	0.06	0.30	0.70	-0.01	-0.00	-0.06	-0.06	0.04	0.12	0.0	1.33	-0.98	0.90	0.10	1.19	0.56
19	35	0.0	0.03	0.0	0.02	0.11	0.47	-0.00	0.01	-0.02	-0.04	-0.02	0.09	0.0	0.54	-1.09	0.85	0.14	1.13	0.49
20	30	0.0	0.0	0.0	0.02	0.09	0.38	-0.00	-0.01	-0.02	-0.02	-0.00	0.07	0.0	-0.31	-0.06	0.98	0.16	1.13	0.44
21	38	0.0	0.0	0.0	0.0	0.02	0.17	-0.00	-0.00	-0.01	-0.01	-0.01	0.04	0.05	-1.05	0.64	1.16	0.28	1.12	0.28
22	37	0.0	0.0	0.0	0.0	0.02	0.23	-0.00	-0.00	-0.01	-0.02	-0.02	0.06	0.0	-0.38	0.08	1.00	0.24	1.14	0.35
23	40	0.0	0.0	0.0	0.0	0.0	0.13	-0.00	-0.00	-0.00	-0.01	-0.02	0.04	0.0	-0.68	0.16	1.02	0.34	1.14	0.27
24	39	0.0	0.0	0.0	0.0	0.0	0.13	-0.00	-0.00	-0.00	-0.01	-0.02	0.04	0.0	-0.68	0.16	1.02	0.34	1.14	0.27
25	36	0.0	0.0	0.0	0.0	0.0	0.15	-0.00	-0.00	-0.00	-0.01	-0.02	0.05	0.0	-0.46	-0.41	0.85	0.32	1.14	0.34
26	41	0.0	0.0	0.0	0.0	0.0	0.11	-0.00	-0.00	-0.00	-0.01	-0.02	0.03	0.0	-0.91	-0.01	0.95	0.37	1.13	0.28

SCORE RANGE	1- 8	9-10	11-11	12-13	14-15	16-24
MEAN ABILITY	-3.25	-1.37	-0.62	0.13	1.07	2.53
MEAN Z-TEST	-0.3	-0.1	-0.3	-0.1	-0.2	-2.7
SD(Z-TEST)	0.5	0.8	0.8	0.8	0.8	7.3
GROUP COUNT	60	38	46	65	47	53

PLUS=TOO MANY RIGHT
MINUS=TOO MANY WRONG

*ERROR IMPACT = PROPORTION ERROR INCREASE DUE TO THIS MISFIT

26 ITEMS CALIBRATED ON 309 PERSONS
311 MEASURABLE PERSONS WITH MEAN ABILITY = -0.22 AND STD. DEV. = 1.91

SERIAL ORDER SEQ NUM	ITEM NAME	ITEM DIFF	STD ERROR	DISC INDX	FIT TTEST	DIFFICULTY ORDER SEQ NUM	ITEM NAME	ITEM DIFF	DISC INDX	FIT TTEST	FIT ORDER SEQ NUM	ITEM NAME	ITEM DIFF	ERR IMPAC	FIT T-TESTS BETWN	TOTAL	WTD MNSQ	MNSQ SD	DISC INDX	POINT BISER
1	1	-8.15	1.24	0.21	0.69	1	1	-8.15	0.21	0.69	19	35	2.95	0.0	0.54	-1.09	0.85	0.14	1.13	0.49
2	3	-7.40	0.81	0.63	0.40	2	3	-7.40	0.63	0.40	17	29	2.19	0.0	1.97	-1.03	0.89	0.11	1.21	0.56
3	5	-6.30	0.59	0.89	0.93	3	5	-6.30	0.89	0.93	18	31	2.07	0.0	1.33	-0.98	0.90	0.10	1.19	0.56
4	4	-4.80	0.35	0.98	1.08	4	4	-4.80	0.98	1.08	25	36	4.72	0.0	-0.46	-0.41	0.85	0.32	1.14	0.34
5	6	-3.03	0.22	0.99	0.92	5	6	-3.03	0.99	0.92	15	24	0.96	0.0	0.29	-0.22	0.98	0.07	1.07	0.58
6	7	-2.87	0.20	1.07	-0.01	7	10	-2.95	0.98	0.03	20	30	3.30	0.0	-0.31	-0.06	0.98	0.16	1.13	0.44
7	10	-2.95	0.20	0.98	0.03	6	7	-2.87	1.07	-0.01	6	7	-2.87	0.0	1.61	-0.01	0.99	0.13	1.07	0.52
8	11	-2.28	0.18	1.05	0.01	8	11	-2.28	1.05	0.01	26	41	5.04	0.0	-0.91	-0.01	0.95	0.37	1.13	0.28
9	12	-1.64	0.17	0.96	0.79	11	26	-1.87	0.90	0.43	16	28	1.83	0.0	-0.32	-0.00	1.00	0.10	1.10	0.53
10	14	-0.84	0.15	0.95	0.93	9	12	-1.64	0.96	0.79	8	11	-2.28	0.0	0.97	0.01	1.00	0.11	1.05	0.58
11	26	-1.87	0.17	0.90	0.43	10	14	-0.84	0.95	0.93	7	10	-2.95	0.0	2.11	0.03	1.00	0.13	0.98	0.49
12	25	-0.61	0.15	1.01	0.28	12	25	-0.61	1.01	0.28	22	37	4.15	0.0	-0.38	0.08	1.00	0.24	1.14	0.35
13	20	0.68	0.16	0.94	1.37	14	21	0.66	0.90	2.24	23	40	4.87	0.0	-0.68	0.16	1.02	0.34	1.14	0.27
14	21	0.66	0.16	0.90	2.24	13	20	0.68	0.94	1.37	24	39	4.87	0.0	-0.68	0.16	1.02	0.34	1.14	0.27
15	24	0.96	0.15	1.07	-0.22	15	24	0.96	1.07	-0.22	12	25	-0.61	0.0	1.31	0.28	1.02	0.07	1.01	0.59
16	28	1.83	0.17	1.10	-0.00	16	28	1.83	1.10	-0.00	2	3	-7.40	0.04	16.02	0.40	1.12	0.66	0.63	0.03
17	29	2.19	0.18	1.21	-1.03	18	31	2.07	1.19	-0.98	11	26	-1.87	0.0	2.62	0.43	1.04	0.09	0.90	0.55
18	31	2.07	0.18	1.19	-0.98	17	29	2.19	1.21	-1.03	21	38	4.46	0.05	-1.05	0.64	1.16	0.28	1.12	0.28
19	35	2.95	0.22	1.13	-1.09	19	35	2.95	1.13	-1.09	1	1	-8.15	0.16	22.21	0.69	1.40	0.99	0.21	-0.10
20	30	3.30	0.24	1.13	-0.06	20	30	3.30	1.13	-0.06	9	12	-1.64	0.01	-1.10	0.79	1.07	0.09	0.96	0.56
21	38	4.46	0.37	1.12	0.64	22	37	4.15	1.14	0.08	5	6	-3.03	0.04	-0.03	0.92	1.12	0.13	0.99	0.48
22	37	4.15	0.31	1.14	0.08	21	38	4.46	1.12	0.64	3	5	-6.30	0.14	9.47	0.93	1.37	0.41	0.89	0.13
23	40	4.87	0.41	1.14	0.16	25	36	4.72	1.14	-0.41	10	14	-0.84	0.01	0.27	0.93	1.07	0.07	0.95	0.58
24	39	4.87	0.41	1.14	0.16	24	39	4.87	1.14	0.16	4	4	-4.80	0.09	3.71	1.08	1.25	0.23	0.98	0.29
25	36	4.72	0.38	1.14	-0.41	23	40	4.87	1.14	0.16	13	20	0.68	0.03	-1.47	1.37	1.10	0.07	0.94	0.54
26	41	5.04	0.43	1.13	-0.01	26	41	5.04	1.13	-0.01	14	21	0.66	0.06	1.00	2.24	1.16	0.07	0.90	0.52
MEAN		0.00		1.00	0.28										2.23	0.28	1.05	0.22		
S.D.		3.92		0.21	0.75										5.49	0.75	0.14	0.21		

26 ITEMS CALIBRATED ON 309 PERSONS
311 MEASURABLE PERSONS WITH MEAN ABILITY = -0.22 AND STD. DEV. = 1.91
END OF FORM 1005 26KCTE
26 ITEMS CALIBRATED ON 309 PERSONS
311 MEASURABLE PERSONS WITH MEAN ABILITY = -0.22 AND STD. DEV. = 1.91
END OF JOB

range. The count of "measurable" persons excludes only those persons who make a zero or perfect raw score, as a person's ability cannot be estimated when they answer all the items either incorrectly or correctly. One of the 35 persons who took form KCTA was eliminated for this reason, leaving 34 measurable persons used to calibrate 14 items. The mean ability of the 34 persons was -0.18, indicating that the center of the test was slightly above the average ability of the persons who took it. The standard deviation of ability, corrected for the error of measurement, was 1.58.

The calibration of items within each form follows this summary information. Form KCTE will be described for the purposes of this example. At the top of each page is a user-supplied title, which is a description of the analysis, followed by a program-supplied title, which identifies the form being calibrated.

Page 1 of the FORCAL analysis of each form shows the input control parameters. The first six parameters define the incoming person records: the length of each person record; the position of the first item response; the length, in columns, of the identification field; the starting column of the identification field; the column length of the form number field; and, the starting position of the form number field.

Other control parameters include: the minimum and maximum percent correct allowed for a person to be included in the calibration sample; the minimum average group size for the between group fit analysis; the unit number from which the incoming person file is to be read; the estimation procedure selected; the maximum total fit-t criterion allowed for a person to be included in the recalibration sample; the item delete code; the item scoring code; and the minimum item status code.

The minimum and maximum score specifications allow the elimination of persons who score too low or too high. This parameter is included because extremely high and extremely low scorers frequently behave erratically. In achievement testing with multiple choice items, the minimum score for item calibration should be set well above the chance level to avoid contaminating the estimates of item difficulty with the effects of whatever random guessing may have occurred.

This analysis used persons whose raw scores were between 5% and 95% of the total possible score for any one form. The number of items in form KCTE was 26, therefore, the minimum score was 1 and the maximum 24. (Random guessing is not an issue on this test because the items are individually administered and scored.)

The minimum group size used was ten. The between group fit analysis is based on up to six score groups. For small data sets this parameter can be used to restrict the program to fewer than six groups.

The calibration procedure chosen was the unconditional method, UCON. The person screening code allows the deletion of persons who misfit on the basis of their response pattern fit statistic. When this option is requested and misfitting persons are encountered, the items in the form are recalibrated with the misfitting persons eliminated. No persons were eliminated for misfit in this analysis. The items were already scored, coded 1 for pass and 0 for fail.

The item status code was 1, meaning that all items with a status code greater than or equal to 1 were to be used in the analysis. The status code is a user controlled switch to delete items from the analysis. Acceptable codes are integers 0 to 9. In this case, only items coded "0" will be deleted.

Page 2 of the FORCAL analysis lists the items in the form, by number and first four characters of their name, position in the form, status in the form, and correct answer. As all items in form KCTE had a status code of 9, all the items were included in the analysis.

Following this item list is the number of accepted items, that is, the number of items with status codes greater than 0. Then the first person record for the form is listed for verification, along with this person's scored responses (in this case, they are identical to the input record), and their number of correct responses. This output helps verify that the analysis being performed is the one intended.

The next piece of output reviews the editing process. The editing routine successively removes persons with zero or perfect raw scores and items with zero or perfect item scores. A zero raw score occurs when a person correctly answers no items in the test and a perfect raw score occurs when a person correctly answers all items in the test. A zero item score occurs when an item is not answered correctly by any persons and a perfect item score occurs when an item is answered correctly by all persons. Since the deletion of a person may change the status of some items and the deletion of an item may change the status of some persons, this editing is iterated until no more persons or items can be deleted for zero or perfect scores.

In form KCTE, two persons were edited out for making a zero raw score but no persons made a perfect raw score of 26. No one received a raw score below the minimum of 1 but two persons were deleted from calibration for making a raw

score above the maximum of 24. No items were rejected for zero or perfect item scores. This left 309 persons to be used in the calibration of 26 items.

Page 3 of the output contains the item difficulty estimates in logits and centered on zero. At the top of this page the estimation procedure used is shown along with the PROX difficulty and ability expansion factors. These expansion factors are used to scale the initial logit estimates for the normal approximation method, PROX. In addition to the difficulty estimates and their standard errors, the table displays: the magnitude of improvement in the last UCON iteration, labelled "LAST DIFF CHANGE"; the PROX difficulty estimates, "PROX DIFF"; and the difficulty estimates after one iteration of UCON, "FIRST CYCLE."

The 26 item difficulties in form KCTE have a range of more than 13 logits, from -8.15 logits, for Item #1, to 5.04 logits, for Item #41. Although this is an extremely wide test, with an item standard deviation of 3.92, these difficulties are rather evenly spaced, as the map on output Page 5 shows.

Page 4 of the output provides the conversion of person raw scores to person ability measures with their associated standard errors. This table also shows the number of per-

sons in the sample who received each of the 25 scores in form KCTE. Accompanying this conversion table is the test characteristic curve which pictures the range of ability of these 25 raw scores. It is provided to show the extent to which the relation between scores and measures is nonlinear. On this wide test the nonlinearity between scores 6 and 24 is almost invisible. However, the increment between adjacent scores varies from 0.44 between scores 19 and 20 to 1.31 between scores 1 and 2.

Page 5 of the output displays a "map" of the variable defined by these persons and these items. It shows the distribution of persons and items along the variable. Values for the variable are shown in the center column labelled "measure midpoint." These values are the midpoints of the intervals covered by that row of the map. Because the persons and items use the same measurement scale, these midpoints simultaneously describe the persons and the items. The measure midpoints are accompanied, in parentheses, by the standard errors of person measurement associated with that position on the scale based on this form. (Consult output Page 3 for the item difficulty standard errors.)

The person section, to the left of the midpoint column, shows the ability mean and standard deviation of the calibrating sample and the number of persons obtaining each raw

score. The item section, to the right of the midpoint column, shows the distribution of the items over the variable with each item identified by name.

When applied to the persons the "measure midpoints" and their standard errors refer to person ability. For example, in form KCTE, the mean ability of the 311 persons is shown in the interval between -0.10 and -0.30 logits. The exact sample mean ability is -0.22 with a standard deviation of 1.91. This is shown at the bottom of every page.

Forty-six persons, with raw scores of 11 and a corresponding ability between -0.50 and -0.70 logits (with a standard error of 0.72), stand 0.40 logits below the mean. These forty-six persons are also shown on Page 4, where their estimated ability is -0.62 logits with a standard error of 0.72.

When applied to the items the "measure midpoints" refer to item difficulty. For example, Item #41, the hardest item in form KCTE, is positioned at a difficulty between 5.10 and 4.90 logits. The exact difficulty estimate of Item #41 is shown on Page 3 as 5.038 logits with a standard error of 0.432.

Page 6 plots person abilities against person fit-t's. The mean and standard deviation of person fit-t's over all measured persons is shown at the top of the plot. This mean has an expected value near 0.0 with a standard deviation of about 1.0. In form KCTE, the mean total fit-t is -0.15 with a standard deviation of 1.29.

The numbers within the plot represent the count of persons at any particular coordinate. An asterisk indicates 10 or more persons occurred at that point. The plot shows that the persons who took form KCTE have a wide range of ability. In addition, there are a number of persons who show some signs of misfit. Fifteen persons had total fit-t's above 2.0 and might be excluded from item calibration because of this misfit. However, our experience has been that removing this small number of misfitting persons, on the basis of their total fit-t, does not change the results of the calibration significantly. Thus, these persons have been left in this analysis.

Of the 15 misfits, 11 are on the able side of the plot, indicating that their misfit must be due to lapses on tasks which should have been easy for them. The appearance of improbable failures in an individual's response record can have significant diagnostic implications. The diagnostic aspects of fit analysis make up a rich and important topic

which unfortunately goes beyond the scope of the immediate study.

The fit of the items within the form can be examined on Pages 7 and 8 of each FORCAL form analysis. Page 7 is divided into three sections. Section 1 shows the item characteristic curve (ICC) observed for each item. This curve is formed by dividing the persons, by score level, into up to six groups of approximately equal size, and finding the proportion of correct answers to each item given by each score group. It is expected that persons in the lower score groups will have less success in answering correctly and those in the higher score groups will have more. The proportion correct is expected to increase as the groups become more able, that is, from left to right in this section. The score range, mean ability, and count of persons in each group are given at the bottom of the first panel.

The second panel on Page 7 shows the proportional departures of the observed item characteristic curves from the ICC's expected by the model using estimates from all persons. Ideally, these departures should be small, indicating that the observed ICC is close to the expected.

Panel 3 on Page 7 gives fit statistics associated with each item. The error impact, between and total fit-t,

weighted mean square, its standard error, a discrimination index, and a point biserial are shown for each item.

The table on Page 8 of the FORCAL output is also divided into three panels. The first panel lists the items in serial order, by their sequence number, name, difficulty and standard error, discrimination index, and total fit-t. The second panel lists the items arranged in difficulty order, by name, discrimination index, and total fit-t. The third panel lists the items in order of total fit-t and shows for each item: item name, item difficulty, error impact, between and total fit-t, weighted mean square, its standard error, a discrimination index, and a point biserial.

Items with a between fit-t larger than 3 or 4 or a total fit-t greater than 2 or 3 should be examined for miskeying, misprinting or other inconsistencies that might cause misfit. If the large fit statistic cannot be traced to a mechanical failure, the content of the item should be scrutinized to make sure it belongs within the realm of the variable intended. Thorough examination of the content of misfitting items is essential in item fit analysis and is almost always rewarding.

There are four items in form KCTE with a between fit-t larger than 3.0. These are: Items #1, #3, #5, and #4, with

between fit-t's of 22.21, 16.02, 9.47, and 3.71, respectively. The total fit-t has a mean of 0.28 and a standard deviation of 0.75. There is one item with a total fit-t greater than 2.0, #21.

The four items with between fit-t's larger than 3.0 are the four easiest items in form KCTE. Item #1, the easiest item in the form and the one with the most misfit, also has a very low discrimination index and a negative point biserial. If we examine the characteristic curve for this item, we see that this item does not differentiate between ability groups. Everyone in the first five ability groups answered this item correctly and only two percent (one person) in the highest ability group failed the item. It is likely that this incorrect answer was the result of carelessness, scoring error, or confusion about the task, as the item required the very simple two-tap sequence 1-4.

Item #3, the two-tap sequence 2-3, is slightly more difficult than Item #1 and discriminated slightly better, with a discrimination index of 0.63, but it is still far too easy for this sample. The ICC for Item #3 shows that two percent (one person) of the lowest ability group and two percent (one person) of the highest ability group incorrectly answered this item. All the persons in the middle four ability groups answered the item correctly. Again, it is

easy to explain failure in the highest ability group as a result of carelessness or scoring error. The "misfit" was the result of just one able person unexpectedly failing this item.

Items #5 and #4 are not quite as easy as the two preceeding ones. Both items required a three-tap sequence with one jump. Of the few persons who missed these items, again only one was in the highest ability group. But because these items were so easy, this single unexpected failure was surprising enough to produce the large between group fit statistics.

All four of the items showing between group misfit were very easy and appeared early in the test. In each case, the misfit was caused by one high ability person failing the item. Whether this single exception to expectation is due to carelessness or misunderstanding instructions on the part of the examinee or scoring error on the part of the examiner, it has little effect on the calibrations of the items.

5.3 SHIFT

5.3.1 Description

Once items have been calibrated within forms there are as many difficulty estimates for each item as there are forms in which it appears. The items that appear in more than one

form are the items which provide linking data. Computer program, SHIFT, uses the observed differences between these within form item calibrations and the fundamental requirement that each item be characterized by a single difficulty, regardless of form or sample, to calculate a difficulty for each form. This form difficulty is then added to the within form item calibrations to place every administration of every item onto the common bank scale.

The fit analysis produced by SHIFT begins with the links between forms. Each item in each link is evaluated for its fit to that link. Next, there is an analysis of the fit of each link into the web structure. Finally, there is an analysis of the fit of each form into the bank. These statistics were described in Chapter 3.

The only items used to calibrate the forms are those which satisfy the fit analyses. However, all items are reported on.

5.3.2 Example of Output

The first page of the SHIFT output reviews the control parameters used to describe the run. These parameters define the incoming item records: the length of the item record; the unit number from which the incoming item file is to be read; the starting column of the form name field; the column

length of the form name field; the first column of the form number field; the length, in columns, of the form number field, and; the identification number of the lowest form number.

The second page of output is a summary table which provides for each form, its form number and name, the number of links to other forms, its translation constant and standard error, the form fit statistic, and the between form and within form fit statistics.

This table is divided into two panels. The first is based on all available item links and the second is computed after deleting item links with a logit residual larger than a specified value. In this analysis, the cut-off value is 0.7 logits. The translation constant for every form and the linking constants are recomputed after deleting those item links which do not satisfy this criterion.

The last column on the second page provides the difference between the translation constants before and after editing. These differences will be small if the deletion of misfitting links does not affect the relative positions of the forms.

LENGTH OF ITEM RECORD	80
INPUT UNIT NUMBER	1
START OF FORM NAME	15
LENGTH OF FORM NAME	4
START OF FORM NUMBER	6
LENGTH OF FORM NUMBER	3
LOWEST FORM NUMBER	1

41 ORIGINAL ITEMS
4 ITEMS NOT USED IN LINK DESIGN
43 LINKS NOT USED IN LINK DESIGN
106 LINKS USED IN DESIGN
79 LINKS DROPPED FROM DESIGN

FORM NMBR	FORM NAME	NMBR OF LINKS	TRANS CONST	TRANS S.E.	FORM FIT	BETWN FORM	WTHN FORM	FORM NMBR	NMBR OF LINKS	TRANS CONST	TRANS S.E.	FORM FIT	BETWN FORM	WTHN FORM	TRANS CONST DIFF
1	KCTA	4.	-0.00	0.03	4.32	-0.66	-1.12	1	4.	0.00	0.07	-0.30	3.33	-1.05	-0.01
2	KCTB	4.	-0.04	0.02	3.15	2.22	5.13	2	4.	-0.04	0.03	3.69	3.41	1.59	-0.01
3	KCTC	4.	-0.48	0.02	12.18	7.21	8.33	3	4.	-0.47	0.04	5.50	1.56	1.30	-0.01
4	KCTD	4.	-0.00	0.02	12.98	5.64	6.69	4	4.	-0.08	0.03	4.87	4.68	2.54	0.08
5	KCTE	4.	0.53	0.02	5.39	3.81	2.00	5	4.	0.58	0.06	2.31	3.08	0.90	-0.06

ITEMS NOT USED IN LINK DESIGN		LINKS NOT USED IN LINK DESIGN		LINKS DROPPED FROM DESIGN DUE TO LARGE RESIDUAL			LINKS USED IN LINK DESIGN		
ITEM NAME	FORM NAME	ITEM NAME	LINK AFFECTED	ITEM NAME	LINK AFFECTED	RESIDUAL	ITEM NAME	LINK AFFECTED	RESIDUA
		1 2101	1, 2						
		1 2101	1, 3						
		1 2101	1, 4						
		1 2101	1, 5						
		1 2101	2, 3						
		1 2101	2, 4						
		1 2101	2, 5						
		1 2101	3, 4						
		1 2101	3, 5						
							1 2101	4, 5	-0.40
2 2102	1 KCTA								
		3 2201	3, 2						
		3 2201	3, 4						
		3 2201	3, 5						
							3 2201	2, 4	0.45
							3 2201	2, 5	0.18
							3 2201	4, 5	-0.27
							4 3102	1, 2	0.28
				4 3102	1, 3	2.91			
							4 3102	1, 4	-0.23
							4 3102	1, 5	0.08
				4 3102	2, 3	2.62			
							4 3102	2, 4	-0.51
							4 3102	2, 5	-0.20
				4 3102	3, 4	-3.14			
				4 3102	3, 5	-2.83			
							4 3102	4, 5	0.31
		5 3104	1, 2						
		5 3104	1, 3						
		5 3104	1, 4						
		5 3104	1, 5						
				5 3104	2, 3	2.66			
				5 3104	2, 4	-0.89			
							5 3104	2, 5	-0.52
				5 3104	3, 4	-3.56			
				5 3104	3, 5	-3.19			
							5 3104	4, 5	0.37
							6 3201	1, 2	-0.69
							6 3201	1, 3	0.05
							6 3201	1, 4	-0.34
				6 3201	1, 5	-1.15			
				6 3201	2, 3	0.73			
							6 3201	2, 4	0.35
							6 3201	2, 5	-0.46
							6 3201	3, 4	-0.39
				6 3201	3, 5	-1.20			
				6 3201	4, 5	-0.81			
							7 3301	1, 2	-0.12
				7 3301	1, 4	-0.99	7 3301	1, 3	-0.20
				7 3301	1, 5	-0.88			
							7 3301	2, 3	-0.09
				7 3301	2, 4	-0.87			
				7 3301	2, 5	-0.76			
				7 3301	3, 4	-0.79			
							7 3301	3, 5	-0.68
							7 3301	4, 5	0.11
				8 4101	2, 3	1.72			
				8 4101	2, 4	2.46			
				8 4101	3, 4	0.73			
							9 4102	1, 2	0.20
				9 4102	1, 3	0.99			
							9 4102	1, 4	-0.29
				9 4102	2, 3	0.78			
							9 4102	2, 4	-0.49
				9 4102	3, 4	-1.28			
							10 4103	1, 2	-0.69
							10 4103	1, 3	-0.13
							10 4103	1, 4	-0.56
				10 4103	1, 5	-1.23			
							10 4103	2, 3	0.55
							10 4103	2, 4	0.13
							10 4103	2, 5	-0.54
							10 4103	3, 4	-0.43
				10 4103	3, 5	-1.10			
							10 4103	4, 5	-0.67
				11 4104	1, 2	-0.78			
				11 4104	1, 3	-1.74			
				11 4104	1, 4	-1.34			
				11 4104	1, 5	-1.47			
				11 4104	2, 3	-0.97			
							11 4104	2, 4	-0.56
							11 4104	2, 5	-0.69
							11 4104	3, 4	0.40
							11 4104	3, 5	0.27
							11 4104	4, 5	-0.13
							12 4202	1, 2	-0.10
							12 4202	1, 3	0.49
							12 4202	1, 4	-0.27
							12 4202	1, 5	-0.39
							12 4202	2, 3	0.58
							12 4202	2, 4	-0.17
							12 4202	2, 5	-0.29
				12 4202	3, 4	-0.76			
				12 4202	3, 5	-0.88			
							12 4202	4, 5	-0.12
							13 4203	2, 3	0.20
							13 4203	2, 4	0.05
							13 4203	3, 4	-0.16
				14 4301	2, 3	0.70			
				14 4301	2, 4	0.90			
							14 4301	2, 5	-0.18
							14 4301	3, 4	0.19
				14 4301	3, 5	-0.89			
				14 4301	4, 5	-1.08			
				15 4401	2, 3	0.72			
							15 4401	2, 4	-0.13
				15 4401	3, 4	-0.86			
							16 4402	3, 4	-0.21
				17 5101	2, 3	-1.14			
				17 5101	2, 4	1.57			
				17 5101	3, 4	2.70			
				18 5102	2, 3	-1.95			
							18 5102	2, 4	0.27
				18 5102	3, 4	2.21			
				19 5103	1, 2	-0.82			
							19 5103	1, 3	-0.19

								19	5103	1, 4	-0.93			
												19	5103	2, 3	0.62
												19	5103	2, 4	-0.11
								19	5103	3, 4	-0.74				
								20	5104	1, 2	1.13				
								20	5104	1, 3	0.89				
								20	5104	1, 4	0.94				
								20	5104	1, 5	0.93				
												20	5104	2, 3	-0.25
												20	5104	2, 4	-0.19
												20	5104	2, 5	-0.20
												20	5104	3, 4	0.05
												20	5104	3, 5	0.04
												20	5104	4, 5	-0.01
												21	5105	1, 2	0.19
												21	5105	1, 3	0.11
												21	5105	1, 4	-0.14
												21	5105	1, 5	0.67
												21	5105	2, 3	-0.09
												21	5105	2, 4	-0.33
												21	5105	2, 5	0.48
												21	5105	3, 4	-0.25
												21	5105	3, 5	0.56
								21	5105	4, 5	0.81				
												22	5106	3, 4	0.55
								23	5109	3, 4	1.65				
												24	5110	3, 4	0.54
								24	5110	3, 5	1.16				
												24	5110	4, 5	0.62
								25	5201	3, 4	0.72				
								25	5201	3, 5	1.03				
												25	5201	4, 5	0.31
								26	5202	3, 4	1.55				
								26	5202	3, 5	1.82				
												26	5202	4, 5	0.27
								27	6101	1, 2	1.55				
								27	6101	1, 3	1.76				
								27	6101	1, 4	2.05				
												27	6101	2, 3	0.20
												27	6101	2, 4	0.50
												27	6101	3, 4	0.29
												28	6102	1, 2	0.67
												28	6102	1, 3	-0.12
												28	6102	1, 4	0.46
								28	6102	1, 5	0.85				
								28	6102	2, 3	-0.80				
												28	6102	2, 4	-0.21
												28	6102	2, 5	0.18
												28	6102	3, 4	0.58
								28	6102	3, 5	0.97				
												28	6102	4, 5	0.39
								29	6103	1, 2	1.55				
								29	6103	1, 3	1.23				
								29	6103	1, 4	1.61				
								29	6103	1, 5	1.84				
												29	6103	2, 3	-0.33
												29	6103	2, 4	0.06
												29	6103	2, 5	0.29
												29	6103	3, 4	0.38
												29	6103	3, 5	0.61
												29	6103	4, 5	0.23
				30	6104	3, 1									
				30	6104	3, 2									
				30	6104	3, 4									
				30	6104	3, 5									
												30	6104	1, 2	0.32
												30	6104	1, 4	-0.17
								30	6104	1, 5	0.73				
												30	6104	2, 4	-0.49
												30	6104	2, 5	0.41
								30	6104	4, 5	0.90				
								31	6108	3, 4	1.59				
								31	6108	3, 5	2.01				
												31	6108	4, 5	0.42
								32	6109	3, 4	2.26				
33	6110	3	KCTC												
34	6111	4	KCTD												
				35	6401	3, 4									
				35	6401	3, 5									
								35	6401	4, 5	0.75				
				36	7104	3, 2									
				36	7104	3, 4									
				36	7104	3, 5									
								36	7104	2, 4	1.57				
								36	7104	2, 5	1.74				
												36	7104	4, 5	0.17
				37	7105	2, 3									
				37	7105	2, 4									
				37	7105	2, 5									
				37	7105	3, 4									
				37	7105	3, 5									
								37	7105	4, 5	0.74				
				38	7201	3, 2									
				38	7201	3, 4									
				38	7201	3, 5									
								38	7201	2, 4	1.58				
								38	7201	2, 5	1.18				
												38	7201	4, 5	-0.40
				39	7301	3, 2									
				39	7301	3, 4									
				39	7301	3, 5									
												39	7301	2, 4	-0.26
												39	7301	2, 5	0.25
												39	7301	4, 5	0.51
				40	7402	1, 2									
				40	7402	1, 3									
				40	7402	1, 4									
				40	7402	1, 5									
				40	7402	3, 2									
				40	7402	3, 4									
				40	7402	3, 5									
												40	7402	2, 4	0.53
								40	7402	2, 5	0.77				
												40	7402	4, 5	0.24
41	8103	5	KCTE												

FORM NAME	TRANSLATION CONSTANT	TRANSLATION S.E.	FORM FIT STATISTIC
KCTA	0.00	0.07	-0.30

AFTER ITEM DELETIONS

LINKED TO FORM	NO. OF ITEMS	OBSERVED SHIFT	SHIFT S.E.	LOGIT LINK RESIDUAL	STDZD LINK RESIDUAL	LINK MISFIT BTWN FORM	LINK MISFIT WTHN FORM*
KCTB	9.	-0.03	0.10	0.01	0.13	-2.34	-0.60
KCTC	7.	-0.44	0.11	0.04	0.34	-2.39	0.61
KCTD	8.	-0.21	0.11	-0.13	-1.25	-3.07	-0.46
KCTE	3.	0.66	0.24	0.08	0.35	0.02	0.64

* SIGNIFIES OVERFIT AS WELL AS UNDERFIT. CHECK FORCAL OUTPUT FOR DETAILS.

FORM NAME	TRANSLATION CONSTANT	TRANSLATION S.E.	FORM FIT STATISTIC
KCTB	-0.04	0.03	3.69

AFTER ITEM DELETIONS

LINKED TO FORM	NO. OF ITEMS	OBSERVED SHIFT	SHIFT S.E.	LOGIT LINK RESIDUAL	STDZD LINK RESIDUAL	LINK MISFIT BTWN FORM	LINK MISFIT WTHN FORM*
KCTA	9.	0.03	0.10	-0.01	-0.13	-2.34	-0.60
KCTC	9.	-0.26	0.05	0.17	3.30	-0.81	1.55
KCTD	19.	-0.06	0.03	-0.02	-0.61	-3.39	2.47
KCTE	14.	0.48	0.04	-0.14	-3.21	-1.93	-0.61

* SIGNIFIES OVERFIT AS WELL AS UNDERFIT. CHECK FORCAL OUTPUT FOR DETAILS.

FORM NAME	TRANSLATION CONSTANT	TRANSLATION S.E.	FORM FIT STATISTIC
KCTC	-0.47	0.04	5.50

AFTER ITEM DELETIONS

LINKED TO FORM	NO. OF ITEMS	OBSERVED SHIFT	SHIFT S.E.	LOGIT LINK RESIDUAL	STDZD LINK RESIDUAL	LINK MISFIT BTWN FORM	LINK MISFIT WTHN FORM*
KCTA	7.	0.44	0.11	-0.04	-0.34	-2.39	0.61
KCTB	9.	0.26	0.05	-0.17	-3.30	-0.81	1.55
KCTD	13.	0.58	0.03	0.19	5.62	-0.77	2.01
KCTE	5.	1.08	0.07	0.02	0.34	1.43	1.01

* SIGNIFIES OVERFIT AS WELL AS UNDERFIT. CHECK FORCAL OUTPUT FOR DETAILS.

FORM NAME	TRANSLATION CONSTANT	TRANSLATION S.E.	FORM FIT STATISTIC
KCTD	-0.08	0.03	4.87

AFTER ITEM DELETIONS

LINKED TO FORM	NO. OF ITEMS	OBSERVED SHIFT	SHIFT S.E.	LOGIT LINK RESIDUAL	STDZD LINK RESIDUAL	LINK MISFIT BTWN FORM	LINK MISFIT WTHN FORM*
KCTA	8.	0.21	0.11	0.13	1.25	-3.07	-0.46
KCTB	19.	0.06	0.03	0.02	0.61	-3.39	2.47
KCTC	13.	-0.58	0.03	-0.19	-5.62	-0.77	2.01
KCTE	19.	0.69	0.04	0.03	0.89	-3.59	-2.09

* SIGNIFIES OVERFIT AS WELL AS UNDERFIT. CHECK FORCAL OUTPUT FOR DETAILS.

FORM NAME	TRANSLATION CONSTANT	TRANSLATION S.E.	FORM FIT STATISTIC
KCTE	0.58	0.06	2.31

AFTER ITEM DELETIONS

LINKED TO FORM	NO. OF ITEMS	OBSERVED SHIFT	SHIFT S.E.	LOGIT LINK RESIDUAL	STDZD LINK RESIDUAL	LINK MISFIT BTWN FORM	LINK MISFIT WTHN FORM*
KCTA	3.	-0.66	0.24	-0.08	-0.35	0.02	0.64
KCTB	14.	-0.48	0.04	0.14	3.21	-1.93	-0.61
KCTC	5.	-1.08	0.07	-0.02	-0.34	1.43	1.01
KCTD	19.	-0.69	0.04	-0.03	-0.89	-3.59	-2.09

* SIGNIFIES OVERFIT AS WELL AS UNDERFIT. CHECK FORCAL OUTPUT FOR DETAILS.

THIS ENDS THE LINKS WHICH HAVE BEEN CALIBRATED AFTER ITEM DELETIONS.

FOLLOWING IS THE LINKS WHICH HAVE BEEN CALIBRATED BEFORE ITEM DELETIONS.

FORM NAME	TRANSLATION CONSTANT	TRANSLATION S.E.	FORM FIT STATISTIC
KCTA	-0.00	0.03	4.32

BEFORE ITEM DELETIONS

LINKED TO FORM	NO. OF ITEMS	OBSERVED SHIFT	SHIFT S.E.	LOGIT LINK RESIDUAL	STDZD LINK RESIDUAL	LINK MISFIT BTWN FORM	LINK MISFIT WTHN FORM*
KCTB	14.	0.01	0.06	0.06	0.89	-0.79	-0.51
KCTC	13.	-0.23	0.07	0.25	3.68	0.62	0.80
KCTD	14.	-0.21	0.06	-0.21	-3.48	-0.06	0.19
KCTE	11.	0.44	0.08	-0.09	-1.17	0.98	-0.56

* SIGNIFIES OVERFIT AS WELL AS UNDERFIT. CHECK FORCAL OUTPUT FOR DETAILS.

FORM NAME	TRANSLATION CONSTANT	TRANSLATION S.E.	FORM FIT STATISTIC
KCTB	-0.04	0.02	3.15

BEFORE ITEM DELETIONS

LINKED TO FORM	NO. OF ITEMS	OBSERVED SHIFT	SHIFT S.E.	LOGIT LINK RESIDUAL	STDZD LINK RESIDUAL	LINK MISFIT BTWN FORM	LINK MISFIT WTHN FORM*
KCTA	14.	-0.01	0.06	-0.06	-0.89	-0.79	-0.51
KCTC	20.	-0.37	0.03	0.06	2.01	2.95	5.84
KCTD	26.	0.12	0.03	0.07	2.81	-0.27	1.96
KCTE	18.	0.49	0.04	-0.08	-2.07	-1.79	-1.33

* SIGNIFIES OVERFIT AS WELL AS UNDERFIT. CHECK FORCAL OUTPUT FOR DETAILS.

FORM NAME	TRANSLATION CONSTANT	TRANSLATION S.E.	FORM FIT STATISTIC
KCTC	-0.48	0.02	12.18

BEFORE ITEM DELETIONS

LINKED TO FORM	NO. OF ITEMS	OBSERVED SHIFT	SHIFT S.E.	LOGIT LINK RESIDUAL	STDZD LINK RESIDUAL	LINK MISFIT BTWN FORM	LINK MISFIT WTHN FORM*
KCTA	13.	0.23	0.07	-0.25	-3.68	0.62	0.80
KCTB	20.	0.37	0.03	-0.06	-2.01	2.95	5.84
KCTD	28.	0.77	0.02	0.29	14.52	6.75	7.57
KCTE	16.	1.03	0.04	0.02	0.51	4.40	2.51

* SIGNIFIES OVERFIT AS WELL AS UNDERFIT. CHECK FORCAL OUTPUT FOR DETAILS.

FORM NAME	TRANSLATION CONSTANT	TRANSLATION S.E.	FORM FIT STATISTIC
KCTD	-0.00	0.02	12.98

BEFORE ITEM DELETIONS

LINKED TO FORM	NO. OF ITEMS	OBSERVED SHIFT	SHIFT S.E.	LOGIT LINK RESIDUAL	STDZD LINK RESIDUAL	LINK MISFIT BTWN FORM	LINK MISFIT WTHN FORM*
KCTA	14.	0.21	0.06	0.21	3.48	-0.06	0.19
KCTB	26.	-0.12	0.03	-0.07	-2.81	-0.27	1.96
KCTC	28.	-0.77	0.02	-0.29	-14.52	6.75	7.57
KCTE	25.	0.68	0.03	0.15	5.94	-1.21	-1.71

* SIGNIFIES OVERFIT AS WELL AS UNDERFIT. CHECK FORCAL OUTPUT FOR DETAILS.

FORM NAME	TRANSLATION CONSTANT	TRANSLATION S.E.	FORM FIT STATISTIC
KCTE	0.53	0.02	5.39

BEFORE ITEM DELETIONS

LINKED TO FORM	NO. OF ITEMS	OBSERVED SHIFT	SHIFT S.E.	LOGIT LINK RESIDUAL	STDZD LINK RESIDUAL	LINK MISFIT BTWN FORM	LINK MISFIT WTHN FORM*
KCTA	11.	-0.44	0.08	0.09	1.17	0.98	-0.56
KCTB	18.	-0.49	0.04	0.08	2.07	-1.79	-1.33
KCTC	16.	-1.03	0.04	-0.02	-0.51	4.40	2.51
KCTD	25.	-0.68	0.03	-0.15	-5.94	-1.21	-1.71

* SIGNIFIES OVERFIT AS WELL AS UNDERFIT. CHECK FORCAL OUTPUT FOR DETAILS.

In this analysis, form KCTC with a form difficulty (translation constant) of -0.48 and standard error 0.02, is the easiest form and form KCTE with a form difficulty (translation constant) of 0.53 and standard error 0.02 is the hardest form. After item link deletions, form KCTC remains the easiest form with a form difficulty of -0.47 and standard error 0.04, and form KCTE is still the hardest form, with a form difficulty of 0.58 and standard error 0.06.

Before deletions all five forms have very large form fit statistics. In particular, the fits of forms KCTC and KCTD, with 12.18 and 12.98, respectively, suggest that some items are causing severe misfit of these forms into the bank.

Before link deletions both the between form fit and within form fit are large for all forms except KCTA, with form KCTC exhibiting the most misfit. After deleting the links with large residuals, the within form fit statistic decreases drastically for all forms. Deleting the links with large residuals has removed the within form misfit.

The between form fit, after item link editing, confounds over and under fit and is rarely useful in judging the editing. This statistic decreased for three out of the five forms, increasing from -0.66 to 3.33 in KCTA and from 2.22

to 3.41 in KCTB. Form KCTD now exhibits the greatest between form misfit at 4.68.

The small differences in translation constants demonstrate that the presence of disturbances had little effect on the estimation of form difficulties. The number of links between forms remains at four for all forms indicating that an entire link between two forms was not eliminated.

The third page of the SHIFT output lists the items and item links not used in the linking structure, the item links dropped from the design due to a large residual, and the item links used in the final link structure. The items in the first panel of the page are not used in the link design either before or after the item link deletions, because they only appear in one form. This was true for four of the 41 original items: #2, #33, #34, and #41.

The next panel shows the items not used in the link analysis, either before or after item link deletions even though they appeared in more than one form. This occurs when the item was not calibrated in either one or both of the forms in which it appears because everyone either passed or failed the item. These items are listed by their name and link affected, that is, the two forms in which this item appears. Of the 228 possible links in this example, 43 were not used for this reason.

The third panel on Page 3 shows the item links used in the SHIFT analysis before deletions, but which are subsequently eliminated from the link design because their logit residual is larger than the specified value. This list gives the item names, the item links affected, and the residuals. Seventy-nine item links were dropped from the analysis due to a large residual.

The last panel on this page shows the item links that were used in the link design, both before and after item link deletions. These item links are listed by item name, link affected, and residual. There were 106 item links remaining in the analysis after the item link deletions.

Pages 4 through 8 show the calibration of the forms by links after the deletion of item links with large residuals. A table is printed for each form listing the form name and number, the translation constant and standard error and the form fit statistic. Then for every form to which this form is linked, the table shows the number of items linking the two forms, the observed shift and standard error between the two forms, the logit link residual and standardized link residual, the link misfit between the two forms, and the link misfit within the form. Pages 9 through 13 show the same tables for the before item link deletions analysis.

5.4 ITEMLIST

After the items have been calibrated and the persons measured, the program, ITEMLIST, enables examination of all the items in the bank. The only input to the program is a title which is printed at the top of every page of output.

The first section of ITEMLIST lists the items in the bank sorted by sequence number. Each item is listed by its sequence number, legitimate alternatives, correct response, item name, final bank difficulty, status, between difficulty root mean square, and within fit mean square. Final bank difficulty is an average of the item's difficulties within the forms in which it was calibrated, adjusted for form difficulty. An item's status will always be "9" unless the item was not calibrated in any of the forms in which it was placed, in which case it will be "0."

The between difficulty root mean square (RMS) is formed by summing the squared difference between an item's bank equated difficulty within each form and its final averaged bank difficulty over all calibrations of the item, and taking the square root of the average of this sum of squares. For example, the between difficulty root mean square of Item #1 on Page 1 of the ITEMLIST output is

$$[[(-8.10 - (-7.81))^2 + (-7.57 - (-7.81))^2]/2]^{**}1/2 = 0.27.$$

ITEM NUMBER	LEGITIMATE ANSWERS	KEY	ITEM NAME	FORM NAME	FORM NUMBER	ITEMS IN FORM	ITEM POS	ITEM STATS IN FORM DIFF	ERROR	FIT T	BANK DIFF	STATUS	BETWEEN DIFF RMS	WITHIN FIT MS	ITEM NMBR
1	01	1	2101								-7.81	9	0.27	0.56	1
				KCTA	1	18	1	0.0				0			
				KCTB	2	28	1	0.0				0			
				KCTC	3	38	1	0.0				0			
				KCTD	4	38	1	-8.10	***	0.8		9			
				KCTE	5	26	1	-7.57	***	0.7		9			
2	01	1	2102								0.0	9	0.0	0.0	2
				KCTA	1	18	2	0.0				0			
3	01	1	2201								-6.85	9	0.23	1.27	3
				KCTB	2	28	2	-6.69	.88	1.3		9			
				KCTC	3	38	2	0.0				0			
				KCTD	4	38	2	-7.22	***	1.4		9			
				KCTE	5	26	2	-6.82	.81	0.4		9			
4	01	1	3102								-4.40	9	1.22*	0.75	4
				KCTA	1	18	4	-4.19	.87	0.4		9			
				KCTB	2	28	4	-4.47	.48	0.3		9			
				KCTC	3	38	4	-7.09	.83	1.5		9			
				KCTD	4	38	4	-4.04	.39	-0.2		9			
				KCTE	5	26	4	-4.22	.35	1.1		9			
5	01	1	3104								-5.94	9	1.54*	0.71	5
				KCTA	1	18	3	0.0				0			
				KCTB	2	28	3	-6.29	.74	1.1		9			
				KCTC	3	38	3	-8.95	***	0.9		9			
				KCTD	4	38	3	-5.48	.55	-0.1		9			
				KCTE	5	26	3	-5.72	.59	0.9		9			
6	01	1	3201								-3.06	9	0.50	0.74	6
				KCTA	1	18	5	-3.65	.79	0.7		9			
				KCTB	2	28	5	-2.96	.44	1.5		9			
				KCTC	3	38	5	-3.69	.26	-0.4		9			
				KCTD	4	38	6	-3.39	.34	0.1		9			
				KCTE	5	26	5	-2.45	.22	0.9		9			
7	01	1	3301								-2.60	9	0.44	0.38	7
				KCTA	1	18	6	-3.22	.75	0.9		9			
				KCTB	2	28	6	-3.10	.43	0.9		9			
				KCTC	3	38	6	-3.01	.23	-0.1		9			
				KCTD	4	38	10	-2.31	.28	0.5		9			
				KCTE	5	26	6	-2.29	.20	0.0		9			
8	01	1	4101								-5.69	9	1.14*	-1.68	8
				KCTB	2	28	7	-4.68	.47	0.0		9			
				KCTC	3	38	9	-6.40	.55	-2.1		9			
				KCTD	4	38	5	-7.22	.84	-0.8		9			
9	01	1	4102								-2.60	9	0.46	1.21	9
				KCTA	1	18	8	-2.24	.55	-1.0		9			
				KCTB	2	28	9	-2.44	.37	0.4		9			
				KCTC	3	38	11	-3.22	.26	1.5		9			
				KCTD	4	38	9	-2.03	.28	1.2		9			
10	01	1	4103								-2.88	9	0.52*	1.33	10
				KCTA	1	18	7	-3.65	.88	1.2		9			
				KCTB	2	28	8	-2.98	.42	1.0		9			
				KCTC	3	38	8	-3.51	.27	1.4		9			
				KCTD	4	38	8	-3.17	.37	1.5		9			
				KCTE	5	26	7	-2.37	.20	0.0		9			
11	01	1	4104								-1.82	9	0.71*	1.40	11
				KCTA	1	18	9	-3.22	.65	-0.5		9			
				KCTB	2	28	10	-2.44	.38	0.7		9			
				KCTC	3	38	13	-1.47	.25	2.5		9			
				KCTD	4	38	14	-1.96	.26	0.0		9			
				KCTE	5	26	8	-1.70	.18	0.0		9			
12	01	1	4202								-1.35	9	0.32	2.44*	12
				KCTA	1	18	10	-1.50	.53	0.7		9			
				KCTB	2	28	11	-1.40	.35	1.5		9			
				KCTC	3	38	10	-1.98	.25	2.8		9			
				KCTD	4	38	11	-1.31	.24	-1.0		9			
				KCTE	5	26	9	-1.06	.17	0.8		9			
13	01	1	4203								-1.35	9	0.08	2.05*	13
				KCTB	2	28	13	-1.23	.30	-0.7		9			
				KCTC	3	38	14	-1.43	.24	2.2		9			
				KCTD	4	38	13	-1.36	.24	-0.9		9			
14	01	1	4301								-0.67	9	0.53*	2.14*	14
				KCTB	2	28	12	-0.49	.28	-0.2		9			
				KCTC	3	38	12	-1.19	.25	2.2		9			
				KCTD	4	38	15	-1.47	.27	1.7		9			
				KCTE	5	26	10	-0.26	.15	0.9		9			
15	01	1	4401								-1.20	9	0.36	4.71*	15
				KCTB	2	28	14	-0.89	.29	-1.8		9			
				KCTC	3	38	15	-1.61	.22	-0.1		9			
				KCTD	4	38	12	-0.84	.27	3.3		9			
16	01	1	4402								-3.46	9	0.06	6.02*	16
				KCTC	3	38	7	-3.51	.30	3.4		9			
				KCTD	4	38	7	-3.39	.36	0.7		9			
17	01	1	5101								-1.12	9	1.16*	-10.17*	17
				KCTB	2	28	16	-1.14	.30	-0.3		9			
				KCTC	3	38	20	0.0	.25	-5.1		9			
				KCTD	4	38	16	-2.79	.30	-2.1		9			
18	01	1	5102								-1.19	9	1.03*	-11.35*	18
				KCTB	2	28	15	-1.89	.33	-3.0		9			
				KCTC	3	38	19	0.06	.25	-5.0		9			
				KCTD	4	38	17	-2.24	.27	-0.2		9			
19	01	1	5103								1.29	9	0.38	4.26*	19
				KCTA	1	18	11	0.76	.49	0.8		9			
				KCTB	2	28	17	1.58	.35	2.0		9			
				KCTC	3	38	16	0.96	.27	-0.4		9			
				KCTD	4	38	19	1.61	.27	3.5		9			
20	01	1	5104								1.22	9	0.43	0.70	20
				KCTA	1	18	12	2.14	.63	1.0		9			
				KCTB	2	28	18	1.01	.29	-0.2		9			
				KCTC	3	38	17	1.26	.29	0.4		9			
				KCTD	4	38	21	1.12	.22	-0.6		9			
				KCTE	5	26	13	1.26	.16	1.4		9			
21	01	1	5105								1.54	9	0.28	1.08	21
				KCTA	1	18	13	1.86	.53	-0.6		9			
				KCTB	2	28	19	1.67	.31	0.3		9			
				KCTC	3	38	18	1.76	.31	-0.3		9			
				KCTD	4	38	24	1.92	.23	-0.1		9			
				KCTE	5	26	14	1.24	.16	2.2		9			
22	01	1	5108								1.33	9	0.34	-1.49	22
				KCTC	3	38	21	1.76	.31	0.3		9			
				KCTD	4	38	23	1.12	.22	-1.7		9			

ITEM NUMBER	LEGITIMATE ANSWERS	KEY	ITEM NAME	FORM NAME	FORM NUMBER	ITEMS IN FORM	ITEM POS	ITEM STATS IN FORM: DIFF	ERROR	FIT T	BANK DIFF	STATUS	BETWEEN DIFF RMS	WITHIN FIT MS	ITEM NMBR
23	01	1	5109								0.52	9	0.91*	-3.80*	23
				KCTC	3	38	25	1.67	.30	-1.9		9			
				KCTD	4	38	22	-0.07	.22	-2.0		9			
24	01	1	5110								1.79	9	0.54*	-0.03	24
				KCTC	3	38	24	2.66	.38	0.0		9			
				KCTD	4	38	25	2.03	.24	-0.2		9			
				KCTE	5	26	15	1.54	.15	-0.2		9			
25	01	1	5201								0.19	9	0.46	-1.22	25
				KCTC	3	38	22	0.96	.27	-1.6		9			
				KCTD	4	38	20	0.15	.22	-1.0		9			
				KCTE	5	26	12	-0.03	.15	0.3		9			
26	01	1	5202								-0.83	9	0.83*	-6.46*	26
				KCTC	3	38	23	0.49	.26	-3.9		9			
				KCTD	4	38	18	-1.15	.26	2.0		9			
				KCTE	5	26	11	-1.29	.17	0.4		9			
27	01	1	6101								2.71	9	0.95*	-0.90	27
				KCTA	1	18	16	4.56	***	0.3		9			
				KCTB	2	28	22	3.01	.36	-1.5		9			
				KCTC	3	38	28	2.81	.40	-1.1		9			
				KCTD	4	38	26	2.43	.25	0.2		9			
28	01	1	6102								2.57	9	0.46	0.79	28
				KCTA	1	18	14	3.21	.94	1.5		9			
				KCTB	2	28	20	2.54	.35	0.5		9			
				KCTC	3	38	26	3.34	.46	-0.1		9			
				KCTD	4	38	28	2.67	.28	1.2		9			
				KCTE	5	26	16	2.41	.17	0.0		9			
29	01	1	6103								2.90	9	0.77*	-0.39	29
				KCTA	1	18	15	4.56	***	0.3		9			
				KCTB	2	28	21	3.01	.36	-0.9		9			
				KCTC	3	38	27	3.34	.46	-0.2		9			
				KCTD	4	38	29	2.87	.26	0.0		9			
				KCTE	5	26	17	2.77	.18	-1.0		9			
30	01	1	6104								4.14	9	0.36	-0.17	30
				KCTA	1	18	17	4.56	***	0.3		9			
				KCTB	2	28	23	4.24	.46	0.3		9			
				KCTC	3	38	29	0.0				0			
				KCTD	4	38	33	4.65	.38	-0.7		9			
				KCTE	5	26	20	3.88	.24	-0.1		9			
31	01	1	6108								2.82	9	1.05*	-0.35	31
				KCTC	3	38	30	4.62	.74	-0.2		9			
				KCTD	4	38	30	2.94	.27	-0.1		9			
				KCTE	5	26	18	2.65	.18	-1.0		9			
32	01	1	6109								1.26	9	1.37*	-3.12*	32
				KCTC	3	38	31	3.14	.44	-0.7		9			
				KCTD	4	38	27	0.79	.22	-2.4		9			
33	01	1	6110								4.62	9	0.0	0.0	33
				KCTC	3	38	32	4.62	.74	0.0		9			
34	01	1	6111								1.46	9	0.0	-6.76*	34
				KCTD	4	38	31	1.46	.23	-2.6		9			
35	01	1	6401								3.69	9	0.34	-1.10	35
				KCTC	3	38	33	0.0				0			
				KCTD	4	38	32	4.15	.37	1.0		9			
				KCTE	5	26	19	3.53	.22	-1.1		9			
36	01	1	7104								5.44	9	0.90*	-0.11	36
				KCTB	2	28	27	6.99	***	0.0		9			
				KCTC	3	38	37	0.0				0			
				KCTD	4	38	38	5.34	.46	-0.4		9			
				KCTE	5	26	26	5.30	.38	-0.4		9			
37	01	1	7105								4.92	9	0.33	0.05	37
				KCTB	2	28	28	0.0				0			
				KCTC	3	38	38	0.0				0			
				KCTD	4	38	35	5.34	.47	0.3		9			
				KCTE	5	26	22	4.73	.31	0.1		9			
38	01	1	7201								4.91	9	0.77*	-0.25	38
				KCTB	2	28	26	6.17	.80	-0.2		9			
				KCTC	3	38	36	0.0				0			
				KCTD	4	38	37	4.51	.37	-0.6		9			
				KCTE	5	26	21	5.04	.37	0.6		9			
39	01	1	7301								5.60	9	0.16	0.18	39
				KCTB	2	28	25	5.65	.77	0.7		9			
				KCTC	3	38	35	0.0				0			
				KCTD	4	38	36	5.83	.54	0.1		9			
				KCTE	5	26	23	5.45	.41	0.2		9			
40	01	1	7402								5.56	9	0.36	0.40	40
				KCTA	1	18	18	0.0				0			
				KCTB	2	28	24	6.17	.97	0.9		9			
				KCTC	3	38	34	0.0				0			
				KCTD	4	38	34	5.56	.54	0.6		9			
				KCTE	5	26	23	5.45	.41	0.2		9			
41	01	1	8103								5.62	9	0.0	0.0	41
				KCTE	5	26	26	5.62	.43	0.0		9			

MEAN	0.28
S.D.	3.60
MIN.	-7.81
MAX.	5.62

ITEM LIST SORTED BY DIFFICULTY RMS

ITEM NUMBER	LEGITIMATE ANSWERS	KEY	ITEM NAME	FORM NAME	FORM NUMBER	ITEMS IN FORM	ITEM POS	ITEM STATS IN FORM: DIFF	ERROR	FIT T	BANK DIFF	STATUS	BETWEEN DIFF RMS	WITHIN FIT MS	ITEM NMBR
5	01	1	3104								-5.94	9	1.54	0.71	5
				KCTA	1	18	3	0.0				0			
				KCTB	2	28	3	-6.29	.74	1.1		9			
				KCTC	3	38	3	-8.95	***	0.9		9			
				KCTD	4	38	3	-5.48	.55	-0.1		9			
				KCTE	5	26	3	-5.72	.59	0.9		9			
32	01	1	6109								1.26	9	1.37	-3.12*	32
				KCTC	3	38	31	3.14	.44	-0.7		9			
				KCTD	4	38	27	0.79	.22	-2.4		9			
4	01	1	3102								-4.40	9	1.22	0.75	4
				KCTA	1	18	4	-4.19	.87	0.4		9			
				KCTB	2	28	4	-4.47	.48	0.3		9			
				KCTC	3	38	4	-7.09	.83	1.5		9			
				KCTD	4	38	4	-4.04	.39	-0.2		9			
				KCTE	5	26	4	-4.22	.35	1.1		9			
17	01	1	5101								-1.12	9	1.16	-10.17*	17
				KCTB	2	28	16	-1.14	.30	-0.3		9			
				KCTC	3	38	20	0.0	.25	-5.1		9			
				KCTD	4	38	16	-2.79	.30	-2.1		9			
8	01	1	4101								-5.69	9	1.14	-1.68	8
				KCTB	2	28	7	-4.68	.47	0.0		9			
				KCTC	3	38	9	-6.40	.55	-2.1		9			
				KCTD	4	38	5	-7.22	.84	-0.8		9			
31	01	1	6108								2.82	9	1.05	-0.35	31
				KCTC	3	38	30	4.62	.74	-0.2		9			
				KCTD	4	38	30	2.94	.27	-0.1		9			
				KCTE	5	26	18	2.65	.18	-1.0		9			
18	01	1	5102								-1.19	9	1.03	-11.35*	18
				KCTB	2	28	15	-1.89	.33	-3.0		9			
				KCTC	3	38	19	0.06	.25	-5.0		9			
				KCTD	4	38	17	-2.24	.27	-0.2		9			
27	01	1	6101								2.71	9	0.95	-0.90	27
				KCTA	1	18	16	4.56	***	0.3		9			
				KCTB	2	28	22	3.01	.36	-1.5		9			
				KCTC	3	38	28	2.81	.40	-1.1		9			
				KCTD	4	38	26	2.43	.25	0.2		9			
23	01	1	5109								0.52	9	0.91	-3.80*	23
				KCTC	3	38	25	1.67	.30	-1.9		9			
				KCTD	4	38	22	-0.07	.22	-2.0		9			
36	01	1	7104								5.44	9	0.90	-0.11	36
				KCTB	2	28	27	6.99	***	0.0		9			
				KCTC	3	38	37	0.0				0			
				KCTD	4	38	38	5.34	.46	-0.4		9			
				KCTE	5	26	25	5.30	.38	-0.4		9			
26	01	1	5202								-0.83	9	0.83	-6.46*	26
				KCTC	3	38	23	0.49	.26	-3.9		9			
				KCTD	4	38	18	-1.15	.26	2.0		9			
				KCTE	5	26	11	-1.29	.17	0.4		9			
29	01	1	6103								2.90	9	0.77	-0.39	29
				KCTA	1	18	15	4.56	***	0.3		9			
				KCTB	2	28	21	3.01	.36	-0.9		9			
				KCTC	3	38	27	3.34	.46	-0.2		9			
				KCTD	4	38	29	2.87	.26	0.0		9			
				KCTE	5	26	17	2.77	.18	-1.0		9			

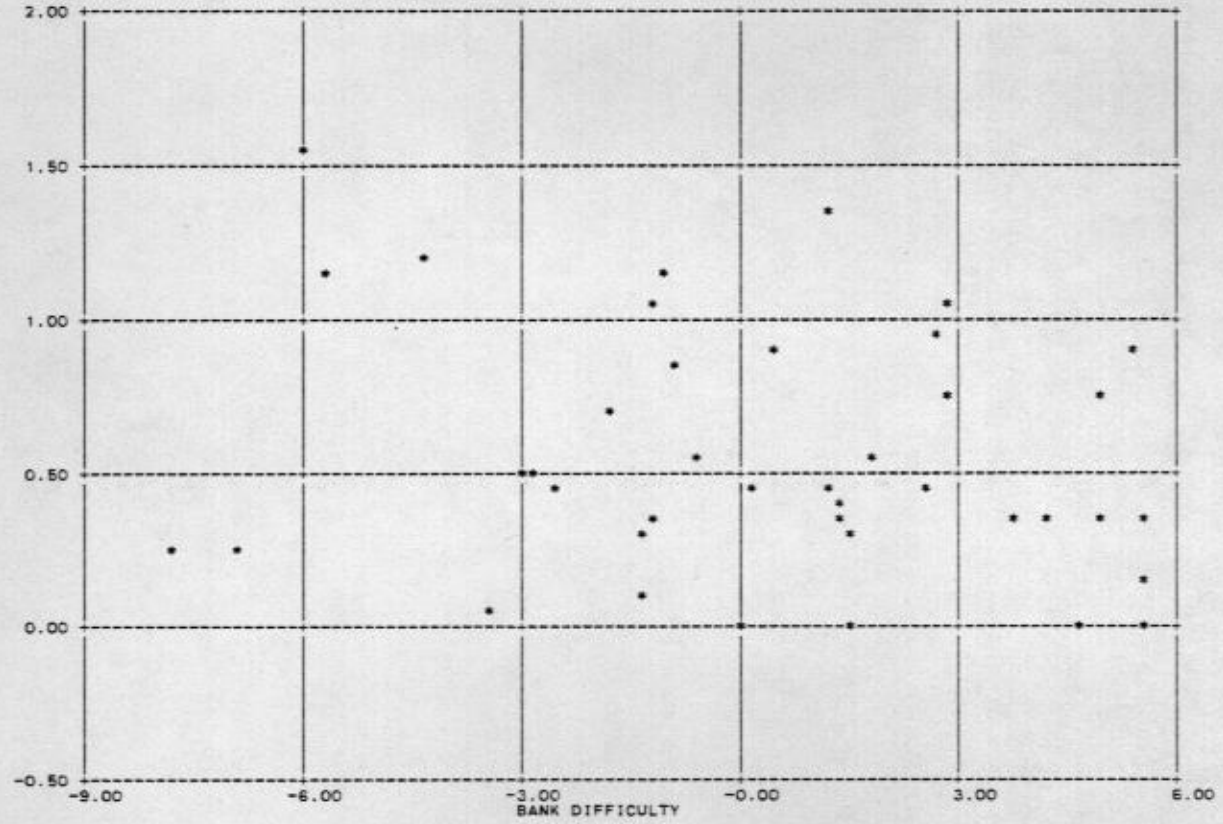
ITEM LIST SORTED BY DIFFICULTY RMS

ITEM NUMBER	LEGITIMATE ANSWERS	KEY	ITEM NAME	FORM NAME	FORM NUMBER	ITEMS IN FORM	ITEM POS	ITEM STATS IN FORM: DIFF	ERROR	FIT T	BANK DIFF	STATUS	BETWEEN DIFF RMS	WITHIN FIT MS	ITEM NMBR
38	01	1	7201								4.91	9	0.77	-0.25	38
				KCTB	2	28	26	6.17	.80	-0.2		9			
				KCTC	3	38	36	0.0				0			
				KCTD	4	38	37	4.51	.37	-0.6		9			
				KCTE	5	26	21	5.04	.37	0.6		9			
11	01	1	4104								-1.82	9	0.71	1.40	11
				KCTA	1	18	9	-3.22	.65	-0.5		9			
				KCTB	2	28	10	-2.44	.38	0.7		9			
				KCTC	3	38	13	-1.47	.25	2.5		9			
				KCTD	4	38	14	-1.96	.26	0.0		9			
				KCTE	5	26	8	-1.70	.18	0.0		9			
24	01	1	5110								1.79	9	0.54	-0.03	24
				KCTC	3	38	24	2.66	.38	0.0		9			
				KCTD	4	38	25	2.03	.24	-0.2		9			
				KCTE	5	26	15	1.54	.15	-0.2		9			
14	01	1	4301								-0.67	9	0.53	2.14*	14
				KCTB	2	28	12	-0.49	.26	-0.2		9			
				KCTC	3	38	12	-1.19	.25	2.2		9			
				KCTD	4	38	15	-1.47	.27	1.7		9			
				KCTE	5	26	10	-0.26	.15	0.9		9			
10	01	1	4103								-2.88	9	0.52	1.33	10
				KCTA	1	18	7	-3.65	.88	1.2		9			
				KCTB	2	28	8	-2.96	.42	1.0		9			
				KCTC	3	38	8	-3.51	.27	1.4		9			
				KCTD	4	38	8	-3.17	.37	1.5		9			
				KCTE	5	26	7	-2.37	.20	0.0		9			
6	01	1	3201								-3.06	9	0.50	0.74	6
				KCTA	1	18	5	-3.65	.79	0.7		9			
				KCTB	2	28	5	-2.96	.44	1.5		9			
				KCTC	3	38	5	-3.69	.26	-0.4		9			
				KCTD	4	38	6	-3.38	.34	0.1		9			
				KCTE	5	26	5	-2.45	.22	0.9		9			
9	01	1	4102								-2.60	9	0.46	1.21	9
				KCTA	1	18	8	-2.24	.55	-1.0		9			
				KCTB	2	28	9	-2.44	.37	0.4		9			
				KCTC	3	38	11	-3.22	.26	1.5		9			
				KCTD	4	38	9	-2.03	.28	1.2		9			
25	01	1	5201								0.19	9	0.46	-1.22	25
				KCTC	3	38	22	0.96	.27	-1.6		9			
				KCTD	4	38	20	0.15	.22	-1.0		9			
				KCTE	5	26	12	-0.03	.15	0.3		9			
28	01	1	6102								2.57	9	0.46	0.79	28
				KCTA	1	18	14	3.21	.94	1.5		9			
				KCTB	2	28	20	2.54	.35	0.5		9			
				KCTC	3	38	26	3.34	.46	-0.1		9			
				KCTD	4	38	28	2.67	.28	1.2		9			
				KCTE	5	26	16	2.41	.17	0.0		9			
7	01	1	3301								-2.60	9	0.44	0.38	7
				KCTA	1	18	6	-3.22	.75	0.9		9			
				KCTB	2	28	6	-3.10	.43	0.9		9			
				KCTC	3	38	6	-3.01	.23	-0.1		9			
				KCTD	4	38	10	-2.31	.28	0.5		9			
				KCTE	5	26	6	-2.29	.20	0.0		9			

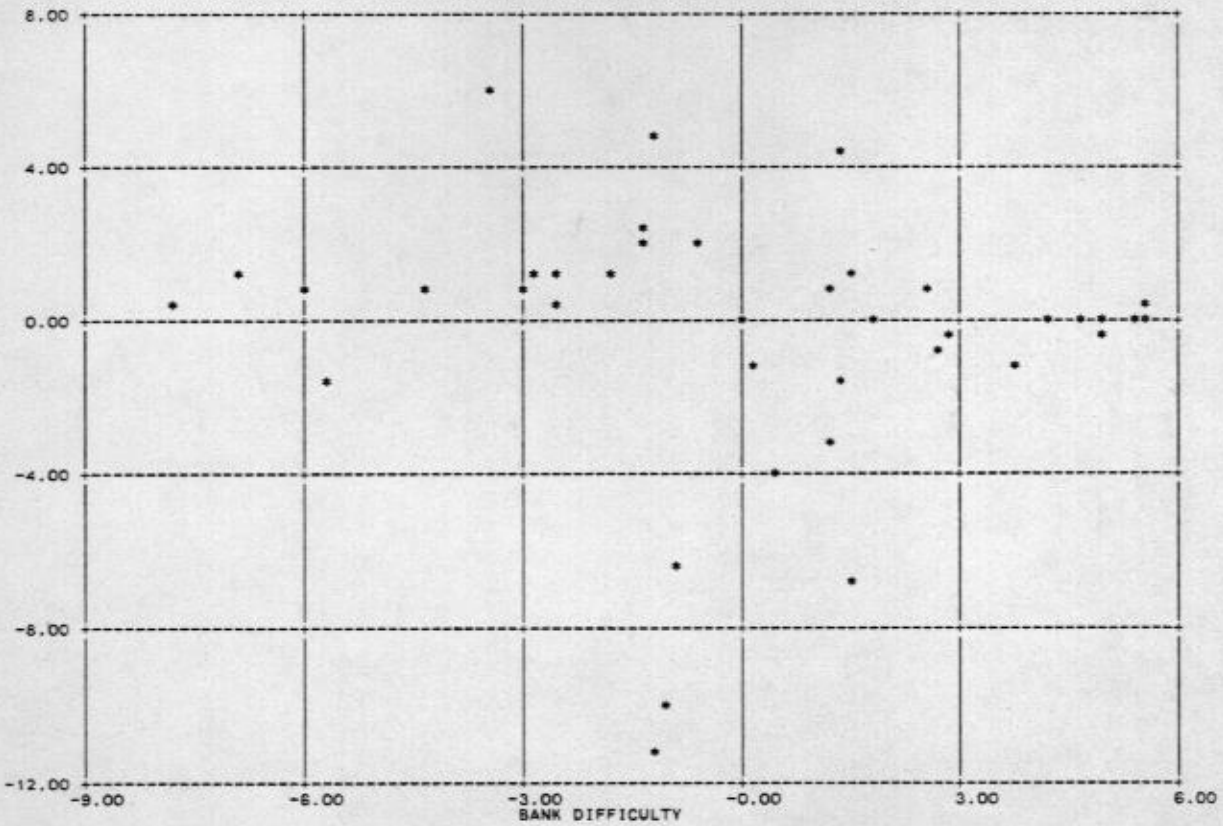
ITEM LIST SORTED BY DIFFICULTY RMS

ITEM NUMBER	LEGITIMATE ANSWERS	KEY	ITEM NAME	FORM NAME	FORM NUMBER	ITEMS IN FORM	ITEM POS	ITEM STATS IN FORM DIFF	ERROR	FIT T	BANK DIFF	STATUS	BETWEEN DIFF RMS	WITHIN FIT MS	ITEM NMBR
20	01	1	5104								1.22	9	0.43	0.70	20
				KCTA	1	18	12	2.14	.63	1.0		9			
				KCTB	2	28	18	1.01	.29	-0.2		9			
				KCTC	3	38	17	1.26	.29	0.4		9			
				KCTD	4	38	21	1.12	.22	-0.6		9			
				KCTE	5	26	13	1.26	.16	1.4		9			
19	01	1	5103								1.29	9	0.38	4.26*	19
				KCTA	1	18	11	0.76	.49	0.8		9			
				KCTB	2	28	17	1.58	.35	2.0		9			
				KCTC	3	38	16	0.96	.27	-0.4		9			
				KCTD	4	38	19	1.61	.27	3.5		9			
15	01	1	4401								-1.20	9	0.36	4.71*	15
				KCTB	2	28	14	-0.89	.29	-1.8		9			
				KCTC	3	38	15	-1.61	.22	-0.1		9			
				KCTD	4	38	12	-0.84	.27	3.3		9			
30	01	1	6104								4.14	9	0.36	-0.17	30
				KCTA	1	18	17	4.56	***	0.3		9			
				KCTB	2	28	23	4.24	.46	0.3		9			
				KCTC	3	38	29	0.0				0			
				KCTD	4	38	33	4.65	.38	-0.7		9			
				KCTE	5	26	20	3.88	.24	-0.1		9			
40	01	1	7402								5.56	9	0.36	0.40	40
				KCTA	1	18	16	0.0				0			
				KCTB	2	28	24	6.17	.97	0.9		9			
				KCTC	3	38	34	0.0				0			
				KCTD	4	38	34	5.56	.54	0.6		9			
				KCTE	5	26	23	5.45	.41	0.2		9			
35	01	1	6401								3.69	9	0.34	-1.10	35
				KCTC	3	38	33	0.0				0			
				KCTD	4	38	32	4.15	.37	1.0		9			
				KCTE	5	26	19	3.53	.22	-1.1		9			
22	01	1	5108								1.33	9	0.34	-1.49	22
				KCTC	3	38	21	1.76	.31	0.3		9			
				KCTD	4	38	23	1.12	.22	-1.7		9			
37	01	1	7105								4.92	9	0.33	0.05	37
				KCTB	2	28	28	0.0				0			
				KCTC	3	38	38	0.0				0			
				KCTD	4	38	35	5.34	.47	0.3		9			
				KCTE	5	26	22	4.73	.31	0.1		9			
12	01	1	4202								-1.35	9	0.32	2.44*	12
				KCTA	1	18	10	-1.50	.53	0.7		9			
				KCTB	2	28	11	-1.40	.35	1.5		9			
				KCTC	3	38	10	-1.98	.25	2.8		9			
				KCTD	4	38	11	-1.31	.24	-1.0		9			
				KCTE	5	26	9	-1.06	.17	0.8		9			
21	01	1	5105								1.54	9	0.28	1.08	21
				KCTA	1	18	13	1.86	.53	-0.6		9			
				KCTB	2	28	19	1.67	.31	0.3		9			
				KCTC	3	38	18	1.76	.31	-0.3		9			
				KCTD	4	38	24	1.92	.23	-0.1		9			
				KCTE	5	26	14	1.24	.16	2.2		9			
1	01	1	2101								-7.81	9	0.27	0.56	1
				KCTA	1	18	1	0.0				0			
				KCTB	2	28	1	0.0				0			
				KCTC	3	38	1	0.0				0			
				KCTD	4	38	1	-8.10	***	0.8		9			
				KCTE	5	26	1	-7.57	***	0.7		9			

BANK DIFFICULTY (X) VS BETWEEN DIFFICULTY ROOT MEAN SQUARE (Y)



BANK DIFFICULTY (X) VS WITHIN FIT MEAN SQUARE (Y)



If an item occurred in only one form its between difficulty RMS is 0.0, as illustrated by Item #2. Items with a root mean square greater than 0.5 are asterisked to identify them as varying in difficulty from form to form. These items should be examined to ensure that they haven't been miskeyed or misprinted in one of the forms in which they appear.

Examination of Page 1 of the ITEMLIST output shows that Items #4, #5, #8, #10, and #11 have between difficulty root mean squares larger than 0.5. The item difficulty of Item #4 in form KCTC is very different from the other difficulties estimated for this item. While the difficulty of Item #4 in the other forms ranges from -4.47 to -4.04, its estimated difficulty in form KCTC is -7.09. For Item #5, the calibration in form KCTC is again causing this item to misfit. The calibration of Item #5 in form KCTC is more than 2.5 logits easier than the calibrations of the item in the other forms. While these discrepancies are large in terms of logits, the standard errors are quite large for these extreme items with a relatively small sample.

The within fit mean square (FMS) is the average of an item's squared fit-t's for every form in which the item was calibrated. Using Item #1 as an example, the within fit mean square is calculated as

$$[(0.8^2 + 0.7^2)/2] = 0.56.$$

The sign of the within fit mean square is retained from the sign of the fit-t with the largest absolute value to distinguish between the misfit of an item caused by poor fit, indicated by large positive fit-t's, or by excessive good fit, indicated by large negative fit-t's. For example, Item #9 has three positive fit-t's and one negative fit-t associated with it. Since the negative fit-t has an absolute value greater than any of the four positive fit-t's, the within fit mean square is shown as a negative number. This indicates that the greatest source of the misfit for Item #9 is due to a pattern that fits too well, rather than to one that fits too poorly. Items with fit mean squares greater than 2.0 or less than -2.0 have asterisks beside them to mark their deviant fit within forms.

The forms in which an item was used are listed for every item. The list for each occurrence of an item includes the form name and number, number of items in the form, position of the item within the form, the difficulty of the item after being translated to the bank origin, the standard error associated with this difficulty, and the fit-t statistic of the item. The standard error and fit-t were calculated in FORCAL. If the standard error is greater than or equal to 1.0, three asterisks are printed instead of the

number. The status for the item within the form is also shown: a "9" means the item was calibrated in FORCAL and a "0" indicates the item was not calibrated.

A second list of the items, sorted by between difficulty root mean square follows the list of items by sequence number. The same information is printed as in the first list, but only those items with an RMS greater than 0.25 are shown. This list is not shown for this example, because no items had an RMS greater than 0.25.

The last section of ITEMLIST produces two plots: bank difficulty against between difficulty RMS and bank difficulty against within fit mean square. These plots are useful for reviewing the pattern of fit and misfit and for finding extreme items which may have been overlooked in the examination of the printed lists.

5.5 FORMLIST

The fifth program, FORMLIST, lists all the items in the bank organized by form. All the user supplies is a title which is printed at the top of every FORMLIST page.

Each form listing shows the form number and form name, the number of items in the form, and the form's bank difficulty (translation constant). Following this header infor-

mation, the form's items are listed in order of their position in the form. Each item is listed by its position in the form, number, key, within form difficulty and its standard error, total fit-t, and its difficulty with form translation constant added. The local form difficulty, standard error, and total fit-t were calculated in FORCAL. The item difficulty with translation constant is its within form difficulty with the translation constant added and is the difficulty shown in ITEMLIST. If an item was not calibrated in FORCAL, its local form difficulty, standard error, total fit-t, and item difficulty on the bank are shown as 0.0. For example, there are 18 items in form KCTA, and items #1, #2, #5, and #40 were not calibrated in FORCAL.

5.6 ITEMMAP

The bank items must be arranged in their order of difficulty to check that the items in a bank define a variable which agrees with substantive expectations. Computer program, ITEMMAP displays the variable graphically by locating the items, according to their bank difficulties, along the line of the variable which they define. This enables educators to examine the relationship between the content of the items and their difficulties and provides the framework to choose items for the design of new tests.

FORM NUMBER	ITEMS IN FORM	FORM NAME	ITEM POS	ITEM NUMBER	KEY	ITEM STATS IN FORM DIFF	ERROR	FIT T	BANK DIFF	ITEM DIFF ON BANK
1	18	KCTA							0.0	
			1	1	1	0.0	0.0	0.0		0.0
			2	2	1	0.0	0.0	0.0		0.0
			3	5	1	0.0	0.0	0.0		0.0
			4	4	1	-4.19	0.87	0.4		-4.19
			5	6	1	-3.65	0.79	0.7		-3.65
			6	7	1	-3.22	0.75	0.9		-3.22
			7	10	1	-3.65	0.88	1.2		-3.65
			8	9	1	-2.24	0.55	-1.0		-2.24
			9	11	1	-3.22	0.65	-0.5		-3.22
			10	12	1	-1.50	0.53	0.7		-1.50
			11	19	1	0.76	0.49	0.8		0.76
			12	20	1	2.14	0.63	1.0		2.14
			13	21	1	1.86	0.53	-0.6		1.86
			14	28	1	3.21	0.94	1.5		3.21
			15	29	1	4.56	1.09	0.3		4.56
			16	27	1	4.56	1.09	0.3		4.56
			17	30	1	4.56	1.09	0.3		4.56
			18	40	1	0.0	0.0	0.0		0.0

FORM NUMBER	ITEMS IN FORM	FORM NAME	ITEM POS	ITEM NUMBER	KEY	ITEM STATS IN FORM DIFF	ERROR	FIT T	BANK DIFF	ITEM DIFF ON BANK
2	28	KCTB							-0.04	
			1	1	1	0.0	0.0	0.0		0.0
			2	3	1	-6.65	0.88	1.3		-6.69
			3	5	1	-6.25	0.74	1.1		-6.29
			4	4	1	-4.43	0.48	0.3		-4.47
			5	6	1	-2.92	0.44	1.5		-2.96
			6	7	1	-3.06	0.43	0.9		-3.10
			7	8	1	-4.64	0.47	0.0		-4.68
			8	10	1	-2.92	0.42	1.0		-2.96
			9	9	1	-2.40	0.37	0.4		-2.44
			10	11	1	-2.40	0.38	0.7		-2.44
			11	12	1	-1.36	0.35	1.5		-1.40
			12	14	1	-0.45	0.28	-0.2		-0.49
			13	13	1	-1.19	0.30	-0.7		-1.23
			14	15	1	-0.85	0.29	-1.8		-0.89
			15	18	1	-1.85	0.33	-3.0		-1.89
			16	17	1	-1.10	0.30	-0.3		-1.14
			17	19	1	1.62	0.35	2.0		1.58
			18	20	1	1.05	0.29	-0.2		1.01
			19	21	1	1.71	0.31	0.3		1.67
			20	28	1	2.58	0.35	0.5		2.54
			21	29	1	3.05	0.36	-0.9		3.01
			22	27	1	3.05	0.36	-1.5		3.01
			23	30	1	4.28	0.46	0.3		4.24
			24	40	1	6.21	0.97	0.9		6.17
			25	39	1	5.69	0.77	0.7		5.65
			26	38	1	6.21	0.80	-0.2		6.17
			27	36	1	7.03	1.07	0.0		6.99
			28	37	1	0.0	0.0	0.0		0.0

FORM NUMBER	ITEMS IN FORM	FORM NAME	ITEM POS	ITEM NUMBER	KEY	ITEM STATS IN FORM DIFF	ERROR	FIT T	BANK DIFF	ITEM DIFF ON BANK
3	38	KCTC							-0.47	
			1	1	1	0.0	0.0	0.0		0.0
			2	3	1	0.0	0.0	0.0		0.0
			3	5	1	-6.48	1.39	0.9		-6.95
			4	4	1	-6.62	0.83	1.5		-7.09
			5	6	1	-3.22	0.26	-0.4		-3.69
			6	7	1	-2.54	0.23	-0.1		-3.01
			7	16	1	-3.04	0.30	3.4		-3.51
			8	10	1	-3.04	0.27	1.4		-3.51
			9	8	1	-5.93	0.55	-2.1		-6.40
			10	12	1	-1.51	0.25	2.8		-1.98
			11	9	1	-2.75	0.26	1.5		-3.22
			12	14	1	-0.72	0.25	2.2		-1.19
			13	11	1	-1.00	0.25	2.5		-1.47
			14	13	1	-0.98	0.24	2.2		-1.43
			15	15	1	-1.14	0.22	-0.1		-1.61
			16	19	1	1.43	0.27	-0.4		0.98
			17	20	1	1.73	0.29	0.4		1.26
			18	21	1	2.23	0.31	-0.3		1.76
			19	18	1	0.53	0.25	-5.0		0.06
			20	17	1	0.47	0.25	-5.1		0.0
			21	22	1	2.23	0.31	0.3		1.76
			22	25	1	1.43	0.27	-1.6		0.96
			23	26	1	0.96	0.26	-3.9		0.49
			24	24	1	3.13	0.38	0.0		2.66
			25	23	1	2.14	0.30	-1.9		1.67
			26	28	1	3.81	0.46	-0.1		3.34
			27	29	1	3.81	0.46	-0.2		3.34
			28	27	1	3.28	0.40	-1.1		2.81
			29	30	1	0.0	0.0	0.0		0.0
			30	31	1	5.09	0.74	-0.2		4.62
			31	32	1	3.61	0.44	-0.7		3.14
			32	33	1	5.09	0.74	0.0		4.62
			33	35	1	0.0	0.0	0.0		0.0
			34	40	1	0.0	0.0	0.0		0.0
			35	39	1	0.0	0.0	0.0		0.0
			36	38	1	0.0	0.0	0.0		0.0
			37	36	1	0.0	0.0	0.0		0.0
			38	37	1	0.0	0.0	0.0		0.0

FORM NUMBER	ITEMS IN FORM	FORM NAME	ITEM POS	ITEM NUMBER	KEY	ITEM STATS IN FORM DIFF	ERROR	FIT T	BANK DIFF	ITEM DIFF ON BANK
4	38	KCTD							-0.08	
			1	1	1	-8.02	1.36	0.8		-8.10
			2	3	1	-7.14	1.15	1.4		-7.22
			3	5	1	-5.40	0.55	-0.1		-5.48
			4	4	1	-3.98	0.39	-0.2		-4.04
			5	8	1	-7.14	0.84	-0.8		-7.22
			6	6	1	-3.31	0.34	0.1		-3.39
			7	16	1	-3.31	0.36	0.7		-3.39
			8	10	1	-3.09	0.37	1.5		-3.17
			9	9	1	-1.95	0.28	1.2		-2.03
			10	7	1	-2.23	0.28	0.5		-2.31
			11	12	1	-1.23	0.24	-1.0		-1.31
			12	15	1	-0.76	0.27	3.3		-0.84
			13	13	1	-1.28	0.24	-0.9		-1.36
			14	11	1	-1.88	0.26	0.0		-1.96
			15	14	1	-1.39	0.27	1.7		-1.47
			16	17	1	-2.71	0.30	-2.1		-2.79
			17	18	1	-2.16	0.27	-0.2		-2.24
			18	26	1	-1.07	0.26	2.0		-1.15
			19	19	1	1.69	0.27	3.5		1.61
			20	25	1	0.23	0.22	-1.0		0.15
			21	20	1	1.20	0.22	-0.6		1.12
			22	23	1	0.01	0.22	-2.0		-0.07
			23	22	1	1.20	0.22	-1.7		1.12
			24	21	1	2.00	0.23	-0.1		1.92
			25	24	1	2.11	0.24	-0.2		2.03
			26	27	1	2.51	0.25	0.2		2.43
			27	32	1	0.87	0.22	-2.4		0.79
			28	28	1	2.75	0.28	1.2		2.67
			29	29	1	2.95	0.26	0.0		2.87
			30	31	1	3.02	0.27	-0.1		2.94
			31	34	1	1.54	0.23	-2.6		1.46
			32	35	1	4.23	0.37	1.0		4.15
			33	30	1	4.73	0.38	-0.7		4.65
			34	40	1	5.64	0.54	0.6		5.56
			35	37	1	5.42	0.47	0.3		5.34
			36	39	1	5.91	0.54	0.1		5.83
			37	38	1	4.59	0.37	-0.6		4.51
			38	36	1	5.42	0.46	-0.4		5.34

FORM NUMBER	ITEMS IN FORM	FORM NAME	ITEM POS	ITEM NUMBER	KEY	ITEM STATS IN FORM DIFF	ERROR	FIT T	BANK DIFF	ITEM DIFF ON BANK
5	26	KCTE							0.58	
			1	1	1	-8.15	1.24	0.7		-7.57
			2	3	1	-7.40	0.81	0.4		-6.82
			3	5	1	-6.30	0.59	0.9		-5.72
			4	4	1	-4.80	0.35	1.1		-4.22
			5	6	1	-3.03	0.22	0.9		-2.45
			6	7	1	-2.87	0.20	0.0		-2.29
			7	10	1	-2.95	0.20	0.0		-2.37
			8	11	1	-2.28	0.18	0.0		-1.70
			9	12	1	-1.64	0.17	0.8		-1.06
			10	14	1	-0.84	0.15	0.9		-0.26
			11	26	1	-1.87	0.17	0.4		-1.29
			12	25	1	-0.61	0.15	0.3		-0.03
			13	20	1	0.68	0.16	1.4		1.26
			14	21	1	0.66	0.16	2.2		1.24
			15	24	1	0.96	0.15	-0.2		1.54
			16	28	1	1.83	0.17	0.0		2.41
			17	29	1	2.19	0.18	-1.0		2.77
			18	31	1	2.07	0.18	-1.0		2.65
			19	35	1	2.95	0.22	-1.1		3.53
			20	30	1	3.30	0.24	-0.1		3.88
			21	38	1	4.46	0.37	0.6		5.04
			22	37	1	4.15	0.31	0.1		4.73
			23	40	1	4.87	0.41	0.2		5.45
			23	39	1	4.87	0.41	0.2		5.45
			25	36	1	4.72	0.38	-0.4		5.30
			26	41	1	5.04	0.43	0.0		5.62

The control parameters for ITEMMAP include the number of items in the bank and the values of the easiest and hardest items. A constant is added to all item difficulties so that the numbers will have positive values. In the KCT example, this constant was +8.00.

The ITEMMAP output is divided into three sections. The first map is spaced at 0.20 logits per line, the second at 0.10 logits, and the third map at 0.05 logits. The third map is especially useful when there are a large number of items in the bank or when there are many items with approximately the same difficulty. As there are only 41 items in the KCT analysis, only the first map, spaced at 0.20 logits, is shown.

The ITEMMAP shows that Item #1, a two-tap sequence of 1-4, is the easist item in the bank, with a rescaled difficulty of about 0.20 logits (-7.8+8.0=0.2). The next four easiest items are #3, #5, #8, and #4. The tapping sequences associated with these items are 2-3, 1-2-4, 1-2-3-4, and 1-3-4. Thus, we see that the four-tap sequence, 1-2-3-4, which uses all the cubes in sequential order, is easier than the three-tap sequence, 1-3-4, which requires a jump of one cube.

Items #40 and #36 are the most difficult items in the bank. The tapping sequences associated with these items are 4-1-3-4-2-1-4 and 1-4-2-3-2-4-3. Both require seven taps with three reversals, and at least one jump of three cubes. The two items next in difficulty are #41, the only eight-tap sequence in the bank, and #39. These have tapping sequences of 1-3-2-4-2-3-1-2 and 3-2-4-1-3-4-2. Both require three reversals and Item #39 has one three-cube jump. Although Item #41 has no jumps of more than two, its extra tap makes it as difficult as the seven-tap sequence with one three-cube jump.

Examination of the entire ITEMMAP shows that the easier items are those requiring two, three, or four taps with either no reversals or one reversal, and the harder items are the five, six, seven, and eight tap sequences with two or three reversals and jumps of two or three cubes. This is consistent with our expectation that items requiring more complicated tasks are more difficult.

This ends the analysis of the KCT example and program description. Now that the steps for constructing a bank have been described, the next three chapters present analyses of three item banks constructed from three data sets.

LOGIT SCALE	ITEMS COUNT	ITEM NAMES		
0.0				
0.20	1	1		
0.40				
0.60				
0.80				
1.00				
1.20	1	3		
1.40				
1.60				
1.80				
2.00	1	5		
2.20				
2.40	1	8		
2.60				
2.80				
3.00				
3.20				
3.40				
3.60	1	4		
3.80				
4.00				
4.20				
4.40				
4.60	1	16		
4.80				
5.00	1	6		
5.20	1	10		
5.40	2	9	7	
5.60				
5.80				
6.00				
6.20	1	11		
6.40				
6.60	2	12	13	
6.80	3	15	18	17
7.00				
7.20	1	26		
7.40	1	14		
7.60				
7.80				
8.00	1	2		
8.20	1	25		
8.40				
8.60	1	23		
8.80				
9.00				
9.20	3	20	32	19
9.40	2	22	34	
9.60	1	21		
9.80	1	24		
10.00				
10.20				
10.40				
10.60	1	28		
10.80	2	27	31	
11.00	1	29		
11.20				
11.40				
11.60	1	35		
11.80				
12.00				
12.20	1	30		
12.40				
12.60	1	33		
12.80				
13.00	2	38	37	
13.20				
13.40	1	36		
13.60	3	40	39	41
13.80				
14.00				
14.20				
14.40				
14.60				
14.80				

LOGIT SCALE ITEMS COUNT ITEM NAMES

OF 41 ITEMS READ 41 APPEAR ON THIS MAP
THE CONSTANT ADDED IS 8.00
MEAN = 8.26
S.D. = 3.59
MIN. = 0.19
MAX. =13.62
MAP OF KCT5 BANK - FIVE FORMS - 0.7 CUT-OFF

Chapter VI

STATE OF CALIFORNIA COMPUTATION ITEM BANK

California state law requires periodic assessment of pupil proficiency in reading comprehension, writing, and computation in grades four through eleven. As a result, a range of assessment items in each of these three areas is essential. As a part of this proficiency assessment program the California State Department of Education developed a pool of computation items appropriate for elementary school children.

The 300 items are grouped into eight basic content areas:

1. knowledge of arithmetic facts (55 items);
2. arithmetic computation (73 items);
3. arithmetic comprehension (36 items);
4. arithmetic applications (44 items);
5. expressions, equations, and formulas (17 items);
6. intuitive geometry (25 items);
7. measurement (35 items), and;
8. interpreting data from tables and charts (15 items).

The 300 items in this pool were organized into a web of twenty-five parallel forms with twenty-four items comprising

each form. The twenty-five forms were then administered to school children in grades four and six with between 87 and 97 persons taking each form.

The FORCAL analysis used persons whose raw score was between 30% and 95% of the total possible score for any one form. As all forms originally had 24 items, the minimum score for every form was 7 and the maximum was 22. The calibration procedure used was the unconditional method, UCON, and all forms used two iterations to arrive at the final item difficulty and person ability estimates. No persons were eliminated for misfit.

The pertinent information from the results of the analysis are summarized in two tables. Table 6.1 shows form number and difficulty, total number of persons who took the form, number of persons measured, with ability mean and standard deviation, the number of persons used to calibrate the items, and the minimum, maximum, and standard deviation of the item difficulties. For each form, Table 6.2 lists form number and difficulty, number of persons used to calibrate the items, with the minimum, maximum, and standard deviation of the item difficulties, and the mean and standard deviation of the fit statistics: point biserial, discrimination index, between fit-t, and total fit-t.

TABLE 6.1

Summary Characteristics by Form

Form		Sample				Calibration				
		Persons		Ability		Per Item		Difficulty		
Num	Diff	Tot	Meas	Mean	S.D.	Used	Cnt	Min.	Max.	S.D.
301	0.11	94	93	1.16	0.96	86	24	-2.01	2.39	1.23
302	-0.44	94	88	0.77	1.17	79	24	-2.33	1.99	1.16
303	0.06	92	89	1.07	1.13	78	24	-1.98	2.17	1.23
304	0.64	94	94	1.28	0.92	93	24	-2.83	3.46	0.80
305	-0.28	93	92	0.71	1.02	86	24	-2.57	1.91	1.22
306	0.07	95	93	0.99	1.14	83	24	-2.36	2.01	1.16
307	0.18	90	88	0.79	1.15	80	24	-2.19	2.16	1.14
308	-0.22	96	96	0.77	1.08	88	24	-2.66	2.41	1.33
309	-0.24	93	88	0.94	1.19	77	24	-1.95	1.81	1.03
310	-0.14	88	88	1.06	0.76	87	24	-2.22	4.41	1.68
311	0.29	91	91	0.97	0.98	87	24	-2.51	2.75	1.46
312	-0.10	97	95	1.02	1.05	90	24	-2.46	1.88	1.21
313	0.19	96	94	1.10	1.17	88	24	-2.24	2.32	1.45
314	-0.02	92	92	0.92	1.02	89	24	-2.21	2.77	1.35
315	-0.34	92	90	0.93	1.17	82	24	-3.88	2.37	1.55
316	-0.43	92	92	0.77	0.99	85	24	-2.70	1.16	1.23
317	0.03	92	92	0.87	1.11	85	24	-2.05	2.92	1.15
318	0.37	93	92	1.21	0.81	91	24	-2.09	2.85	1.37
319	-0.45	94	89	0.97	0.91	84	24	-3.12	2.27	1.32
320	0.07	94	94	1.17	0.98	91	24	-3.02	3.69	1.60
321	0.15	94	92	0.93	1.23	83	24	-1.72	3.02	1.20
322	-0.02	90	85	1.16	0.94	79	24	-2.69	2.15	1.31
323	0.12	89	87	1.05	1.05	80	24	-2.49	2.38	1.41
324	0.15	89	89	1.33	1.00	87	24	-2.79	4.04	1.48
325	0.24	93	88	1.26	1.08	80	24	-1.60	2.83	1.17

Examination of Tables 6.1 and 6.2 shows each form's dispersion of item difficulties, range of items, and where the misfit occurs. For example, Form 310 has the largest difficulty standard deviation, 1.68, the hardest item in the bank with a difficulty of 4.41 logits, and the second largest between fit-t mean, 1.80. These statistics suggest that the item(s) causing misfit in this form may be at the extremes of the variable.

TABLE 6.2

Summary Fit Statistics by Form

Form		Point Biserial		Discrimin		Between Fit-t		Total Fit-t	
Num	Diff	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
301	0.11	0.34	0.14	1.02	0.39	0.52	2.20	0.05	0.57
302	-0.44	0.42	0.13	1.03	0.34	1.00	2.64	-0.07	0.63
303	0.06	0.38	0.16	1.06	0.48	0.77	2.57	-0.18	0.97
304	0.64	0.38	0.15	1.04	0.41	1.56	3.20	0.12	0.80
305	-0.28	0.37	0.14	1.03	0.37	0.98	2.10	-0.05	0.75
306	0.07	0.39	0.12	1.03	0.32	1.11	2.09	-0.16	0.67
307	0.18	0.38	0.13	1.03	0.40	1.11	2.37	-0.08	0.59
308	-0.22	0.37	0.17	1.03	0.50	1.66	2.57	-0.02	0.90
309	-0.24	0.43	0.14	1.04	0.42	1.08	2.39	-0.22	0.96
310	-0.14	0.33	0.16	0.99	0.51	1.80	2.48	0.04	0.69
311	0.29	0.37	0.11	1.06	0.36	0.83	1.89	-0.01	0.77
312	-0.10	0.63	0.13	1.06	0.40	0.75	2.65	0.04	0.99
313	0.19	0.38	0.11	1.04	0.28	0.45	2.25	-0.04	0.64
314	-0.02	0.38	0.17	1.02	0.43	1.57	2.55	0.04	0.85
315	-0.34	0.39	0.12	1.05	0129	0.24	2.78	-0.11	0.69
316	-0.43	0.35	0.15	1.04	0.48	0.67	2.39	-0.14	0.94
317	0.03	0.42	0.12	1.06	0.34	1.16	2.02	-0.10	0.89
318	0.37	0.35	0.13	1.03	0.37	0.89	1.99	0.09	0.62
319	-0.45	0.36	0.18	1.04	0.50	1.30	2.21	-0.01	0.76
320	0.07	0.39	0.14	1.08	0.44	1.52	2.29	0.06	0.89
321	0.15	0.42	0.13	1.04	0.36	1.86	2.83	-0.15	0.76
322	-0.02	0.36	0.13	1.08	0.48	0.38	2.65	0.02	0.80
323	0.12	0.36	0.13	1.06	0.38	1.30	2.34	-0.05	0.77
324	0.15	0.37	0.16	1.00	0.36	0.64	2.74	0.07	0.52
325	0.24	0.38	0.13	1.05	0.36	0.58	2.06	-0.08	0.71

6.1 EXTREME ITEMS

Since items at the extremes, either very easy or very hard, are frequently different than the other items, and hence often a source of misfit, they should be identified and their content investigated. In an item bank this size the most efficient way to find such items is through the variable map. As this bank is divided into eight content areas,

separate variable maps for each area were produced. These maps are examined next.

6.1.1 Knowledge of Arithmetic Facts

The first content area is divided into eight skills:

1. recognizing addition or subtraction terminology;
2. recognizing multiplication or division terminology;
3. recalling basic addition facts;
4. recalling basic subtraction facts;
5. recalling basic multiplication facts;
6. recalling basic division facts;
7. recognizing addition, subtraction, multiplication, and division symbols; and,
8. recognizing symbols for equality and relationship.

The most extreme item in this content area, #277, was the hardest item in both forms in which it occurred and was also the most difficult item in the entire item bank. Since this item requires a relatively easy task, "What does the symbol "-" mean?" it was surprising that students found it extremely difficult. In addition, other items in the same skill, recognizing addition and subtraction symbols, appeared easy to the students. However, further examination revealed that Item #277 was incorrectly keyed, so that "division" rather than "subtraction" was designated the right answer. Thus, the cause for its extreme difficulty was either that the item was miskeyed or that the symbol was

misprinted in the forms as a subtraction symbol rather than a division sign.

One other item in knowledge of arithmetic facts, #5, from recognizing multiplication and division terminology, was extremely difficult. Item #5 reads: "To find the product of 4 and 7, you must:" with alternatives: add, subtract, multiply, and divide. No keying or printing error was found to explain why this item was so hard. A possible explanation for the excessive difficulty of Item #5 is that students do not associate "product" with multiplication.

Of the other three items in this skill, only one, Item #6, was easy. It reads: "To find what 3 times another number is, you must:" with the alternatives: add, subtract, multiply, divide. This item uses the teacher-oriented term, "times," to describe the multiplication process. The other two items in this skill, Item #7, "To find out how many times 4 goes into 12, you must:" and Item #8, "To find out how many groups of 3 there are in 15, you must:" use the familiar division terms "goes into" and "how many groups of." The alternatives for both these items are the same as for Items #5 and #6. These items were difficult but not extremely so. No items in this content area were easy, suggesting terminology is not an area that is stressed.

6.1.2 Arithmetic Computation

The fifteen skills in this content area are:

1. adding whole numbers without renaming;
2. adding whole numbers with renaming;
3. subtracting whole numbers without renaming;
4. subtracting whole numbers with renaming;
5. multiplying whole numbers without renaming;
6. multiplying whole numbers with renaming;
7. dividing whole numbers with one-digit divisors;
8. adding common fractions with like denominators;
9. subtracting common fractions with like denominators;
10. adding or subtracting mixed numbers with like denominators;
11. adding decimal fractions;
12. subtracting decimal fractions;
13. multiplying decimal fractions;
14. estimating whole number sums and differences; and,
15. estimating whole number products and quotients.

There were seven items which stood apart from the other arithmetic computation items. Of these items, two, #37 and #39, deal with adding mixed numbers with like denominators. The two items requiring subtraction of mixed numbers were difficult, but not extreme. Item #39, "4 3/4 + 3 2/4" written vertically and with the alternatives: 1 1/4, 7 5/8, 8 1/4, 8 5/4, was the most difficult item in both forms in which it appears and was the most difficult item in this

content area. It was also the second hardest item in the whole bank. Item #37, "2 3/5 + 4 1/5" (aligned vertically) with the alternatives: 6 4/10, 6 4/5, 8 3/25, 8 4/5, was the hardest item in one of its two forms. The addition items in this skill provide an incorrect alternative in which denominators are added: 6 4/10 for Item #37 and 7 5/8 for Item #39. To correctly answer this type of item the student must know enough to avoid these alternatives. Item #39 is even more difficult because it requires renaming as well.

Of the five other extreme items in this content area, three items, #85, #86, and #87, are from the adding common fractions with like denominators skill. Items #85 and #86 were aligned vertically and Items #86 and #87 use three addends. Item #85 reads: "2/5 + 1/5" and uses the alternatives: 2/25, 1/5, 3/10, and 3/5. Items #86 is: "1/4 + 3/4 + 2/4" with alternatives: 6/16, 6/12, 6/8, 6/4; and Item #87 reads: "3/8 + 5/8 + 4/8 =" with the alternatives: 12/24, 12/16, 12/8, and 13/8. The difficulty of these items can again be attributed to the temptation to add denominators. Support for this interpretation comes from the remaining item in this skill, Item #88, "2/10 + 4/10 + 3/10 =" with the alternatives: 9/10, 10/10, 11/10, 24/10. This item does not include a distractor inviting this error and was relatively easy.

The other two extremely difficult items in this content area, #42 and #52, require adding and multiplying decimal fractions, and both are written vertically. The difficulty of Item #42, "68.47 + 31.53" with alternatives: .0100; .100; 99.91; 100; may be due to a failure to rename or the result of not being certain that the correct answer, "100.00", is equivalent to the printed alternative "100". The second most difficult item, #52, was also the most difficult item in both of the forms in which it appeared. This item, "8.06 x 30" with alternatives: 241.8; 2,418; 8,060; 24,180;, may owe its extreme difficulty to requiring multiplication by zero. There were no extremely easy items in this content area.

6.1.3 Arithmetic Comprehension

This content area consists of nine skills:

1. reading, writing, and expressing place value;
2. recognizing place value of a given digit in a numeral;
3. identifying ordinal positions;
4. ordering and comparing whole numbers;
5. identifying multiples of a given number;
6. recognizing and extending number patterns;
7. recognizing money values and money notation;
8. recognizing fractions; and,
9. identifying equivalent fractions.

The items at the hard end of this content area were evenly dispersed, except for Item #128, which was somewhat more difficult than the others. Item #128 requires identifying an equivalent fraction for 2/3 and uses the alternatives: 2/5, 4/5, 4/6, 3/6. This is the only item in which the numerator is not "1." The other three items in this skill, which require picking the fraction with the same name as 1/2, 1/5, and 1/4, all had approximately the same difficulty and were at the hard end of this content area.

There were two very easy items in the arithmetic comprehension area, although they were not distinguishable as such on the map of the whole item bank. Both of these items, #114 and #116, are part of the recognizing and extending number patterns skill. Two items in this skill have the missing number in the last position of the sequence, e.g. 1, 3, 4, 7, _; and two have it in the second to last position, e.g. 0, 2, 4, 6, _, 10. Items #114 and #116 are of the latter type. The reason picking the next to last number is easier than picking the last may be because it is easier to determine the correct pattern when there is information both before and after the missing term.

6.1.4 Arithmetic Applications

This content area is divided among eleven skills very similar to those of arithmetic computation:

1. adding whole numbers;

2. subtracting whole numbers;
3. adding and subtracting whole numbers;
4. multiplying whole numbers;
5. dividing whole numbers;
6. adding decimal fractions;
7. adding and subtracting decimal fractions;
8. multiplying decimal fractions;
9. estimating answers to word problems;
10. using problem analysis techniques; and,
11. using problem checking techniques.

The difference between arithmetic computation and applications is that in applications the student is given a "story" problem in which the numbers are embedded, instead of just the numbers needed for the calculation.

There were two items in this content area which are much harder than the other arithmetic applications items: #152 and #155, multiplying whole numbers and dividing whole numbers, respectively. The other three items in each of these skills were far easier than these extreme items. The wording of the problem in Item #155 might be the reason for its difficulty. This item reads: "There are 12 pencils in each box and 360 pencils in each carton. How many boxes of pencils are there in a full carton?" with alternatives: 30 boxes, 60 boxes, 120 boxes, 360 boxes. In the other three items in this skill the form of the problem is X - Y = ?,

where X is the first number in the word problem and Y is the second number. In Item #155, however, X is the second number and Y the first number in the story, thus the student may make an error by failing to follow the basic facts of the story.

Item #152's difficulty may be the result of requiring the multiplication of two two-digit numbers, from the story: "The bakery had 58 packages of eggs. Each package had 36 eggs in it. How many eggs did the bakery have?" with alternatives: 94 eggs; 1,548 eggs; 2,088 eggs; 8,088 eggs. There is one extremely easy item in content area four, #141, subtracting whole numbers, and reads: "There were 10 boys in the class, but 6 of them went home. How many were left?" and uses the alternatives: 4 boys, 6 boys, 10 boys, and 16 boys. This item is the easiest item in the bank, as well as the easiest item in each of the two forms in which it occurs.

6.1.5 Expressions, Equations, and Formulas

This content area consists of four skills:

1. evaluating simple expressions involving addition or subtraction;
2. solving equations involving addition or subtraction;
3. recognizing simple equations; and,
4. evaluating simple algebraic formulas.

The items in this content area were evenly, but widely dispersed. However, two of the four items in the fourth skill, items #185 and #188, stand apart from the other items. Item #185, the hardest item in this content area, was the second most difficult item in both forms in which it appeared and was among the ten hardest items in the bank. In addition to these extreme items, a third item, #187, was the next most difficult item, although not as extreme. The three items in this skill, evaluating simple algebraic formulas, all involve multiplication and items #185 and #188 used the same formula, A = L x W, where L and W are given and A must be determined. However, the form of these two problems may be giving students trouble. Item #185 reads:

A = L x W
The length is 8 feet.
The width is 3 feet.
The area is _____.

The alternatives are: 5 square feet, 11 square feet, 24 square feet, and 38 square feet. Item #188 is of this same form and has the alternatives: 5 square feet, 9 square feet, 20 square feet, and 25 square feet. Item #187, however, reads:

D = T x R
Time = 3 hours
Rate = 60 kilometres per hour
Distance = ?

and uses the alternatives: 18 kilometres, 57 kilometres, 63 kilometres, and 180 kilometres. The other item in this skill, #186, is of this form, but requires division, rather than multiplication.

Examination of these four problems shows that the stories that use the word "is" in place of "=" signs, #185 and #188, were the most difficult items. The next hardest item, #187, is the one that uses "=" signs, but also uses kilometres in its formula, whereas the easiest item by more than a logit, #186, uses "=" signs and the more familiar "miles" in its story.

Item #181 is much easier than the other items. This item requires students to match the appropriate number sentence equation with the story. This item reads:

> Kim had 9 apples.
> She picked 4 more.
> How many apples does she have in all?
>
> Which number sentence matches the story?

The number sentence alternatives are: 9 - 4 = __; 9 + 4 = __; 4 + __ = 9; 9 X 4 = __. The two easiest items in this content area, #181 and #183, are addition problems of this kind. The others in this skill require subtraction and multiplication equations and are much more difficult.

6.1.6 Intuitive Geometry

There are six skills in this content area:

1. identifying the line segments of geometric figures;
2. identifying relationships between geometric figures;
3. recognizing symmetry;

4. identifying basic two-dimensional geometric figures;
5. identifying basic three-dimensional geometric figures; and,
6. identifying coordinates and locations of points on the coordinate plane (first quadrant).

All but one of the items in this content area were evenly spaced along the variable line. No items stand out as extremely difficult, but there is one easy item, #199. This item is a logit easier than any other intuitive geometry item and more than two logits easier than any other item in its skill, identifying two-dimensional figures. Item #199 requires the student to identify the number of triangles in an octagon that has been divided into six equilateral triangles. The other items in this skill, from the easiest to the hardest, involve identifying the number of triangles in a rectangle, where the triangles are not all the same shape, identifying which figure is not a rectangle, and identifying a right angle.

Although no items are extremely hard with respect to this content area, the two most difficult items in this area are among the ten hardest items in the bank. Item #252 asks: "Which parts of this cube remind you of line segments?" and Item #253 reads: "Which figure shows perpendicular lines?" Both items were used in previous administrations and the characteristics of these item stems were changed for the

current administration. The item stem for the other items in the same skill as #252 consist of a two-dimensional drawing in which individual line segments are identified by a capital letter. Item #252, in contrast, is a three-dimensional figure with no identifying marks on its line segments. This produces a different type of item which is more difficult. Item #253 uses the easier item stem so its difficulty must be due to not knowing what "perpendicular" means.

6.1.7 Measurement

This content area is divided into eight skills:

1. estimating and choosing the measure of familiar objects or distances;
2. renaming within the U.S. Customary and Standard International Metric system of measurement;
3. taking the linear measure of geometric shapes;
4. using measurement instruments;
5. interpreting maps and scale drawings;
6. calculating with units of time;
7. interpreting calendars; and,
8. telling time.

The items in this content area were evenly spaced, except for Item #234, which is a logit easier than any other measurement item. This item is from the interpreting calendars skill and requires the student to determine on what day of the week July 4th falls. This should be an easy item and it

is, yet a similar item in this skill, "March 18th falls on what day?" is much more difficult. The difference between these items is that July 4th falls on a Tuesday while March 18th falls on a Saturday. Students will find the March 18th item difficult to answer correctly if they fail to note the calendar convention of placing Saturday at the right end of the week and Sunday at the left, and hence, determine incorrectly that March 18th falls on a Sunday.

The two most difficult items in this content area are not extreme with respect to the measurement variable, but they are among the five hardest items in the bank. Item #256 involves interpreting maps and scale drawings. It shows a drawing and asks that the distance between two points on a map be calculated. It reads: "Bill followed the path and went from home to the mountains for a picnic. How far was his round trip if he went the shortest way possible?" This item is the same as others in its skill, but requires the student to know the meaning of "round trip", rather than "from X to Y and back" as another, easier, item in this skill asks.

The other item, #255, reads: "2 1/2 tons equals how many pounds?" with alternatives: 40 pounds; 250 pounds; 2,000 pounds; 5,000 pounds. It is a straightforward problem, but not of the same form as the others in skill two. This may

be due to the use of this item in previous administrations and a subsequent change in item stems. For this item to conform to the others in its skill it should read: "The truck weighed 2 1/2 tons. How many pounds is that?"

6.1.8 Interpreting Data from Tables and Graphs

This content area in computation consists of three skills:

1. interpreting data tables;
2. interpreting bar graphs; and,
3. interpreting picture graphs.

The fifteen items in this content area are spread over three and a half logits, but none are extremely difficult, extremely easy, or stand out from the rest of the items.

6.2 MISFITTING ITEMS

Once the items have been banked, the identification and study of misfitting items is the next step. These items are listed in Tables 6.3 through 6.10. They are arranged by the magnitude of the between fit-t statistic within content area. Item number, form number(s) in which they misfit, bank difficulty and standard error, point biserial, discrimination index, error impact, and between and total fit-t are listed for each misfitting item.

Some general sources of misfit, such as random guessing, carelessness, and miskeying, occasionally occur in test data

regardless of the content area. Examination of the fit statistics cannot diagnose these disturbances unambiguously, but it can call attention to problem items. Understanding how the disturbances would appear in the data can suggest hypotheses that can be checked by examining the item content, scoring, key, etc.

Random guessing is only a problem with relatively difficult items, when more low ability persons give the correct answer than expected, based on their abilities. The result on the calibration is that the item difficulty is under estimated due to the unwarranted successes. The symptoms in the fit statistics are: high difficulty, high between fit-t, and low discrimination index. The statistics for Items #253, #209, #256, and #229 in Table 6.8 and 6.9 are consistent with this pattern.

When examining departures from the estimated item characteristic curve, it frequently happens with guessing data that there are too many successes in the low ability region, as we would anticipate, and too few successes in the higher ability region. The too few successes is due to our underestimation of the item difficulty; the high ability persons would respond to the true difficulty rather than our estimate of it, giving the appearance of carelessness.

Carelessness has an analogous but almost opposite effect. It occurs when high ability persons fail items they should have been able to pass. The effect on the calibration is to overestimate the item difficulty. The pattern produced in the statistics is low difficulty, high between fit-t, and low discrimination. Items #258, #9, #280, #192, and #248 are examples of this pattern.

The typical pattern in the item characteristic curve would be a fairly high proportion passing in every ability group, but which levels off or decreases in the highest groups. There may also appear to be a surprising number of passes in the low group, due to the overestimation of the item's difficulty.

Miskeying produces a very obvious pattern of misfit. When this occurs, the item will appear very difficult,the between fit-t will be very large and the discrimination will often be negative. Item #277 illustrates this pattern. After changing the key, this item was rather easy and had no evidence of misfit.

Interactions with exposure can also affect the shape of the ICC. If, for example, the item involves a special skill or a piece of information that only the high ability students have been exposed to, then these student's will find

the item relatively "easier" than do low ability students, and the item will have a high index of discrimination. On the other hand, a skill that is unrelated to instruction, so that students at any ability level are equally likely to acquire it, will make the item relatively "easier" for the low ability students and, hence, to have a low discrimination index.

6.2.1 Knowledge of Arithmetic Facts

TABLE 6.3

Misfitting Items from Knowledge of Arithmetic Facts

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
277	310	4.41	0.52	-0.06	-0.50	0.16	9.50	0.38
	324	4.04	0.39	-0.11	-0.19	0.21	10.59	1.09
258	304	-2.83	0.72	-0.03	0.24	0.0	6.48	-0.05
9	319	-1.49	0.37	-0.06	-0.01	0.16	6.23	0.82
30	308	1.28	0.24	0.08	-0.03	0.15	6.23	2.00
32	307	0.17	0.25	0.14	0.03	0.13	5.69	1.51

In this content area, five out of fifty-five items have between fit-t's greater than 5.0. Item #277 showed the most misfit, which was explained as miskeying.

The four other items in this content area with between fit-t's larger than 5.0 come from two skill areas. Items #258 and #9 deal with recalling basic addition facts, and items #30 and #32 involve recognizing symbols for equality and relationship.

Item #258, "6 + 9" (aligned vertically), and with alternatives: 69, 14, 15, 3; was an extremely easy item in both forms 304 and 308. It was also the sixth easiest item in the bank and the easiest item in Form 304, the form in which it misfit. Everyone in all six ability groups correctly answered the item, except for the fourth ability group, where 14% (two persons) failed the item. Two persons in a higher ability group incorrectly answering #258, account for its misfit in this form.

Item #9 reads: "The answer to this problem is about how many?" with alternatives: 3, 40, 70, 80. Item #9 probably has more to do with introducing the extraneous skill of estimating. The high ability students may have been reluctant to "guess." They would tend to perform the calculation and then look for the closest answer, making the item more difficult than it was intended.

The two remaining misfitting items in this area require the student to recognize a "less than" symbol, Item #30, and

a "greater than" symbol, Item #32. The four items in this skill use the format: "In X < Y, what does "<" mean?" where X and Y are whole numbers and the alternatives are the same for all items: Is equal to, Is less than, Is greater than, Is not equal to. The other two items in this skill, recognizing an "equals" sign, Item #29, and an "inequality" symbol, Item #31, have between fit-t's greater than 3.0.

Items #30 and #32 are neither very hard nor extremely easy, but in the forms in which they misfit, persons in the higher ability groups tend to fail these items more often than those in the lower ability groups. These items also have the largest total fit-t's in the forms in which they misfit, their ICC's are flat and do not discriminate between ability groups. This suggests that knowledge of these symbols is not closely related to the other areas of math competency. High ability students may also have realized that one of the consequences of "less than" is "not equal to" and have selected that wrong alternative.

6.2.2 Arithmetic Computation

Six of the 73 items in this content area misfit. Three of these items, #38, #39, and #40, are from the adding or subtracting mixed numbers with like denominators skill. Item #39 is very difficult for students at this grade level. This item, "4 3/4 + 3 2/4" (aligned vertically) requires

TABLE 6.4

Misfitting Items from Arithmetic Computation

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
39	304	3.46	0.33	0.23	0.73	0.0	11.99	-0.21
	320	3.69	0.37	0.14	0.79	0.03	5.11	0.12
280	314	-1.69	0.38	0.1	0.52	0.05	7.77	0.15
38	303	0.24	0.26	0.11	0.03	0.15	5.97	1.62
44	314	-0.17	0.26	0.11	0.14	0.15	5.50	1.75
55	308	-0.53	0.26	0.21	0.51	0.12	5.21	1.10
40	311	1.57	0.24	0.39	1.04	0.05	4.13	0.34
	323	1.02	0.25	0.23	0.54	0.11	5.09	1.23

students to know the basic facts of adding mixed numbers, as well as how to simplify fractions, 5/4 = 1 1/4. This is the only problem in which students need to simplify. This extra requirement may have introduced a special interaction with experience, if all curricula did not introduce simplification at the same point.

Neither subtraction items #38 or #40 are particularly difficult, but both misfit because of the shape of their ICC's. For Item #38, "6 6/8 - 2 3/8" (aligned vertically) and with the alternatives: 4 3/16, 4 3/8, 8 9/16, 8 9/8, the lowest ability group did as well as the highest, suggesting there may be a naive way to succeed on the item or a clever

way to fail. On Item #40, "5 2/3 - 1 1/3" (aligned vertically) with alternatives: 4, 4 1/3, 4 2/3, 41, the middle ability groups had a higher failure rate than the lowest group.

Item #280, "243 + 55" with alternatives: 289, 212, 887, 298, from the vertical addition of whole numbers without renaming, is the only item of this type which misfits.

The only horizontal addition of decimal fractions problem, Item #44, "3.6 + 1.2 + .4" with alternatives: .520; 4.14; 5.2; 8.8; misfits in one of the forms in which it occurred and has the highest total fit-t in Form 314. Most of the misfit for this item was due to the fourth and fifth ability groups. One hypothesis is that these moderateley able students misread the decimal point in front of the "4" and chose distractor "8.8" as the answer. It is the only item in this skill in which the number of digits to the left of the decimal point are unequal. Another possibility is that renaming on a horizontal problem causes trouble for some students.

The last misfitting item in arithmetic computation, #55, is from the estimating whole number sums and differences skill. This item asks:

The answer to this problem is about how many?

```
 18
- 9
---
```

with alternatives: 1, 10, 20, 100. For this item, the second and third ability groups fail more often than the other ability groups.

6.2.3 Arithmetic Comprehension

TABLE 6.5

Misfitting Items from Arithmetic Comprehension

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
109	321	1.47	0.26	0.23	0.54	0.12	9.15	1.16
	324	1.72	0.25	0.39	0.95	0.07	4.41	0.57
126	318	1.27	0.23	0.22	0.44	0.12	3.17	1.53
	322	1.57	0.25	0.01	-0.20	0.16	7.48	2.08
125	323	1.44	0.25	0.25	0.48	0.11	6.32	1.05

In this content area, three out of thirty-six items misfit. Two of these items, #125 and #126, are from the same skill, identifying equivalent fractions, which had a total of only four items. Both use the same wording: "Pick the fraction with the same value as X." In Item #125, X is 1/2 and its alternatives are: 6/12, 3/4, 2/6, 6/10. In #126, X is 1/5, with alternatives: 3/15, 2/8, 2/7, and 3/10.

Item #109, one of the four items in the identifying multiples of a given number skill, exhibits the most misfit. This item asks: "What are the first 4 multiples of 3?" with alternatives: 0, 3, 6, 9; 3, 4, 5, 6; 3, 6, 9, 12; 6, 9, 12, 15. It misfits in both forms in which it appeared, as well as having the largest total fit-t in Form 321. One hypothesis for this is that there is ambiguity about where to begin. The higher ability students may have thought the beginning with zero was more correct than the one starting with three.

6.2.4 Arithmetic Applications

TABLE 6.6

Misfitting Items from Arithmetic Applications

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
134	314	0.38	0.25	0.58	1.53	0.0	5.38	-0.76
139	306	1.50	0.25	0.41	1.04	0.0	5.24	-0.09

This content area had two of forty-four items with between fit-t's greater than 5.0. All of the items in this area involve giving a "story" problem to the students and then having them perform some operation with the "story."

Item #134, from the using problem analysis techniques skill, misfits, as does Item #139 from the using problem-checking techniques skill. Item #134 asks: "The sixth grade class is having a bake sale. There are 3 classes and each class made 75 cupcakes. How would you find out how many cupcakes they made altogether?" Its alternatives are: Add: 3 + 75; Multiply: 3 x 75; Subtract: 3 - 75; Subtract: 75 - 3. Item #139 asks: "Mr. Green bought 24 cases of beans. There are 12 cans in each case. How many cans did he buy?"

Answer:
```
  24
X 12
----
  48
 24
----
 288 cans
```

How would you check this answer?

These items are unusually difficult for the three lowest ability groups, yet unusually easy for the highest ability groups. The ICC percentages for Item #134 are: 25, 44, 25, 90, 100, 100. One hypothesis to explain this phenomenon is that the lower ability groups do not understand what is required to solve the items, or have learned to be afraid of this type of item. The higher ability groups know how to attack these items and find them almost trivial. The item's observed difficulty is neither extremely easy nor hard, because the two groups balance each other out. Item #139 shows the same pattern, only the

groups are not as disparate. These items have negative total fit-t's because the high groups answer correctly more often than expected and the low groups less often. This also produces a high index of discrimination. However, the discriminating power is in one narrow range of ability and may have more to do with affect toward this type of problem than with ability.

6.2.5 Expressions, Equations, and Formulas

TABLE 6.7

Misfitting Items from Expressions, Equations, and Formulas

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
178	309	1.10	0.26	0.14	0.20	0.13	8.50	1.42
185	320	2.38	0.26	0.14	0.34	0.12	6.56	1.22
187	312	1.42	0.24	0.53	0.53	0.10	5.97	1.05
188	307	2.08	0.28	0.25	0.59	0.09	5.09	0.61

In this content area, four of the seventeen items misfit. Item #178, from solving equations involving addition or subtraction, showed the most misfit. This item, "X - 6 = 3, X = ?" with alternatives: 3, 6, 8, 9, was easier for the lowest ability group than for the middle ability groups.

Three of the four problems in the evaluating simple algebraic formulas skill misfit. These three items, #185, #187, and #188 (shown in section 6.1.5), involve multiplication, whereas the fourth problem in this skill requires division. These items were the hardest in this content area. Their misfit was caused by persons in the middle ability groups tending to fail the item more often than those in the lowest or highest ability groups.

6.2.6 Intuitive Geometry

TABLE 6.8

Misfitting Items from Intuitive Geometry

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
254	312	1.09	0.24	0.30	-0.13	0.17	8.05	2.82
	316	1.11	0.24	0.03	-0.20	0.15	6.28	2.28
197	301	1.58	0.24	0.10	-0.02	0.14	5.05	1.87
	303	1.42	0.26	0.0	-0.17	0.17	7.09	2.37
198	309	0.85	0.26	0.05	-0.17	0.17	6.55	2.31
206	305	1.91	0.27	0.09	0.26	0.14	6.11	1.38
200	311	0.05	0.26	0.21	0.35	0.10	5.57	0.88
192	302	-0.31	0.27	0.17	0.12	0.16	5.36	1.78
	317	-0.10	0.26	0.21	0.51	0.13	3.48	1.32
253	310	2.22	0.26	0.13	0.12	0.11	5.28	1.03
209	304	2.25	0.25	0.18	0.26	0.16	5.18	2.15
	306	2.01	0.27	0.18	0.43	0.13	4.01	1.32
201	315	1.40	0.26	0.22	0.28	0.14	5.01	1.66

This content area had nine of its twenty-five items misfitting. This is the largest proportion of misfit of any area, suggesting a relatively weak relation to the other components of the bank. Three misfitting items were from the identifying basic two-dimensional geometric figures skill. In both forms in which Item #254 misfit it had the largest total fit-t, as well as negative discrimination indices. This resulted from the flat item characteristic curves - between 45% and 55% of the persons in each ability group correctly answered the item. Item #254 asks: "Which of these figures is not a rectangle?" The alternatives are: a horizontal rectangle, a square, a parallelogram, and a vertical rectangle. This item might appear to have two correct answers to some students. To answer the item correctly, you must know that a square is a type of rectangle. There might also be a problem with knowing what a parallelogram is, causing some to avoid this answer entirely.

Two items misfit in the identifying relationships between geometric figures skill: Item #192, identifying parallel lines, and Item #253, identifying perpendicular lines. In skill three, recognizing symmetry, two of the four items misfit: items #197 and #198.

Two of the four items in the last skill, #206 , identifying coordinates and #209, locations of points in the first quadrant of a coordinate plane, also misfit.

6.2.7 Measurement

TABLE 6.9

Misfitting Items from Measurement

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
211	302	1.99	0.29	0.20	0.62	0.02	11.00	0.09
	313	1.88	0.25	0.28	0.74	0.09	4.75	0.82
256	314	2.77	0.30	0.11	0.38	0.12	6.61	0.82
	321	3.02	0.34	0.02	0.05	0.14	10.23	0.77
229	323	2.38	0.29	0.36	1.08	0.0	5.12	-0.34

Three out of thirty-five items in this content area misfit. Two of the misfitting items were in the first skill, estimating and choosing the measure of familiar objects or distances. Item #211, "Which of these measuring units is the smallest?" with alternatives: millimetres, centimetres, decimetres, metres, requires knowledge of the metric system. The extremely large between fit-t associated with this item in Form 302 can be traced to its item characteristic curve. In the lowest ability group, 36% of the persons correctly answered the question, no one in the second and third ability groups answered it correctly, and only 12% of the fourth ability group correctly answered the item. Of the fifth and sixth ability groups, 58% and 56%, answer the item cor-

rectly. This suggests that there is a distraction that is overly attractive to moderately low ability students.

The remaining misfitting item is in the fifth skill, interpreting maps and scale drawings. Item #256 (shown in Section 1.2.7), involving measuring a distance on a map, had a between fit-t of 10.23 in Form 321. Almost no one in four out of the six ability groups answered this item correctly, while 30% or more of the persons in the second and fifth ability groups correctly answered it.

Item #229 from skill six, calculating with units of time, also misfit. It involves adding two time periods, 7 hours 50 minutes + 3 hours 35 minutes and requires renaming in the process. Its alternatives are: 10 hours 25 minutes, 10 hours 75 minutes, 11 hours 15 minutes, 11 hours 25 minutes. It was the hardest item in Form 323.

6.2.8 Interpreting Data from Tables and Charts

TABLE 6.10

Misfitting Items from Interpreting Data from Tables and Charts

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
248	315	-2.23	0.48	0.25	0.84	0.0	5.87	-0.65

One of the fifteen items in this content area misfit. The misfitting item, #248, was the easiest item in the interpreting picture graphs skill. This item misfit because one person in the highest ability group in Form 315 failed the item, probably as a result of carelessness.

Chapter VII

STATE OF CALIFORNIA READING ITEM BANK

A pool of elementary reading items was developed by the California State Department of Education as part of their Proficiency Assessment of Basic Skills project. These reading items were grouped into four basic content areas:

1. phonetic analysis (36 items);
2. structural analysis (61 items);
3. vocabulary (87 items); and,
4. comprehension (200 items).

The reading items from all areas, not just those from comprehension, were passage-related. This means that all items were derived from the language contained in the passages and therefore, all items appeared with their related passages in the test forms. There were 35 passages ranging in readability from the second grade level to the sixth grade level.

The approximate readability level of each passage was calculated on the basis of the Fry Readability Graph.[5] The approximate grade level of single words used in the vocabu-

---

[5]Fry, Edward. A Readability Formula That Saves Time. Journal of Reading, Vol. 2 (April, 1968), 513-16, 575-77.

lary items, was determined using the Core Vocabulary, Consisting of a Basic Vocabulary for Grades 1-8 and an Advanced Vocabulary for Grades 9-13.[1] The passages ranged in length from 75-150 words for the lower grade levels to 150-275 words for the higher grades.

The 384 items in this reading pool were organized into a web of 35 balanced parallel forms. Each test form consisted of two reading passages and from 16 to 28 items. The forms were administered to approximately 3200 elementary school children in grades four and six, with from 89 to 108 children taking each form.

The results of the form calibration are shown in Tables 7.1 and 7.2. Table 7.1 shows the form number and difficulty for each of the 35 forms, and the sample and calibration statistics. The sample statistics consist of the total number of persons who took each form, the number of persons measured, and the mean and standard deviation of ability. The calibration statistics include: the number of persons and items used in the item calibration, the minimum, maximum and standard deviation of the item difficulties. The mean ability and the minimum and maximum item difficulty are

---

[1] Core Vocabulary, Consisting of a Basic Vocabulary for Grades 1-8 and an Advanced Vocabulary for Grades 9-13. Educational Developmental Laboratories Research and Information Bulletin No. 5. Huntington, N.Y.: Educational Developmental Laboratories, n.d.

shown as they appear on the bank, that is, with the form difficulty added, rather than as calculated by FORCAL.

Table 7.1 shows that Form 107 was the easiest form. It was made up of one third grade passage and one fourth grade level passage. The most difficult was Form #130 with two passages from the sixth grade level. Table 7.1 also shows that the easiest item in the bank was in Form 124, with a difficulty of -4.00 and that Form 127 contained the hardest item with a difficulty of 3.80. The form with the most widely spaced item difficulties was Form 118, with an item difficulty standard deviation of 1.76.

Table 7.2 shows the form number and difficulty for the 35 forms and the mean and standard deviation of the fit statistics as calculated by FORCAL. The point biserial, discrimination index, and between and total fit-t are listed. This table shows that Form 111 exhibited the most between fit-t misfit, with a mean of 2.87 and a standard deviation of 2.78.

7.1 EXTREME ITEMS

7.1.1 Phonetic Analysis

This content area has 36 items spread over four skills:

1. decoding consonants;
2. decoding variant consonants (consonants that signal more than one speech sound, such as c, g, and s);

TABLE 7.1

Summary Characteristics by Form

Form		Sample				Calibration				
		Persons		Ability		Per Item		Difficulty		
Num	Diff	Tot	Meas	Mean	S.D.	Used	Cnt	Min.	Max.	S.D.
101	0.75	96	88	1.30	1.05	77	21	-0.55	2.17	0.68
102	-0.19	100	98	1.18	1.24	84	24	-2.79	2.61	1.30
103	-0.93	98	90	1.01	1.29	70	27	-2.89	2.31	1.22
104	-0.58	101	92	1.10	1.40	65	24	-2.95	1.71	1.14
105	0.36	97	93	1.37	1.16	81	22	-1.05	2.20	0.96
106	-0.16	103	98	1.13	1.09	88	23	-2.46	1.84	1.12
107	-1.14	102	77	0.59	1.15	51	22	-2.97	0.69	0.87
108	-0.90	101	92	0.60	1.16	72	22	-2.45	0.91	0.94
109	0.08	99	95	0.95	1.04	86	24	-1.45	1.86	0.84
110	0.54	99	97	1.17	1.18	83	22	-0.89	1.72	0.68
111	0.50	101	101	0.83	1.06	92	20	-1.27	2.80	1.19
112	-0.58	96	94	0.80	1.37	83	25	-2.62	3.04	1.44
113	-0.67	102	98	0.45	1.14	87	24	-2.93	1.26	1.09
114	-0.40	99	99	0.50	1.19	94	20	-2.48	2.12	1.30
115	-0.99	97	88	0.78	1.24	69	23	-2.40	2.26	1.17
116	-0.76	98	92	0.76	1.08	81	27	-2.57	1.70	0.94
117	-0.86	94	90	1.04	1.22	76	28	-3.91	1.51	1.32
118	-0.47	98	92	1.69	1.32	80	23	-3.63	3.26	1.76
119	0.58	95	92	1.53	1.23	79	21	-3.03	3.03	1.40
120	0.57	95	89	1.25	1.07	80	22	-0.57	2.80	0.83
121	0.67	95	95	1.36	1.11	81	19	-0.34	2.14	0.69
122	0.36	96	89	1.30	1.27	72	21	-1.57	1.70	0.93
123	-0.43	108	102	1.09	1.38	79	22	-2.16	2.20	1.23
124	-0.36	102	94	0.89	1.16	90	21	-4.00	2.50	1.68
125	-0.31	99	99	0.81	1.01	85	20	-2.29	2.35	1.28
126	-0.63	95	95	0.88	0.97	77	20	-2.67	1.70	1.01
127	0.12	95	95	1.33	1.20	87	19	-1.40	3.80	1.23
128	0.52	94	94	1.27	1.18	83	16	-0.68	3.45	0.98
129	0.92	95	95	1.44	1.15	85	16	-0.04	2.10	0.78
130	1.39	94	94	1.58	1.04	79	21	-0.04	3.11	0.79
131	0.97	89	85	1.31	1.03	80	23	-0.68	2.77	1.04
132	-0.12	92	92	1.32	1.02	81	20	-2.05	2.12	1.11
133	0.19	89	85	1.44	1.09	76	22	-2.19	2.14	1.32
134	0.95	95	92	1.57	1.08	83	24	-1.95	2.43	0.93
135	1.04	97	95	1.43	1.05	80	21	-0.76	2.26	0.87

TABLE 7.2

Summary Fit Statistics by Form

Form		Point Biserial		Discrimin		Between Fit-t		Total Fit-t	
Num	Diff	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
101	0.75	0.40	0.15	1.05	0.46	0.59	2.43	-0.13	0.86
102	-0.19	0.43	0.13	1.03	0.27	0.69	2.47	-0.23	0.59
103	-0.93	0.43	0.19	1.00	0.35	0.89	2.43	-0.51	0.78
104	-0.58	0.40	0.14	1.02	0.32	-0.27	2.41	-0.67	0.67
105	0.36	0.42	0.09	1.04	0.32	0.47	2.02	-0.22	0.73
106	-0.16	0.42	0.19	1.02	0.47	1.81	2.57	-0.15	1.04
107	-1.14	0.39	0.18	1.02	0.47	0.89	1.70	-0.57	0.73
108	-0.90	0.44	0.13	1.03	0.31	1.18	2.15	-0.28	0.63
109	0.08	0.41	0.12	1.04	0.34	1.00	2.28	-0.09	0.77
110	0.54	0.40	0.19	1.06	0.62	1.84	3.38	-0.28	1.20
111	0.50	0.41	0.15	1.07	0.46	2.87	2.78	-0.26	1.24
112	-0.58	0.44	0.17	1.01	0.35	0.74	2.61	-0.28	0.78
113	-0.67	0.41	0.13	1.04	0.33	0.92	2.02	-0.04	0.67
114	-0.40	0.47	0.10	1.05	0.31	2.30	2.73	-0.01	1.06
115	-0.99	0.43	0.19	1.01	0.42	1.01	2.08	-0.33	0.75
116	-0.76	0.43	0.12	1.03	0.29	0.41	1.73	-0.04	0.71
117	-0.86	0.42	0.13	1.06	0.39	0.49	3.16	-0.24	0.77
118	-0.47	0.40	0.12	1.03	0.29	0.11	2.68	-0.35	0.73
119	0.58	0.40	0.12	1.06	0.34	0.45	2.53	-0.35	0.95
120	0.57	0.42	0.13	1.04	0.35	0.85	1.79	-0.18	0.78
121	0.67	0.45	0.16	1.04	0.43	1.00	3.06	-0.14	1.00
122	0.36	0.46	0.12	1.06	0.36	1.02	2.25	-0.30	0.88
123	-0.43	0.43	0.15	1.03	0.40	1.15	2.54	-0.62	1.00
124	-0.36	0.43	0.13	1.05	0.37	1.23	2.87	-0.06	1.03
125	-0.31	0.42	0.13	1.05	0.33	1.82	2.11	0.02	0.85
126	-0.63	0.42	0.11	1.02	0.25	-0.45	1.78	0.04	0.70
127	0.12	0.45	0.11	1.03	0.20	0.74	2.66	0.01	0.60
128	0.52	0.48	0.14	1.05	0.33	1.69	1.70	-0.03	1.14
129	0.92	0.47	0.13	1.06	0.38	1.89	2.68	-0.11	1.10
130	1.39	0.39	0.15	1.06	0.49	1.09	2.54	-0.47	1.00
131	0.97	0.41	0.13	1.06	0.42	0.87	1.96	0.03	0.84
132	-0.12	0.44	0.13	1.03	0.30	1.51	1.76	-0.00	0.96
133	0.19	0.38	0.15	1.01	0.36	0.99	2.61	-0.09	0.50
134	0.95	0.39	0.12	1.07	0.39	1.12	1.59	-0.09	0.78
135	1.04	0.40	0.18	1.07	0.58	2.10	2.81	-0.58	1.30

3. decoding vowels; and,

4. decoding spelling patterns (common letter and sound combinations used in forming words).

The items in this content area were evenly, yet widely spaced over six logits. Eight of the eleven items from the decoding consonants skill were the easiest in this area. The easiest item in this content area, #6, was the easiest item in one of the two forms in which it occurred as well as in the bank. It comes from a second grade level passage, #4, and asks: "The word that has the same sound as the "cl" in "clowns" is:" The alternatives are: clean, chip, cried, color.

The second easiest item, #7, was the third easiest item in the bank. Its difficulties in the two forms in which it appeared, however, were quite disparate. In Form 124 it was the easiest item by more than a logit, at -4.01, but in Form 123 it was the second easiest item with a difficulty of -1.97. It is not clear why these difficulties were so different, as the difficulties of both forms were approximately the same, -0.43 and -0.36, and the mean ability of the persons taking the forms was not very different. It might be due to sampling variation because the samples are relatively small and the item extreme. Item #7, from a second grade level passage, reads: "The word that has the same sound as the "dr" in "draw" is:", with the alternatives: didn't, drive, bring, and dinner.

Two items in phonetic analysis, #23 and #339, were much harder than the other items. These items were from the decoding vowels skill. Item #339, the most difficult phonetic analysis item, was the hardest item in one of the forms it which it appeared, 111, and was among the ten most difficult items in the bank. This item and Item #23, the hardest item in both of its forms, 114 and 115, have one alternative in common that seems to be the source of their extreme difficulty. Item #339 reads: "The word that has the same sound as the "e" in "problem" is:" with alternatives: ago, eat, out, ink. Item #23 is: "The word that has the same sound as the "ou" in "famous" is:" Its alternatives are: own, you, ago, and odd. The correct answer to both problems is "ago." However, "ago" contains two vowel sounds, whereas the other alternatives to these two items, as well as to every other question in this skill, contains only one vowel sound. Although the "a" in "ago" is the correct vowel sound, the dual vowel sounds created by the "a" and the "o" may cause confusion.

7.1.2 Structural Analysis

This second content area has 61 items divided among seven skills:

1. recognizing the meanings and functions of prefixes;
2. recognizing the meaning and functions of derivational suffixes (suffixes used to change a word from one part of speech to another);

3. recognizing the meanings and functions of inflectional suffixes (suffixes used to form plurals and possessives of nouns, to indicate verb tense, and to form the comparative and superlative of adjectives);

4. recognizing the two component words of compound words;

5. recognizing root words;

6. recognizing the infinitive form of irregualr verbs; and,

7. recognizing the meaning of contractions.

The items in this area were spread over seven logits, and there were three items at the extremes: two very easy items, #64 and #37, and one extremely hard item, #81. The easiest item in this area, #64, comes from the fourth skill. It was the easiest item in both forms in which it occurred, 118 and 119, and the second easiest item in the bank. It requires the student to recognize the two component words of "knock-out", as used in a sixth grade passage.

The nine items in the fourth skill were among the easiest in this area, and they showed a distinct pattern. The compound words in which it is possible to make words in addition to the two components were the hardest items. For example, the two component words in "eyelashes" are "eye" and "lashes." But the word can be split into "eyel" and "ashes." Although "eyel" is not a word, "ashes" is. The words in the easier items in this skill cannot be split in this way.

The other very easy item in structural analysis, #37, was also the least difficult item in each of its two forms, 102 and 103, and was among the five easiest items in the bank. Item #37 is from a passage at the third grade level and reads: "The "un" in "unhappy" makes the word mean:" The alternatives are: too happy, so happy, not happy, very happy.

The most difficult item in structural analysis, #81, was more than 1.5 logits harder than any other item in this area, and was the most difficult item in the bank by more than half a logit. This item requires recognizing the infinitive form of an irregular verb and it is easy to see why this item is extremely difficult. Item #81 reads: "The word "is" comes from:" and has the alternatives: in, it, see, be. All the other items in this skill have base words similar in sound and written form, such as "The word "fought" comes from:" with alternatives: fight, flow, fright, found.

7.1.3 Vocabulary

This content area consists of six skills:

1. recognizing general word meanings;
2. recognizing synonyms;
3. recognizing the meanings of multiple meaning words and homographs (words identical in written form but different in sound and meaning);

4. recognizing the meanings of homophones (words identical in sound but different in written form, origin, and meaning);

5. recognizing antonyms; and,

6. determining word meaning from context.

The 87 items in this area were evenly spaced over 5.5 logits. There were no items much easier than the others, but there were two extremely difficult items, #114 and #354. The most difficult item, #114, was the hardest item by more than 1.5 logits in one of the two forms in which it appeared, 118, and was the second most difficult item in the bank. This item is from the recognizing synonyms skill and uses a passage from the sixth grade. Item #114 reads: "In this story, "cunning" means:" The alternatives are: talent, cleverness, skill, decision. The passage is about a boxer and the sentence from the passage reads: "...,yet his skill and cunning allow him to win a decision on points." If the student had no prior knowledge of the meaning of "cunning," he would use the context of the sentence to determine the meaning and in so doing, might incorrectly choose "talent," rather than "cleverness."

The second extremely difficult item in vocabulary, Item #354, also concerns recognizing synonyms. This item was the most difficult item in Form 130 and the third hardest item in the bank. The passage Item #354 refers to is about wild burros and the item reads: "In this story, "hardy" means:"

The alternatives are: pleasant, difficult, tough, mean. As in Item #114, if the meaning of the word is not previously known, the student would probably find it difficult to choose the correct answer using the context of the passage alone. The other items in this skill provide alternatives in which the context of the passage does not have to be consulted to choose the correct answer.

7.1.4 Comprehension

The largest content area consists of 200 items divided among ten skills:

1. identifying specific details from a single sentence in a passage;
2. identifying specific details from two or three sentences in a passage;
3. using the reference system in connected discourse;
4. identifying the sequence of events described in a passage;
5. recognizing cause-and-effect relationships in a passage;
6. recognizing the main idea of a passage;
7. inferring meaning from information stated in a passage;
8. recognizing facts and opinions in a passage;
9. making critical judgements regarding the author's purpose or attitude; and,
10. making critical judgements regarding the ideas or information presented in a passage.

The items were spaced over a range of 5.5 logits, with no items much easier or more difficult than the others. The easiest item in this area was #305, the fifth easiest item in the bank. It requires inferring meaning from information stated in a third grade level passage.

The two most difficult items in comprehension also require inferring meaning from information stated in a passage. Item #318 was the most difficult item in both forms in which it occurred, 119 and 120, and Item #341 was the most difficult item in Form 312, by almost one logit. Both of these items were among the five hardest items in the bank.

Item #318, from a fifth grade passage about putting an unmanned aircraft on Mars, reads: "You can tell from the story that the snooper plane:" Its alternatives are: is not difficult to build, will find life on Mars, will scoop up samples of dirt, has not yet been built. The reason for its extreme difficulty is that it has no simple answer. Alternative four is the most correct answer, but given the context fo the passage, choice two is also a possibility.

Item #341 reads: "You can tell from the story that the old mouse:" and has alternatives: really thinks that the young mouse's plan is smart and simple; doesn't really think

the plan is smart, although he says it; wants the young mouse to bell the cat because it is his plan; does not think that the mice should harm the cat. The second alternative is the correct response. The reason it was extremely difficult is the nature of what must be inferred. The passage is from the fourth grade level, yet the students must infer sarcasm and mockery in the old mouse, because he does not mean what he says. This subtlety will be lost on many fourth and sixth grade children.

7.2 MISFITTING ITEMS

The analysis presented here is subjective. Subject matter specialists should use these explanations as examples of how to use the statistics and should be encouraged to formulate and investigate their own hypotheses.

7.2.1 Phonetic Analysis

There were five items with between fit-t's greater than 5.0. The item with the largest between fit-t was #11, a decoding consonants item which reads: "The word that has the same sound as the "gr" in "gray" is:" The alternatives are: grass, glass, glad, gas. Item #11 was the easiest item in Form 133, the form in which it misfit. Only two persons answered this item incorrectly in this form, one in the lowest ability group and one in the highest ability group. This high ability person could only have failed such an easy

TABLE 7.3

Misfitting Items from Phonetic Analysis

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
11	133	-2.19	0.60	-0.08	0.06	0.0	10.05	-0.17
339	112	2.07	0.28	0.10	0.11	0.17	6.81	1.85
23	115	2.28	0.34	-0.02	0.06	0.0	6.56	-0.35
22	114	0.73	0.25	0.35	0.55	0.13	5.69	1.51
26	125	0.02	0.26	0.53	1.26	0.0	5.41	-0.23

item as a result of carelessness. It was this single failure which caused the extreme between fit-t misfit, because it was a very unlikely event according to the model. A general discussion of the effects of carelessness, random guessing, etc. is given in Chapter 6.

The other four items in phonetic analysis with between fit-t's larger than 5.0 are from the decoding vowels skill: #339, #23, #22, and #26. Items #339 and #23 (see 1.1.1) were the hardest items in this content area. This set up their misfit. The item characteristic curves for Item #339 in Form 112 and for Item #23 in Form 115, were almost flat, indicating that the persons in the lower ability groups answered about as many items correctly as the students in the higher ability groups. Because only three or four stu-

dents in each ability group answered the question correctly it is reasonable to suspect that most of the successes were due to random guessing. If this is the case, these items are not useful for measurement at this level.

Neither Item #22, "The word that has the same sound as the "o" in "monkey" is:" nor Item #26, "The word that has the same sound as the "o" in "moving" is:" were as difficult as items #339 or #23. The alternatives for Item #22 are: up, out, all, old; for Item #26 they are: shot, show, shut, shoe. The item characteristic curve of Item #22 in Form 114 showed that students in the two lowest ability groups answered the item correctly more than expected and that the students in the third and fifth ability groups had a surprising number of incorrect answers. This could be due in part to an overestimate of the item's difficulty due to the position of Item #22 as the first item in Form 114.

In Form 125, Item #26 received too many correct answers from the first ability group, whereas the second and third groups produced more incorrect answers than expected. Additionally, all the students in the fifth ability group answered the item correctly which was two persons more than expected. This pattern would occur if the very lowest ability students guessed randomly while the slightly more able students tried to solve the item and failed or if they omitted it entirely.

7.2.2 Structural Analysis

TABLE 7.4

Misfitting Items from Structural Analysis

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
47	121	2.11	0.26	-0.04	-0.38	0.21	11.12	2.98
80	124	1.24	0.24	0.15	-0.06	0.17	7.76	2.57
58	103	-1.02	0.35	0.07	0.10	0.20	7.21	1.42
86	106	-1.07	0.33	0.11	0.47	0.15	6.10	0.97
76	111	-0.31	0.26	0.54	1.44	0.0	5.86	-1.59
81	127	3.80	0.34	0.15	0.68	0.19	5.86	1.23
	128	3.45	0.33	0.14	0.44	0.20	5.01	1.62
54	108	0.21	0.27	0.30	0.43	0.11	5.85	0.87
77	113	1.21	0.26	0.17	0.29	0.13	5.27	1.39
92	132	-0.85	0.35	0.32	0.84	0.0	5.15	0.05
343	134	1.88	0.25	0.26	0.56	0.09	5.07	0.89

Ten items in this area had between fit-t's larger than 5.0. There was one misfitting item from each of the first four skills. Item #343, from recognizing the meanings and functions of prefixes, reads: "The "ir" in "irreversible" makes the word mean:" The alternatives are: more reversible, very reversible, not reversible, partly reversible.

This item misfit in Form 134 because the students in the two lowest ability groups answered the item correctly more often than expected, but those in the fourth and fifth ability groups answered the item less often than expected. The "partly reversible" distractor may have been more attractive to the higher ability students than to the lower ability students.

Item #47 in Form 121 exhibited the most misfit in this area. In addition to an extremely large between fit-t in this form, Item #47 was the most difficult item, had a negative discrimination index, and a large total fit-t of 2.98. This item is from the second skill and reads: "The "er" in "laborer" makes the word mean:" The alternatives are: less labor, more labor, the job of laboring, a person who labors. The item characteristic curve for Item #47 showed that more than 30% of the students in the two lowest ability groups had too many correct answers and about 25% of those in the two highest ability groups had too few correct answers. This pattern might be interpreted as indicating the higher ability students confused the "er" ending with the comparative form of adjectives, thus choosing "more labor." Lower ability students did not have enough knowledge to be misled in this way.

Item #54 from the third skill, recognizing the meanings and functions of inflectional suffixes, reads: "The "s" in "results" makes the word mean:" The alternatives are: result, more than one result, belonging to the result, result is. The item characteristic curve for this item in Form 108 was: 33, 25, 80, 45, 100, 60. The first, third, and fifth ability groups had too many correct answers, and the second, fourth, and last groups too few. The erratic ICC suggests that the persons in the lower ability groups may have been guessing, and the others confused by the possessive distractor "result is."

Item #58, from the fourth skill, involves recognizing the two component words of compound words. The item reads: "The two words in "butterfly" are:" with alternatives: but + terfly, butt + erfly, butter + fly, butterf + ly. In Form 103, this item proved to be easier for the three lower ability groups than for the three higher groups. Although two words can be made from only one alternative, two of the distractors for this item, but + terfly and butt + erfly, contain one legitimate word. Some of the higher ability students may have failed to read all the alternatives carefully before answering.

Four of the items with large between fit-t's, items #76, #77, #80, and #81, come from the sixth skill, recognizing

the infinitive form of irregular verbs. Item #80, "The word "made" comes from:" with alternatives: mate, make, mad, many, was not extremely easy or difficult, but was the worst fitting item in Form 124. Its negative discrimination index reflects its flat item characteristic curve: 44, 43, 33, 50, 50, 64. All groups appear to have been able to eliminate one or two alternatives and then were unable to distinguish among the others.

Item #81 (see 1.1.2), was the hardest item in the structural analysis area, and the most difficult item in both forms in which it appeared, 127 and 128. Examination of the item characteristic curve in Form 127 showed no more than four persons answered the item correctly in any group, and only two persons in the lowest ability group answered it correctly. This was almost certainly due to lucky guesses. Item #81 was not quite as difficult in Form 128, but the misfit was still related to its difficuly. Four of the six ability groups had too many incorrect answers.

Item #77 was not a very difficult item in Form 113, but it had the largest between and total fit-t in this form. It reads: "The word "rode" comes from:" Its alternatives are: road, ride, role, rodeo. Item #77 did not discriminate well between ability groups, as indicated by its relatively flat item characteristic curve: 8, 36, 38, 23, 30, 42. The

greatest contribution to the between fit-t misfit came from the second and third ability groups where the students answered the item correctly more often than expected. This pattern suggests some influences of random guessing although their probability was some confusing homonyms.

The only item with a large negative total fit-t, yet between fit-t greater than 5.0, was Item #76: "The word "hung" comes from:" The alternatives are: huge, hang, hurt, high. Although this item was the only easy item of this skill six misfitting items. only 7% of the students in the lowest ability group answered correctly. Not only was this 41% less than expected, it was well below the chance level.

The remaining items in structural analysis with large between fit-t misfits are #86 and #92, both from the seventh skill, recognizing the meaning of contractions. Item #86, from a fourth grade passage, reads: "In this story, "doesn't" means the same as:" with alternatives: does to, did not, does not, do to. Item #92, based on a fifth grade passage, reads: "In this story, "don't" means the same as:", with alternatives: do no, do not, did not, does not. The source of the misfit for both Item #86 in Form 106 and Item #92 in Form 132 was a high ability group in which too few students responded correctly to the problem. As these items were relatively easy for the other ability groups, careless-

ness was probably the reason for too many incorrect answers in the high ability groups. For Item #92, this may be due to the distractor "do no", which might have been read quickly as "do not."

7.2.3 Vocabulary

TABLE 7.5

Misfitting Items from Vocabulary

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
357	110	1.72	0.25	-0.23	-1.00	0.21	14.03	3.39
112	117	-0.59	0.32	0.26	0.30	0.12	9.77	0.71
164	124	0.67	0.25	0.29	0.41	0.13	7.99	1.60
102	123	0.37	0.27	0.09	-0.05	0.18	7.61	2.19
114	118	3.26	0.32	0.11	0.30	0.08	7.41	0.35
	119	2.72	0.28	0.12	0.29	0.13	5.86	1.26
122	104	-0.30	0.32	0.16	0.30	0.15	6.48	1.04
107	106	-0.03	0.26	0.15	0.32	0.16	6.34	1.87
	107	0.22	0.31	0.18	0.13	0.09	5.21	0.50
354	130	3.11	0.29	0.07	0.12	0.14	5.89	1.17

This content area had eight items with between fit-t's greater than 5.0. Of these, five were from the second skill, recognizing synonyms: #357, #112, #114, #107, and

#354. In Form 110, Item #357 was the hardest item, had a negative discrimination index, and the largest between and total fit-t statistics. This item is: "In this story, "uninhibited" means:" with alternatives: unrestrained, graceful, unattended, timid. The extreme fit statistics for Item #357 stem from its item characteristic curve, which is the reverse of what is expected. The percentage of persons answering Item #357 correctly in each ability group were: 60, 50, 54, 21, 36, 24. The lowest ability group had 46% more correct answers than expected and the highest ability group 49% more incorrect answers than the model expected. For this item, the students in the lower ability groups did much better than expected and those in the higher groups much worse. One hypothesis is that the lower ability students selected the correct answer without consulting the passage, whereas the higher ability students substituted each alternative into the sentence: "They are very enthusiastic and competitive and completely uninhibited as they dive in the grass for batted balls." Upon substituting the four alternatives the higher ability students may have considered "graceful" or "unattended" more appealing choices for that sentence.

The next worst fitting item from the vocabulary area was Item #112, "In this story, "silent" means:" The alternatives are: happy, sure, afraid, quiet. The large between

fit-t came from the highest ability group, in which only 58% of the persons correctly answered the item, instead of the 94% expected. As this item was not difficult and because it was used in a third grade passage, carelessness could be the reason for a few high ability students failing this item too often. However, if the passage is examined, a different hypothesis emerges. The passage is a conversation between the seeds in a box in a store. The seeds are discussing what will happen when someone buys them. The relevant sentence in the passage is: "No one could answer the question, as they were all silent with worry." A high ability person might interpret this to mean that the seeds were afraid, and consequently incorrectly choose "afraid" as the more correct answer.

The third misfitting item in skill two was Item #114, recognizing a synonym for "cunning" (see 1.1.3). This item was the hardest item in this content area, the second most difficult item in the bank, and had the largest between fit-t in both forms in which it occurred, 118 and 119. In Form 118 where the misfit was greatest, the source of the misfit can be traced to more students in the lower ability groups correctly answering the item than was expected. This was also the case for Item #114 in Form 119, and is the same pattern that occurs with most extremely difficult items. If the high ability students substituted the word into the sen-

tence because they didn't know the answer, this guessing pattern would have resulted.

The fourth item from skill two which misfit was #107, "In this story, "swift" means:" The alternatives are: slow, clever, tame, fast. The primary source of misfit for the item in Form 106 was that the students in the two lowest ability groups answered the item correctly too often, yet those in the fourth ability group answered it correctly much less than expected. In Form 107 Item #107 was much more difficult and had a flat item characteristic curve, except in the second ability group in which everyone failed the item, causing the misfit indicated by the between fit-t statistic.

Item #354, the remaining misfitting item from the recognizing synonyms skill, requiring identifying a synonym for "hardy" (see 1.1.3), misfit to a much lesser degree than Item #114, but fit the pattern for an extremely difficult item. The lowest ability group had 30% too many correct answers and the second ability group had 14% too many.

The remaining three items in the vocabulary area, #164, #102, and #122, are from three different skills. Item #164 involves determining word meanings from context, skill six. This item uses a passage from the fourth grade and reads:

"You can tell from the story that "flit" means:" Its alternatives are: to sit still, to fight bravely, to move quickly, to begin slowly. The main source of misfit came from the lowest ability group in which there were too many correct answers. Instead of 56% correctly answering the item, only 18% were expected to do so. It is entirely possible that the lowest ability students guessed the correct answer, because none of the distractors were plausible substitutes for the test word "flit." The passage is about aikido, a Japanese way of protecting oneself and the relevant sentence is: "But your mind is still free to flit here and there like a butterfly."

One item from the recognizing general meanings skill, #102, had a large between fit-t, as well as a negative discrimination index and the largest total fit-t in the form in which it misfit, 123. Item #102 is from a fifth grade passage and reads: "In this story, "crew" means:" with alternatives: a group of people working together, a group of sailors, a group of passengers, a group of newspaper reporters. An examination of its item characteristic curve explains its negative discrimination index and large fit statistics. The percentage of persons answering Item #102 correctly in each ability group was: 43, 85, 64, 71, 50, 71. The greatest source of misfit came from the second and fifth ability group where there were 37% too many correct answers and 30% too many incorrect answers, respectively.

The last item which misfit in vocabulary was Item #122, from the recognizing multiple meaning words and homographs skill. This item is: "In this story, "bright" means:" with alternatives: shining, smart, poor, stiff. The correct answer is "shining." The item characteristic curve for Item #122 in Form 104 was relatively flat and thus the reason for its misfit. Too many students in the two lowest ability groups answered the item correctly more often than expected, whereas students in the fourth and fifth ability groups failed the item more than expected.

7.2.4 Comprehension

The fourth content area produced twenty items with between fit-t's greater than 5.0. Four of the misfitting items in this area, #376, #341, #323, and #310, are from the seventh skill, inferring meaning from information stated in a passage.

Item #376 had a between fit-t of 10.62, as well as a negative discrimination index and the largest total fit-t in Form 135. The percentage of persons in each ability group who answered Item #376 correctly in Form 135 was: 46, 67, 33, 31, 25, 45. Instead of the percentages increasing, indicating more success as the students become more able, the higher ability students fail the item more often than expected, while the lower ability groups correctly answered the item much more than expected.

TABLE 7.6

Misfitting Items from Comprehension

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
376	135	2.14	0.25	-0.14	-0.61	0.19	10.62	2.77
341	112	3.04	0.34	-0.05	0.00	0.12	9.99	0.58
382	111	0.88	0.24	0.04	-0.19	0.16	9.23	2.67
	110	0.80	0.25	0.09	0.07	0.14	5.00	1.74
381	101	1.46	0.25	-0.05	-0.36	0.18	9.07	2.50
	135	1.61	0.25	0.04	-0.09	0.14	5.42	1.19
333	111	1.58	0.24	0.37	0.82	0.06	8.88	0.35
264	117	-0.40	0.30	0.09	-0.11	0.18	8.85	1.68
206	114	-0.13	0.25	0.31	0.38	0.14	8.11	1.67
175	110	-0.10	0.28	0.65	1.89	0.0	6.95	-1.94
342	103	2.31	0.30	0.11	0.43	0.0	6.90	-0.30
	102	2.61	0.29	0.07	0.29	0.08	5.24	0.51
182	115	0.06	0.28	0.07	-0.06	0.18	6.73	2.09
380	131	1.72	0.25	0.10	-0.05	0.17	6.71	2.25
	130	2.47	0.26	0.09	0.03	0.15	6.52	1.76
232	129	1.86	0.25	0.19	0.20	0.17	6.53	2.23
262	114	1.20	0.25	0.33	0.69	0.12	6.48	1.24
323	125	1.85	0.26	0.23	0.45	0.15	6.31	1.59
198	129	2.10	0.26	0.38	0.75	0.09	6.10	0.77
195	127	-0.79	0.33	0.47	1.17	0.0	5.39	-0.64
330	122	1.67	0.28	0.22	0.46	0.13	5.41	1.18
279	119	1.69	0.25	0.26	0.53	0.11	5.29	1.06
310	106	1.66	0.25	0.17	0.19	0.14	5.26	1.82
334	116	1.72	0.27	0.26	0.70	0.04	5.10	0.17

Item #376, from a passage about drowning in cold water, asks: "You can tell from the story that the brains of the student and the physician:" The alternatives are: were not deprived of oxygen for more than four minutes, did not need oxygen, were exceptionally active, were not under water for more than four minutes. As the first alternative is correct, lower ability students may have selected it without reading the other three alternatives, whereas higher ability students may have read all alternatives and then were unable to distinguish between alternatives one and four, which are very similar in meaning.

Item #341 was the second most difficult item in this area and among the five hardest items in the bank. It is from a passage about a conversation between mice (see 1.1.4). This item was so difficult that its item characteristic curve in Form 112 was virtually flat. Three persons at most in any one ability group correctly answered the item.

Item #323, also from the seventh skill, is a difficult item in Form 125 and had the largest total fit-t in that form. Its item characteristic curve follows expectation, except for the lowest ability group, which is the source of misfit for this item. Given the item's difficulty, the expected percent of persons who answered Item #323 correctly should have been 6%, instead of the 36% who did so. This may indicate random guessing for this group.

Item #323 is: "You can tell from the story that aikido is:" with alternatives: newer than judo, less well known than judo, more dangerous than judo, more popular than judo. Although the answer to this item is the second alternative, "less well known than judo" some students may have thought the correct answer was "more dangerous than judo," because the passage states: "It is a way of protecting yourself,..."

The last misfitting item in skill seven was Item #310. It is from a sixth grade passage about building a tree house and reads: "Andy probably waited before building the tree house so that:" Its alternatives are: Ryan would not hurt himself trying to climb the tree; he would have time to plan how to build the tree house; he could decide which tree to build it in; Ryan would think he was the best builder in the world. The correct answer is choice one. Its almost flat item characteristic curve coupled with students in the highest ability group answering correctly more than expected, gave this item its large fit statistics.

The next worst fitting items in the comprehension area were #382 and #381, which involve recognizing facts and opinions in a passage, as do misfitting items #380 and #330. Items #380, #381, and #382 all have very low discrimination indices, between fit-t's greater than 5.0 in both forms in which they occur, and large total fit-t's. Each of these

items is of the form: "The statement, "...," is:" and have the alternatives: a fact, an opinion. These three items are the only two-alternative items in the bank and do not conform to the item stem or distractor characteristics for this skill. For each of these items, the lower ability students did better than expected while the higher ability students failed the item more than expected. This happened because some items from the secondary grade manual were included in this elementary grade test administration to provide linkage between the two. Some skills, however, such as recognizing facts and opinions, are not taught in the some way at the elementary level and therefore, the secondary items misfit when included in the elementary forms.

Item #330 is also from the eighth skill, but from the elementary manual. This type of item is of the form: "Which of the following statements from the passage is an opinion rather than a fact?" and four alternatives are given. The alternatives for Item #330 are: Orson Welles wrote a radio play; It seemed that the world had gone mad; Orson Welles was head of radio's "Mercury Theater"; The story describes Earth being attacked by a spaceship. The second alternative is the correct one.

This item was the hardest in Form 122, and also had the largest total fit-t. The source of this misfit was the

fifth ability group in which only 27% of the persons correctly answered Item #330, instead of the 61% expected. One hypothesis for this, since this is the last item in Form 122, is that many of the fairly high ability students who work slowly and carefully would not have reached it. These students would do less well on the end of the test than expected. Alternatively, it may be that anyone not believing in space ships would have thought that the fourth answer was opinion and the second one fact.

The fifth worst fitting item in comprehension was #333 from skill nine. This item and two other misfitting items, #342 and #334, involve making critical judgements regarding the author's purpose or attitude. Item #333 uses the passage about the mice and reads: "The writer of this story is trying to tell us:" with alternatives: to keep completely quiet until someone asks us to speak; what seems like a good idea at first doesn't always turn out to be; old people are always afraid of trying something that is new to them; it's always better to be safe than to be sorry. The seond alternative is the correct one. Item #333 was neither particularly difficult nor easy, yet the first ability group had far more correct answers than expected and in the fourth ability group, no one answered the item correctly. This was the source of the misfit for this item.

Item #342 is from a third grade passage about a hungry bull named Fred. The question is: "The writer probably wrote this story to:" with alternatives: make the reader laugh, make the reader sad, teach the reader a lesson, tell an important fact. This item was the most difficult item in both forms in which it occurred, 102 and 103, and misfit in both forms. In both, the item characteristic curve showed that the persons in the lower ability groups answered the item correctly more often than expected, while those in the higher groups failed the item more than expected. This passage is apt to make the reader laugh and so the lower ability students may have read no further than the first alternative and chose it as the correct answer that it is. However, the higher ability children may have chosen "teach the reader a lesson," incorrectly, because they know from past experience that even when a story is simple and silly it usually has a moral.

Item #334, also from skill nine, uses the third grade passage about the box of seeds in a store, and reads: "The writer probably wrote the story to tell us:" The alternatives are: if we listen carefully we can hear flowers talk; to take better care of our seeds and flowers; not to worry about what will happen to us later; not to believe what others tell us. The correct answer is choice three. This item was the most difficult item in both of the forms in which it

appeared, but had a large between fit-t in Form 116. Given the difficulty of the item, the students in the second ability group performed better than expected, introducing between group misfit into responses. This may have been due to guessing, as the other fit statistics are typical for this item. It may also have been due to high ability students missing the item consistently because they did not expect a story to tell them not to worry about their future.

The two misfitting items in comprehension, #264 and #262, come from the fourth skill. In Form 117, the form in which Item #264 had a large between fit-t, it also had a negative discrimination index and the largest total fit-t. This item asks: "What happened first?" and the alternatives are: the ringmaster started talking, the boys went inside the tent, the people were quiet, there was a real bear on a chain. The second alternative is the correct one. Item #264 misfit because too many students in the second ability group answered the item correctly and not enough students in the fifth ability group did so. One hypothesis is that some students may have thought there was no correct answer, as the passage states: "The parade was first." and there is no alternative for this.

Item #262, from a fifth grade passage, asks: "What happened after the tube wilted rapidly?" The alternatives are:

Haynes said "Lemme see that tube."; He mended the hole he'd made; They took the tube out for inspection; They took the wheel off. The second alternative is the correct answer. This item had a large between fit-t because only 11% of the fourth ability group answered the question correctly, instead of the expected 42%. These students probably failed due to carelessness, as the rest of the item characteristic curve was close to expectation and its other fit statistics were reasonable. This could have resulted if these students assumed that the first time the tire was flat, it "wilted rapidly." In this case, choice four would have been the correct answer.

The next item with an extremely large between fit-t and the largest total fit-t in Form 114 was Item #206. This item is from skill ten, making critical judgements regarding the ideas and information in a passage and is from the same passage as Item #262. The item is: "How do you think the writer felt about Haynes' method of fixing the tire?" and the alternatives are: He was disappointed about it; He didn't like it because it was unscientific; He was pleased with it because it worked; He thought he could have done a better job. The correct answer is the third alternative. The main source of misfit for this item in this form was the first and last ability groups, in which 25% too many persons answered correctly and 23% too many persons failed the item.

The students in the lowest ability were probably guessing and those in the highest ability were probably careless.

Of the remaining six misfitting comprehension items, four involve identifying specific details from a single sentence: #175, #182, #198, and #195. Although Item #175's item characteristic curve increased as it should, it did so with a sudden shift. The percentage of persons correctly answering this item in each ability group were: 20, 30, 92, 100, 100, 100. Thus, there were too many incorrect answers in the first two ability groups and too many correct answers in the other four higher ability groups. This item reads: "This story says Douglass is still remembered because:" and the answer is the fourth alternative: he worked to abolish slavery and gain equal rights. The other alternatives are: he was a slave during the time of the war between the states; he ran away from a cruel slavemaster; he taught himself to read and write. The answer to this item is given in the last line of the passage. It may be that lower ability students did not read through the entire fifth grade passage before answering the item, and as a result, failed the item, whereas students who read through to the last sentence of the story would find this item easy.

Item #182, also from skill one, uses a second grade passage and asks: "Joe is:" with alternatives: a snake, a

pirate, the writer of the story, the writer's best friend. In the form in which #182 misfits, Form 115, the item has a negative discrimination index and the largest total fit-t. This is because its item characteristic curve is relatively flat and the students in the lowest ability group answered the item correctly more often than expected. At first glance, this item appears to have no correct answer. The story says Joe "is a rat" and "rat" is not one of the alternatives. Upon closer inspection of the passage, however, the entire sentence is: "My best friend, Joe, is a rat." thus making choice four the correct response.

Item #198 is the hardest item in the form in which it misfits, 129. This item was from a passage about bee breeding and reads: "What caused a calm type of bee to develop?" with alternatives: a scientist; the hot, dry climate of Africa; the mild climate of Europe; a beekeeper. The correct answer is alternative three. The reason for its misfit is that the students in the lowest ability group answered this item correctly too often. The other fit statistics for this item were reasonable, so guessing may be the cause for the misfit.

The last misfitting item from the first skill, #195, was a very easy item in Form 127 in which it misfit. It asks: "Why are rats able to gnaw through metal pipes?" with

alternatives: They have unusually sharp teeth; They are always hungry; They can wiggle through holes to find them; They can climb walls to find them. The correct answer is: They have unusually sharp teeth. This item was answered correctly by at least 90% of the persons in each ability group, except for the lowest group in which only 33% answered it correctly.

The remaining misfitting comprehension items are #232, from skill two and #279, from skill five. Item #232 involves identifying specific details from two sentences in a sixth grade passage. The item is: "The 26 queen bees that escaped were from:" and the alternatives are: Brazil, Africa, Europe, South America. The item characteristic curve was not even remotely as expected. This may be because this item has more than one simple answer. The relevant sentence in the passage is: "In 1957, a scientist in Brazil was studying some honeybees from Africa." Thus, although the bees were <u>from</u> Africa, they escaped <u>from</u> Brazil. This undoubtedly confounded answering the item. In addition to a large between fit-t misfit, Item #232 had the largest total fit-t in Form 129.

The last misfitting item in comprehension is #279 from the fifth skill, recognizing cause-and-effect relationships. This item had a large between fit-t because the students in

the lowest ability group answered more items correctly than those in the third ability group. Item #279 is from a sixth grade passage about a boxer named Carlos, and asks: "Why do many experts feel that Carlos is an ideal boxer?" The alternatives are: He has both the knockout power and the ability to win a decision on points; He was 21 years old when he first entered the ring; He is a welterweight who weighs between 136 and 147 pounds; Success in boxing was not his only goal. The correct answer is the first alternative. As it is the first alternative that is correct, lower ability students may not have read the other alternatives and just chose the first one they came to that made sense as a correct answer, correctly so. Higher ability students, on the other hand, may have been confused by the distractors and ultimately chose the wrong response.

Chapter VIII

HOUSTON MATHEMATICS ITEM BANK

The Houston Independent School District compiled a pool of mathematics items organized into twenty-seven objectives of sixteen items each. These objectives are:

1. Add whole numbers and solve simple word problems using addition of whole numbers.
2. Subtract whole numbers and solve simple word problems using subtraction of whole numbers.
3. Multiply whole numbers and solve simple word problems using multiplication of whole numbers.
4. Divide whole numbers and solve simple word problems using division of whole numbers.
5. Read and interpret charts and graphs.
6. Identify the average (mean) of a set of data.
7. Compare fractions and mixed numbers using <, >, or =.
8. Add fractions and mixed numbers.
9. Subtract fractions and mixed numbers.
10. Multiply fractions and mixed numbers.
11. Divide fractions and mixed numbers.
12. Compare two numbers.
13. Round decimal and whole numbers.
14. Add decimals and solve simple word problems using addition of decimals.
15. Subtract decimals and solve simple word problems using subtraction of decimals.

16. Multiply decimals and solve simple word problems using multiplication of decimals.

17. Divide decimals and solve simple word problems using division of decimals.

18. Identify fractions, decimals, and percents which are equivalent.

19. Solve simple word problems involving money including problems requiring the student to identify change from a transaction.

20. Find a given percent of a designated number (i.e., find the percentage when the base and rate are given).

21. Find the base or rate given the percentage and the rate or base respectively.

22. Evaluate simple numerical expressions and formulas, given the formula.

23. Solve an equation of the form $x + a = b$, $x - a = b$, $ax = b$, or $x/a = b$.

24. Identify the standard numeral for a number expressed as a power.

25. Choose an appropriate metric unit to measure length, capacity, and mass.

26. Convert from one unit of measure to another.

27. Solve problems involving scale drawings and map reading.

The 432 items in this pool were placed into a web of eight parallel forms, with 108 items in each form. Each item was used in two forms and each form was connected to two other forms making up the ring of links shown in Figure 8.1. In this figure, an 'x' indicates one link between two forms.

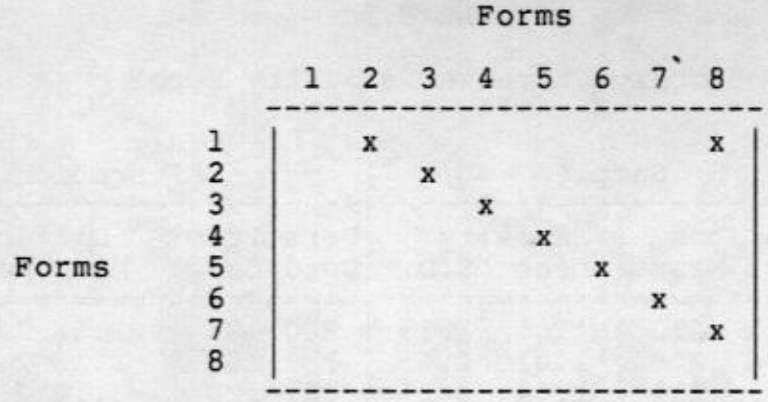


Figure 8.1: A Web for Houston Mathematics Items

The items were administered to 1735 seventh, eighth, and ninth graders in Houston, with from 206 to 232 students taking each form.

The results of the bank calibration are summarized in Tables 8.1 and 8.2. The first two columns of Table 8.1 show each form by number and difficulty. This is the difficulty of the form after all items have been shifted onto one common bank scale. It is called the "translation constant" in the SHIFT program. The next four columns show the sample statistics for each form. The total number of persons who took each form and the number of persons measured, i.e. the number who made a score other than zero or perfect, are shown along with the mean and standard deviation of ability of the measured persons.

TABLE 8.1

Summary Characteristics by Form

Form		Sample				Calibration				
		Persons		Ability		Pers	Item	Difficulty		
Num	Diff	Tot	Meas.	Mean	S.D.	Used	Cnt	Min.	Max.	S.D.
1	0.06	212	212	1.08	1.35	200	108	-2.68	2.58	1.11
2	-0.01	206	205	1.01	1.45	191	108	-3.15	2.79	1.20
3	0.10	211	211	1.04	1.27	207	108	-3.10	2.77	1.30
4	0.12	220	220	0.71	1.29	212	108	-2.37	2.37	1.15
5	-0.03	232	232	0.90	1.33	221	108	-2.80	2.65	1.18
6	-0.10	222	222	0.76	1.33	212	108	-2.83	2.30	1.09
7	-0.12	216	215	0.65	1.25	205	108	-2.82	2.15	1.12
8	-0.02	216	215	0.92	1.29	205	108	-3.24	2.23	1.22

The last five columns of Table 8.1 describe the calibration: the number of persons used to calibrate the items (i.e. those persons whose score was between the designated minimum and maximum scores); the item count (which is the number of items excluding zero or perfect item scores); and the minimum, maximum, and standard deviation of the item difficulties. Item difficulties have been shifted to the bank origin, that is, the form difficulty has been added to the local within-form item difficulty. The standard deviation is not affected by this transformation, so is the same as found in the FORCAL output.

Examination of Table 8.1 shows that the eight forms were all of approximately equal difficulty, with Forms 3 and 4 slightly harder than the others and Forms 6 and 7 a bit

easier. One person in each of Forms 2, 6, and 7 were deleted for making a zero or a perfect score. The mean ability of the measured persons shows that the most able persons took Form 1 and the least able Form 7.

The minimum and maximum score was set at 5% and 95% of the number of items and therefore the minimum raw score was 5 and the maximum was 102 for all forms. As a result, between 4 and 14 persons per form were not used in the calibration of the items, because their scores were less than the minimum or greater than the maximum. No items were answered either all right or all wrong by any persons, so 108 items were calibrated for each of the eight forms.

Table 8.2 shows the results of the fit statistics from the within-form item calibration. These are the number and difficulty of each form, and the mean and standard deviation of the point biserial, discrimination index, and between and total fit-t statistics. While the mean between fit-t and the standard deviations of both fit-t statistics are larger than we would like, all forms behaved in similar ways and the amount of misfit is not unusual for banks of this type. As will be seen in the subsequent sections, this level of disturbance does not prevent the construction of a useful bank.

TABLE 8.2

Summary Fit Statistics by Form

Form		Point Biserial		Discrimin		Between Fit-t		Total Fit-t	
Num	Diff	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	0.06	0.42	0.13	1.01	0.33	1.15	1.50	-0.17	1.78
2	-0.01	0.42	0.13	1.01	0.33	0.94	1.48	-0.14	1.70
3	0.10	0.42	0.15	1.01	0.37	1.34	1.79	-0.21	2.02
4	0.12	0.41	0.12	1.02	0.35	1.26	1.32	-0.19	1.81
5	-0.03	0.41	0.12	1.03	0.32	1.00	1.51	-0.18	1.74
6	-0.10	0.44	0.13	1.02	0.36	1.10	1.64	-0.18	2.04
7	-0.12	0.41	0.13	1.02	0.35	1.07	1.39	-0.17	1.82
8	-0.02	0.41	0.16	1.00	0.38	1.42	1.63	-0.24	2.11

8.1 EXTREME ITEMS

The next step in determining how well the items in a bank define the variable they are intended to measure is examination of the items' difficulties on the ITEMMAP. Extremely easy or extremely hard items on the bank may not be appropriate for inclusion with the other items. Also, the order of the items on this map should coincide with the test constructor's idea of the variable. When items are much easier or harder than educators think they should be, both the content of the item and the educator's preconceived ideas should be reexamined. As the Houston math bank is divided into objectives, the order of the items within each objective can be explored on individual ITEMMAPs of each of the twenty-seven objectives.

The sixteen items in the first objective, adding whole numbers, covered a range of only 1.5 logits. All the items on the objective were of comparable difficulty and the objective itself was extremely easy with respect to the entire item bank. Item #102, 2 + 4 + 3 + 8 + 7 + 1 (aligned vertically) with alternatives: 35, 25, 24, 15, was the easiest item in this objective and the least difficult item in the bank. It was the first item in both forms in which it occurred, Forms 2 and 3, and the easiest item in Form 2. It is one of four items requiring vertical addition of six single digit addends and its sum is the lowest of the four items. The only other item involving all single digit addends is #110, a word problem requiring the addition of three numbers, which was of about average difficulty on the objective.

The second easiest item from the first objective, #108, was the fourth easiest item on the bank, and the least difficult item in each of the forms in which it occurred, 6 and 7. It is one of two problems involving the addition of four addends: a single digit number, a two-digit one, a three-digit one, and a four-digit one. It reads: "3,824 + 7 + 13 + 692" (aligned vertically) and has alternatives: 3,426; 4,435; 4,536; 4,581. The other item of this type, #107, reads: "8 + 5,046 + 27 + 324" (aligned vertically) with alternatives: 5,415; 5,405; 5,385; 905. In

#108, all the places needed in the solution are in the first addend, at the top of the column of numbers. This makes adding each column and regrouping easier than in #107, in which the first addend uses only one of the four places needed.

Objective two, subtracting whole numbers, was spread over three logits with one very easy and one extremely difficult item. The easiest item in this objective, #202, was the easiest item in Form 3, and was the second easiest item in Form 2 and on the bank. It reads: "9,275 - 162" (aligned vertically) with alternatives: 7,655; 9,103; 9,113; 9,437, and is one of two items involving the vertical subtraction of a three-digit number from a four-digit one. However, Item #202 does not require renaming and hence, was easier than #206, "5,008 - 732" (aligned vertically), which does.

The most difficult item from the second objective was #210. It is a simple word problem and reads: "Beefy Burger sold 2,348 hamburgers last week. This week they sold 3,092 hamburgers. How many more hamburgers did they sell this week than last?" The algebraic formula for this problem is Y - X = ?, where X is the first number in the problem and Y the second number. Item #210 and the other word problem with the Y - X format, #209, are much harder than the two word problems in which the numbers could be put into the

equation X - Y = ?. Although Item #210 was almost a logit more difficult than any other item in objective two, it was of average difficulty with respect to the bank.

Both the third and fourth objectives, multiplying and dividing whole numbers, had a range of about two logits. The sixteen items in objective four had an average difficulty of almost one logit more than the items in objective three. No items in either objective were very easy or extremely difficult with respect to their objectives or the bank.

The item difficulties in objective three were divided into two distinct groups of items. The easier items consisted of: three non-zero one-digit factors written horizontally (4 items), problems with two non-zero two-digit factors aligned vertically (2 items), and items involving a multiple of 100 times a multiple of 10 (2 items). The harder items were: four problems in which the multiplicand (top number) is a three-digit number with a '0' only in the tens place and the multiplier (bottom number) is a non-zero two-digit number; and four word problems. The two easier word problems use one one-digit factor and one two-digit factor, and the two harder word problems consist of two two-digit factors.

Interpreting charts and graphs, objective five, was spread over three logits with one item that was extremely easy. Item #505 was the least difficult item in objective five, the easiest and second easiest item in Forms 1 and 8, respectively, and the third easiest item in the bank. The item consists of a bar graph and involves choosing the longest bar and identifying it from the vertical axis. The other item similar in format, #516, was more difficult and of average difficulty for the objective. It requires selecting the shortest bar. Item #505 was easier than #516 because the length of the longest bar was significantly longer than any other bar in the picture, whereas in Item #516 there are three short bars to choose from, none much shorter than the others.

The sixteen items in objective six, identifying the average (mean) of a set of data, had a range of less than a logit. All the items were clustered within a half a logit range, except for Item #608, which was slighltly easier than the others. This item asks: "Find the average (mean) of {100, 100, 200, 300, 800}." It is the only item in which all the numbers are multiples of 100.

The items comparing fractions and mixed numbers using <, >, or =, objective seven, spread over a little more than a logit, with item #707 slightly easier than the other items.

This item is: "1/4 ? 2/8", and has the same alternatives as all the items in this objective: <, =, >, always in that order. Only one other comparison of fractions has the correct answer "=", Item #703, "1/3 ? 3/9", which requires multiplication by three instead of by two.

Addition of fractions and mixed numbers, objective eight, had a range of 3.5 logits, because of two extremely easy items, #802 and #803. Item #802 involves addition of fractions, "4/9 + 1/9" with alternatives: 4/18, 5/18, 4/9, 5/9, and Item #803 consists of addition of mixed numbers, "3 2/7 + 11 1/7" with alternatives: 11 3/7, 14 3/14, 14 2/7, 14 3/7. All the items in this skill were written vertically. The fourteen other problems in this objective require: simplifying after addition, such as, 4/8 = 1/2; finding the lowest common denominator (LCD), 7/8 + 3/4; renaming using either the whole number or the fraction of the mixed number; or some combination of these formats, such as "27 2/5 + 16 1/3", which involves finding the LCD and renaming of the whole number parts.

Subtraction, multiplication, and division of fractions and mixed numbers, objectives nine, ten, and eleven, respectively, each covered a range of about two logits and none had any very easy or extremely difficult items in the objective or with respect to the bank. However, the item diffi-

culties in objective ten showed a definite pattern. The easier items were the six problems involving a simple fraction times a simple fraction, such as Item #1009, "1/2 x 3/4" with alternatives: 3/8, 1/2, 2/3, 1 1/2, and the harder items were those involving a mixed number times a simple fraction (four items) and a mixed number times a mixed number (six items).

Objective twelve, comparing two numbers, had a difficulty range of almost four logits due to one item, #1202, which was 1.5 logits easier than the other items in the objective. This item was one of the easiest items on the bank. The items in this objective are of the format X ? Y, where ? is either <, =, or >. The six items that involve the comparison of two whole numbers were the easiest. Of the six, Item #1202 is the only one that used the comparison X = Y. The two items with the format X > Y, #1204 and #1205, were the next easiest in difficulty and the three items in which X < Y, #1201, #1203, and #1206, were the most difficult whole number comparisons.

The remaining ten items in this objective showed a distinct pattern of item difficulties. The four items comparing a whole number to a decimal number greater than 1.0 were the next in difficulty, with the "=" comparison the easiest, the "<" next, and the ">" comparison the most difficult.

The other comparisons in order of difficulty from easiest to hardest were: a decimal less than 1.0 to a whole number, (.39 ? 5); a decimal greater than 1.0 to a decimal less than 1.0, (3.9 ? .39); two decimals less than 1.0, (.273 ? .7200); two decimals greater than 1.0, (2.93 ? 2.930) and (6.009 ? 6.09); and a whole number to a decimal less than 1.0, (45 ? .45).

Fifteen items in the thirteenth objective, rounding decimal and whole numbers, were evenly dispersed over 2.5 logits. The sixteenth item, #1314, was almost a logit more difficult than any other item in the objective and was the fifth most difficult item in the bank. It is one of three mixed numbers involving rounding and reads: "Round 27.324 to the nearest ten" with alternatives: 20, 27, 27.3, 30. The others are: Item #1315, "Round 567.438 to the nearest tenth" with alternatives: 570, 567.44, 567.4, 567.3; and Item #1316, "Round 3,872.64 to the nearest hundred" with alternatives: 4,000; 3,900; 3,872.64; 3,800. The correct answers to these items are: 30; 567.4; and 3,900, respectively. Only #1314 requires changing the left-most place of the whole number.

Objectives fourteen and fifteen, addition and subtraction of decimals, each had a difficulty range of about 3.5 logits. The items in the fourteenth objective showed two lev-

els of difficulty, with a one logit gap between the two sets of items. There are eight problems written vertically, four word problems, and four problems written horizontally. The easier set of items consisted of six of the eight problems aligned vertically: #1401, #1402, #1404-#1406, and #1408. These six problems involve the addition of two or three mixed numbers (4 items), one mixed number and one decimal (1 item), and two mixed numbers and one decimal (1 item).

Items #1403 and #1407 are the only two vertically aligned addition items which involve only decimal numbers. Item #1403 reads: "Add: .6293 + .804" with alternatives: 14,333; 1.4343; 1.4333; .14333; and Item #1407 is: "Add: .3980 + .6275 + .2184" with alternatives: .12439; 1.2439; 1.2529; 12,439. Students may not have been able to place the decimal point in these problems, making them more difficult than items that used mixed numbers, in which the decimal point is already placed between the whole number part and the decimal part.

A bimodal distribution of item difficulties was also evident in the subtraction of decimals objective. The easier items were the non-word problems in which the two numbers have either the same number of places to the right of the decimal point or those in which the minuend (first or top number) have one more digit to the right of the decimal

point than the subtrahend (second or bottom number.) The word problems and the items in which the subtrahend have one more digit than the minuend, were substantially more difficult. The hardest item in this objective was a word problem, #1513, and reads: "Dana ran around the track in 125 seconds and Pat in 111.42 seconds. How many more seconds did it take Dana to run around the track than Pat?" The alternatives are: 13.58, 14.42, 14.58, 236.42. This problem is the only one in which the subtrahend has two more digits to the right of the decimal point than the minuend.

The item difficulties in objectives sixteen, seventeen, and eighteen, multiplication and division of decimals, and identifying fractions, decimals, and percents which are equivalent, had a range of about 2.5 logits. The average difficulty of the latter two objectives was about a logit harder than objective sixteen. The items were evenly dispersed in all three objectives, except for objective sixteen, in which Item #1606, "36.6 x 5.7" (aligned vertically), was more than half a logit easier than any other item. This was probably due to the alternatives for this problem which were: .4392, 25.62, 43.92, 208.62. They are different enough from one another that the students could easily estimate the correct answer. The alternatives to each of the other problems contained the same digits and differed only in the placement of the decimal point, so

while these items are on the same objectives they use quite different skills.

The sixteen items of objective nineteen, word problems involving money, had a bimodal distribution of item difficulties spread over three logits. The twelve easier items, #1901 through #1912, had a range of less than one logit, and the four harder items, #1913 through #1916, had a range of one logit, with more than a one logit gap between the two sets. The twelve easier items involve making a purchase of one or more items and determining either the total cost of the items or the change received after calculating the total cost. These items require addition and/or subtraction or multiplication. For example, Item #1903 reads: "David bought a record album for $7.99 and a tape cassette for $9.29. If he gives the sales clerk $20, how much change should he get back?" The alternatives are: $37.28, $17.28, $3.28, $2.72, and the correct answer is $2.72. Another item, #1912 reads: "What is the total cost for 5 cans of tennis balls at $3.19 each?" with alternatives: $15.95, $15.55, $8.19, $.64. The correct answer is the first alternative.

The four harder items of objective nineteen all ask: "Which is the best buy?" To determine the correct answer, the unit price has to be calculated given the quantity and

quantity price of an item. For example, Item #1913 asks: "Which is the best buy?" with alternatives: 1 grapefruit for $.30; 3 grapefruit for $1.00; 4 grapefruit for $1.25; They are all the same. To solve each of these problems three division steps are required, rather than one addition, subtraction, or multiplication step as in Items #1901 through #1912. Also, the fourth alternative, which is the same for all four problems and reads: "They are all the same", invariably increases the difficulty of any item.

The sixteen items from finding a given percent of a designated number, objective twenty, were evenly dispersed over a narrow difficulty range of 1.5 logits. The average difficulty of this objective was very high, making it the second hardest objective in the bank. The hardest item, #2004, the fourth most difficult item in the bank, is "18% of 615 =" and has the alternatives: 110.7, 34.16, 11.07, .3416. All other items in this objective were nearly as difficult. In fact, eight items in objective twenty were among the most difficult 10% of all items on the bank.

Objective twenty-one also involves percents and was the most difficult objective in the bank. The item difficulties were evenly spread over two logits, and nine were in the hardest 5% of all items in the bank. In addition, the three most difficult items in this objective were the three hard-

est items on the bank: Items #2116, #2106, and #2102. Item #2102 is one of five items that uses the same format and reads: "X% of 112 = 28. Find X." The correct answer is determined by X = percentage/base, and is 25 in this example. Items #2106 and #2116 are of the format "rate of n = percentage. Find n." where n = percentage/rate.

The sixteen items in objective twenty-two, evaluating formulas, had a range of difficulty of more than three logits. The eleven easier items were spread over 2.5 logits and the remaining five items over 0.3 logits. These five harder items were among the 5% most difficult items in the bank. Four of these items require the squaring of a number to solve the problem. Items #2205 and #2206 involve finding the area of a square and use the formula $A = s^2$, where s is given in the problem. Items #2207 and #2208 involve finding the area of a circle, where $A = \pi r^2$, and π and r are given. The fifth hard item requires finding the area of a triangle, where A = (1/2)*b*h, and b and h are given.

Objective twenty-three, solving an equation for x, had item difficulties evenly distributed over more than four logits, and with a distinct pattern. The four equations of the form x + a = b were the easiest items with #2304, "x + 56 = 86", the easiest by almost one logit. The alternatives are: 29, 30, 56, 142. Item #2304 was also among the

easiest 3% of items in the bank. It may have been easier than the other three items that used this format because no regrouping was necessary. The hardest items in this objective, #2313 through #2316, use the format x/a = b. The remaining eight items of average difficulty are of the form x - a = b and ax = b.

The items requiring the calculation of n where n = x**a, objective twenty-four, had a bimodal distribution of item difficulties spread over two logits. Of the five easier items, the easiest, #2401, reads: "2^2 = n. Find n." with alternatives: 0, 1, 2, 4. The four other easy items all used the exponent "1". The eleven harder items have bases ranging from two to five, and exponents ranging from two to four.

The sixteen item difficulties of objective twenty-five, choosing the appropriate metric unit, were evenly dispersed over about two logits. Converting from one unit of measurement to another, objective twenty-six, was spread over more than three logits with two items, #2612 and #2614, almost one logit more difficult than the other items. These two items were also among the fifteen most difficult items in the bank. These items and one other, #2610, the next most difficult item in the objective, are the only problems which use decimal notation. For example, Item #2612 reads:

"Change 75.3 centimeters to millimeters. (10 millimeters = 1 centimeter)" with alternatives: 753 millimeters, 75.3 millimeters, 7.53 millimeters, .753 millimeters.

The twenty-seventh objective, solving problems involving scale drawings and map readings, had difficulties evenly dispersed over a range of more than three logits. The seven easier items all involve using coordinates on a map. Items #2702, #2708, #2709, and #2711 are all of the form: "Which place is located at (coordinates) on the map?" and items #2706, #2707, and #2710 are of the form: "Give the coordinates of the location of (place) on the map."

The nine remaining items show a scale map or a drawing and ask for the actual distance between two points given the scaled distance and the scale. For example, Item #2701 reads: "On the map of Wardsville, the distance from the town hall to the library is 2 inches. If 1 inch = 1/2 mile, what is the actual distance between the town hall and the library?" with alternatives: 1/4 mile, 1 mile, 1 1/2 miles, 2 miles. Of this type of problem, three items, #2701, #2714, and #2715 were much harder than the others. These three items all use the fraction "1/2" in either the distance or the scale, and there was confusion about multiplying or dividing by 1/2.

8.2 MISFITTING ITEMS

After the bank has been built, the items which exhibited misfit must be thoroughly explored in order to know whether they are suitable for inclusion with the other items in the bank. The source of the misfit is often miskeying or misprinting of the item. However, when the misfit is not due to such mechanical failures the item's stem and alternatives should be scrutinized to ensure that the process by which of the item is solved is appropriate for the operational definition of the variable the item was meant to measure.

The items with "between fit-t's" greater than 5.0 or "total fit-t's" larger than 4.0 were flagged as misfitting in this analysis. These cut-off levels resulted in twenty-three items being identified as misfitting out of the 432 in the bank. They were from only six of the twenty-seven objectives. These items are listed in Tables 8.3 through 8.8, ordered by the size of the between fit-t within objective. Each misfitting item is shown with the form number(s) in which it exhibited misfit, the difficulty and standard error in the form(s), and the point biserial, discrimination index, error impact, and between and total fit-t.

Three of the six objectives that had items with extreme misfit involve percents. The misfitting items in objective eighteen involve changing a decimal to a percent of equal

value, those in objective twenty require finding the percentage when the base and rate are given, and the misfitting items of objective twenty-one involve finding the base or rate given the percentage and rate or the percentage and base, respectively.

Two other objectives with misfitting items require the comparison of two numbers. Objective seven involves the comparison of fractions and mixed numbers and objective twelve the comparison of whole numbers, mixed numbers in decimal form, and decimal numbers. The remaining objective with misfitting items, twenty-two, involves evaluating formulas.

The items in Table 8.3 are examined within the objectives in which they occurred, because content similarity is often decisive in determining why particular items misfit. This in turn may result in a better understanding of how these topics are learned.

The sixteen items of objective seven are all of the format "Which symbol can be put in the box to make the sentence true?" X ? Y, with three alternatives <, =, >, always in this same order. Eight of the sixteen items compare two fractions and eight compare two mixed numbers. In the group of eight fractions, there are three in which the correct

TABLE 8.3

Misfitting Items from Comparing Fractions and Mixed Numbers

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
706	5	0.31	0.18	0.10	-0.14	0.19	6.71	6.17
	6	0.27	0.19	0.08	-0.19	0.22	6.54	6.63
708	8	0.36	0.18	0.21	0.20	0.14	4.38	4.22
	7	-0.19	0.18	0.14	0.13	0.15	4.23	4.66
710	1	0.33	0.19	0.19	0.17	0.15	4.23	4.42
702	1	0.27	0.19	0.16	0.22	0.16	4.13	4.68
	2	0.15	0.19	0.20	0.29	0.15	3.96	4.11
716	7	0.19	0.18	0.19	0.24	0.14	3.31	4.54

answer is "<", two where it is "=", and three in which it is ">." Of the eight mixed number comparisons, two have correct answers of "<," two have "=" as the correct response, and four have correct answers of ">."

All items in which the correct answer is "=" fit well, regardless of whether fractions or mixed numbers are compared. These items, #703, #707, #711, and #715, were of average difficulty on the bank and had item characteristic curves of the expected shape and slope.

From the eight items comparing fractions, three items misfit in both forms in which they appeared. Item #706 is one of three items in which the correct answer is "<" and

reads: "1/3 ? 3/8." This item was the worst-fitting item in Forms 5 and 6. Its between fit-t and total fit-t were greater than 6.0 in both forms, as well as having a negative discrimination index in both. The primary source of the difficulty was that too many persons in the higher ability groups failed to answer the item correctly.

There were two ways in which students could have answered these questions. They could have found the lowest common denominator (LCD) and converted each fraction or they could have "eyeballed" the item, that is, made an educated guess of which fraction seemed bigger or smaller. One hypothesis to explain the misfit of Item #706 is that it appears that the lower ability persons chose the precise LCD approach and the higher ability persons the less accurate, but quicker, "eyeball" method. If the students used the "eyeball" approach with this item, the fraction with two parts missing, that is 2 out of 3 in the fraction 1/3, may have seemed larger than the fraction with five parts missing, 3/8, and so the higher ability students may have chosen ">", rather than the correct answer "<".

The "<" fraction items that fit were easier to "eyeball": Item #701, "3/5 ? 7/8", and Item #704, "7/9 ? 6/7". These items could be estimated more easily because in each case, one fraction of the pair has two parts missing, 2 out of 5

and 2 out of 9, and one fraction has one part missing, 1 out of 8 and 1 out of 7. Therefore, it might be assumed that the fraction missing two parts is smaller than that missing one part. For Items #701 and #704 this is the case, but as illustrated for Item #706, the fraction missing two parts is less than the one missing five parts, thus duping the higher ability students into the wrong answer.

The same strategy holds for the fractions in which ">" is the correct answer. For misfitting items, #702 and #708, lower ability persons answered correctly more often than expected given the difficulty of the item, and higher ability persons incorrectly answered the items too often. Item #702 is: "4/5 ? 3/4" and Item #708 reads: "3/4 ? 2/3". In each of these items, both fractions have only one part missing, therefore, the "eyeball" approach will not work well. However, in Item #705 which fits, one fraction has three parts missing and the other five missing parts, "5/8 ? 4/9", thus, the "eyeball" method can be used to answer the item correctly.

Of the two mixed number items in which "<" is the correct answer, Item #710 misfit in Form 1 and marginally misfit in Form 2. Item #713 fit in both Forms 4 and 5. For Item #710 more lower ability persons answered the item correctly than expected. If the same strategy is assumed for mixed number

comparisons as for fractions, the "eyeball" approach does not work for #710, because 2 2/3 has only one part missing from its fraction, whereas 2 7/9 has two parts missing. Therefore, according to this method, 2 7/9 would seem smaller, although it is not. It appears that students who used the LCD method answered the item correctly more often than expected by the item's difficulty, but students who "eyeballed" the item had more failures than anticipated.

From the four mixed number comparisons in which ">" is the correct answer, Item #716 misfit in Form 7 and marginally misfit in Form 8, and Item #712 marginally misfit in Form 4. The same pattern holds for these misfitting items as the others and can be attributed to the same phenomenon. Item #716, "1 8/9 ? 1 3/4", has one part missing from each fraction making it difficult to correctly guess the answer. Item #712, "4 3/8 ? 4 1/4", has five parts missing from the larger fraction whereas three parts are missing from the smaller fraction, so the "eyeball" method would produce the wrong answer. Items #709 and #714 which fit reads: "1 1/2 ? 1 1/3" and "2 3/4 ? 2 3/5", respectively. In both of these items the fractions with the least number of parts missing, 1 out of 2 and 1 out of 4, is the larger number.

Items in objective twelve also use the format of comparison of two numbers and use the same item stem as in objec-

tive seven: "Choose the symbol that makes the sentence true. X ? Y" and the alternatives are: <, =, >, always in that order. Six items compare two whole numbers, of which three have the correct answer "<," one "=," and two ">." Four items compare a whole number to a decimal number greater than 1.0, two items compare a whole number to a decimal number less than 1.0, two items compare a pair of decimal numbers greater than 1.0, one item compares a decimal greater than 1.0 to a decimal number less than 1.0, and one item compares two decimal numbers less than 1.0. This lat item, #1208, was the only one of the sixteen items in this objective that misfit.

TABLE 8.4

Misfitting Items from Comparing Two Numbers

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
1208	3	0.76	0.19	0.17	0.14	0.18	4.30	5.25
	4	0.71	0.18	0.22	0.23	0.13	3.74	4.20

Item #1208 compares ".273 ? .7200" and the correct answer is <. It misfit in both forms in which it occurred, with the largest misfit of the 108 items in Form 4. It misfit because the persons in the lower ability groups correctly

answered the item more often than expected, while those in the higher ability groups made more incorrect answers than expected. The result was that in both forms, the fourth ability group had fewer correct answers than did the third ability group, and in Form 3 the fifth ability group also had fewer correct answers than the third group. One hypothesis to explain this is that the lower ability groups used the same rule for decimals less than one as for whole numbers, whereas higher ability persons may have thought there was some "trick" involved, as there would have been had they been comparing two negative numbers less than 1.0.

TABLE 8.5

Identifying Fractions, Decimals, and Percents

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
1804	6	0.64	0.19	0.19	0.18	0.18	4.27	5.26
1805	6	0.74	0.18	0.24	0.31	0.15	3.45	4.41

Two items in the eighteenth objective, identifying fractions, decimals, and percents which are equivalent, misfit. Item #1804 is: "Change .79 to a percent of equal value." with the alternatives: 79%, 7.9%, .79%, .79/100%. Item #1805 is: "Change .51 to a percent of equal value." and has

the alternatives: 51%, 5.1%, .51%, .51/100%. Of the six items involving changing a decimal to a percent, only Items #1804 and #1805 provide the distractor .xx/100%. For these items, which misfit in Form 6 of the two forms in which they appeared, the lower ability groups had more correct answers than expected and the persons in the higher ability groups more incorrect answers than expected, given the difficulty of the items.

The wording of the problems to change a decimal to a percent may have been confusing to some students. Some may have simply left the decimal unchanged and added the percent sign, thus "changing the decimal to a percent." Others, possibly the higher ability persons more so than the others, might have known that % means divide by 100, and as a result may have thought that is what was required and used the alternative .xx/100% in the only two items where it is provided as a distractor. The items involving changing a percent to a decimal (3 items), a decimal to a fraction (3 items), and a fraction to a percent (4 items), all fit in each of the two forms in which they occurred, probably because their meaning is less confusing.

The twentieth objective requires finding a percent of a given number and had two misfitting items. The format of the items in this objective is "A% of B =", where A was

TABLE 8.6

Misfitting Items from Finding a Given Percent of a Designated Number

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
2008	3	0.87	0.20	0.01	-0.41	0.27	7.59	7.62
	2	0.84	0.20	0.12	-0.02	0.20	5.23	5.77
2007	1	0.59	0.18	0.21	0.28	0.14	4.84	4.26
	8	0.76	0.18	0.22	0.28	0.15	3.41	4.37

always a two-digit whole number and B a three-digit whole number. Of the sixteen items in this objective, five are word problems, and of these five, two misfit, #2008 and #2007. Both items misfit in each form in which they appeared. Item #2008 had the largest between fit-t in Form 2 and the largest total fit-t in both Forms 2 and 3, and a negative discrimination index in both forms. Item #2007 had a between fit-t larger than 3.0 and a total fit-t greater than 4.0 on both Forms 1 and 8.

Although #2008 was easy within its objective, it was of average difficulty in its forms. Its severe misfit was caused by far too many persons in the two lowest ability groups correctly answering the item and not enough persons in the two highest ability groups doing so. Item #2008 asks: "The Simmons spend 12% of their weekly income on food. If their weekly income is $400, how much do they spend on

food?" The alternatives are: $33.33, $48.00, $333.33, $480.00.

The item characteristic curve for #2007 indicates the same pattern for low and high ability persons. Item #2007 asks: "Joni saved 25% by buying a $22 radio on sale. How much did she save?" with alternatives: $.55, $5.40, $5.50, $16.50. Although this was the easiest item in the objective, it was of average difficulty in Forms 1 and 8.

Items #2008 and #2007 exhibited the greatest misfit in this objective, however, fourteen out of the sixteen items in this form had total fit-t's greater than 2.0 in at least one of the forms in which they occurred, and eight of these in both forms. The only items which displayed no misfit were #2011 and #2012. Item #2011 asks for "60% of 120" in a word problem, and Item #2012 is: "50% of 212 =."

Eleven out of the sixteen items in another objective involving percents, objective twenty-one, had between fit-t's greater than 4.0 in at least one of its two forms. Additionally, fourteen of the sixteen items in this objective had total fit-t's greater than 2.0 in at least one form, and twelve of these had total fit-t's larger than 2.0 in both forms in which they occurred.

TABLE 8.7

Misfitting Items from Finding the Base or Rate

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
2102	3	2.77	0.23	-0.05	-0.05	0.19	8.59	3.53
2113	5	2.24	0.22	0.01	-0.01	0.23	6.90	4.46
	6	2.04	0.22	0.04	-0.03	0.21	6.62	4.40
2115	3	2.36	0.21	0.05	0.09	0.18	6.60	3.96
2104	6	1.73	0.21	0.05	-0.06	0.23	6.35	5.30
2101	1	2.27	0.21	0.10	0.15	0.17	5.90	3.55
	8	2.07	0.21	0.12	0.21	0.16	5.06	3.53
2110	7	1.91	0.22	0.08	0.11	0.19	5.68	3.74
	8	1.82	0.20	0.14	0.18	0.16	5.64	3.85
2108	5	2.02	0.20	0.17	0.38	0.14	5.61	3.00
2116	6	2.17	0.20	0.18	0.44	0.09	5.30	1.95
2103	5	1.29	0.19	0.12	0.04	0.19	5.27	5.23
2105	1	1.45	0.20	0.13	0.01	0.20	5.26	5.23
	2	1.14	0.19	0.22	0.22	0.15	3.81	4.16
2109	7	0.96	0.19	0.19	0.14	0.16	4.39	4.43
	6	1.24	0.19	0.23	0.33	0.16	3.90	4.35

Eight of the items in this form use the format: "X% of a = b. Find X." and eight are of the form: "a% of X = b. Find X." Three items of each type are word problems. One item from each format fit: Item #2111 is: "There are 40 students in a class, and 38 students are present today. What is the percent of students present?" with alternatives: .95%; 15.2%; 95%; 1,520%. This item can be

put into the format "X% of a = b. Find X." The other item which fit, Item #2106, reads: "40% of n = 70. Find n." with alternatives: 175, 110, 17.5, .005. These two items are not noticeably different than the other items in this objective and there is no clear explanation of why they fit when the others misfit. It appears that percent problems of this kind cause trouble for nearly everyone, producing misfit which cannot be attributable to one source. Probable factors contributing to misfit are: incorrect placement of the decimal point, computational errors, and most of all, confusion about when to multiply and when to divide.

TABLE 8.8

Misfitting Items from Evaluating Simple Numerical Expressions

Item Number	Form Number	Diff	S.E.	Point Biser	Disc. Index	Error Impact	Betw. Fit-t	Total Fit-t
2208	3	1.97	0.20	0.11	0.13	0.16	6.95	3.93
2216	8	1.60	0.20	0.06	-0.12	0.21	6.87	5.29

Item #2208 reads: "Find the area of the circle using the formula $A = \pi r^2$, when r = 10 miles. (Use 3.14 for π.)" The alternatives are: 31.4 sq. mi.; 62.8 sq. mi.; 103.14 sq. mi; 314 sq. mi.

Item #2216 reads: "Find the circumference of the circle using the formula $C = 2\pi r$, when $r = 13$ miles. (Use 3.14 for π.)" with alternatives: 40.82 mi.; 42.82 mi.; 81.64 mi.; 163.28 mi.

Chapter IX

USING ITEM BANKS TO BUILD VARIABLES AND MAKE MEASURES

9.1 CONSTRUCTION

9.1.1 Planning

An item bank is much more than a collection of test items. It is a composition of carefully written questions which pertain to a common theme in an orderly way. The more care that is devoted to the planning and writting, the more useful the bank will be and the more constructive the process of building it.

The first step, of course, is to establish the common theme. This must involve the potential users. Only they can determine which items must be included and which must not. One useful approach to this is a careful specification and ennumeration of the intended domain. The consequent deliberations can help clarify for everyone participating how they want their curriculum to be defined.

It is also useful at this stage to specify as precisely as possible how the items are expected to be ordered. If this is done prior to testing, it provides a judgmental frame of reference for later use in determining the utility and believability of the empirical ordering.

A major problem in every testing program is obtaining high quality items. Fortunately, there are already a number of extensive item banks developed by different school districts for use in a variety of curriculum areas (Hiscox, 1978). These banks may be a reasonable starting point. But it is almost certain that they will not fully meet the needs of a new group. Some items will not be useful in the new context and some important areas will not be adequately covered.

Some, perhaps all, items should be developed locally. While requiring more effort than simply using someone else's items, the gain in involvement and commitment of local educators is worth the effort. A bank which local educators are actively and personally involved in building will be more widely and wisely applied. The resulting items will more accurately represent the local curriculum and everyone will have a better understanding of what the bank can do for them.

9.1.2 Calibrating

The empirical ordering of the items requires that they be arranged in forms and administered to appropriate students. As the data analyses of the preceding chapters indicate, useful item calibrations can be obtained from relatively small samples. With these initial estimates in hand, the

bank can be put into immediate use, if it is monitored carefully and updated as necessary.

In planning a bank, the first questions that must be answered are:

1. How many items are to be calibrated?
2. How many can be administered to one student?
3. What range of difficulties can be given to one student?
4. How many forms can be managed?
5. How many students are available?

The answers to these questions are the parameters needed for designing a linking structure. The linking design should contain as many interconnections as possible so that the consistency of the links can be checked.

Actually obtaining item calibrations, once the forms have been designed, administered, scored and checked, is the easiest part of the process. The program FORCAL can be used to calibrate all items within each form by computing a difficulty for each item relative to the center of its form. This frees each item difficulty from the ability distribution of its calibrating sample. Then the program SHIFT can be used to calibrate each form by using the linking items to translate all difficulties to a common origin. This frees each item difficulty estimate from the difficulty levels of the forms in which it appeared.

9.1.3 Validation

Any assortment of items and forms can be processed through a series of computer programs to produce numbers. Whether or not these numbers have any meaning or represent the original intent of the planners can only be verified by careful inspection of the results. First, the calibrations should arrange the items in a reasonable order, one that agrees more or less with prior expectations. The first grade forms should be easier than the second grade forms; simple addition should be easier than multiplication. If the planning is done well and the expectations reasonable, this step should be a confirmation of what was already known and a reassurance that the process is working.

The next step is an inspection of the statistical analyses produced by the calibration programs. These statistics identify items which represent unique themes not shared by other items, items which suffered mechanical errors in their printing, administration, or processing and items which need revision. Irregularities of this kind are signalled by fit statistics lying outside the general distribution for the majority of items, regardless of how large or small the unusual values are.

If the general distribution is more diffuse than expected but has no obvious outliers, it can be interpreted as sig-

nalling that the system contains more noise than anticipated by the model. This does not mean that the bank cannot be used but it does mean that it should be used with more caution than would otherwise be necessary and that the standard errors of estimation are probably underestimates.

9.2 APPLICATIONS

9.2.1 Test Design

The presence of a carefully organized item bank makes test design more demanding. The versatility it offers enables us to do a much better job of choosing a test. As a result it is no longer defensible to pick a grade level test off the shelf and to administer it to everyone in the district. The teacher can do much better by considering who is to be measured and for what purpose. The bank provides the basis for designing the best possible test for every purpose. It is not necessary for everyone to take the same test to obtain comparable measures. Each student can, in principle, take the unique set of items most appropriate to his development and the reason for testing him. The length of the test, its level and range of difficulty, and the type and content of the items can be determined for each student individually rather than being frozen into a fixed set because of publishing considerations.

9.2.2 Person Fit

Applied psychometrics has concentrated almost exclusively on item analysis. This approach requires the assumption that, if items are developed carefully enough, they will always be appropriate for measuring anyone under any circumstances. This assumption is not justified in practice. Students sometimes guess randomly, make careless errors when tired or pressured, misunderstand instructions, and bring special experience that interacts with some items.

When disturbances like these occur, they can make the student's score a misleading indicator of his ability. To guard against this, a careful analysis of each student's pattern of responses should be a routine part of every testing session. Since the simple logistic model for the interaction between student and item excludes parameters for anything other than person ability and item difficulty, the analysis of residuals from this model provides a strong framework for detecting disturbances. When disturbances are detected, the teacher is then in a position to make an informed decision about the most appropriate action.

Because any subset of items from the bank can produce a comparable measure on the underlying variable, the teacher has a wide choice of possible actions available. The entire testing session can be repeated with another set of items

under better conditions. Or some of the items taken can be disqualified from scoring because they were too easy, too difficult or inappropriate in some other specifiable way. Which action is best in a particular circumstance will depend on what disturbed the original testing session. A detailed person fit analysis can help the teacher make the best choice by calling the problem to his attention and indicating its nature in a useful way.

9.2.3 Computer Assisted Learning

The simplicity of the simple logistic response model and the degree of data control it provides make it feasible and useful to implement a banking and testing package on an inexpensive microcomputer system. Selection of the best next question based on responses to preceding ones can reinstate many of the benefits of tutorial instruction. The questioning process is one-on-one. It can provide immediate feedback on success and failure, guidance toward better answers and pertinent instructional material. The incentive for the student can become improving his own competence rather than complying with authority or competing with peers.

Although the machine lacks the experience or intuition of a teacher to diagnose and adapt to unusual situations, the simple response model provides a precise statement of what is expected from each interaction. It can therefore recog-

nise surprising results and call them to the attention of student and teacher. With a calibrated item bank to draw on, it can handle much routine questioning and reporting thus serving as an assistant to the teacher (rather than a substitute).

9.2.4 National Item Banks

There is and always will be some interest in comparing performances among students. Some degree of competition is useful for stimulating maximum performance. It also helps to know how well others have done in order to judge if you are doing as well as you might. This makes a place for national item banks as supplements to local banks.

Nationally developed item banks, based on core curricula, could result from coalitions of local districts combining their banks and extracting the common elements. In addition to providing an objective basis for the empirical study of curricular structure, this process would have the benefit of provoking discussion of what the common curriculum might contain, thus extending and sharpening everyone's understanding of what is being taught.

An empirically demonstrated common curriculum could provide the basis for national comparisons. It could not, however, replace the local banks. While containing the poten-

tial for a national frame of reference, every test should still be locally controlled and relevant. The emphasis should be on understanding student performances rather than on ranking them.

The primary intent of the testing should be to assess competence rather than to select winners. As every administration of a test could be designed to obtain the most useful information about an individual student, the test would be able to make a fair and useful measurement and the student could not help but become a willing, perhaps eager, participant.

Appendix A

DESCRIPTION OF PROGRAM CONTROL CARDS

FORM DOCUMENTATION

Card Number	Card Name	Format and Description
1	Title Card	(20A4) Description of analysis
2	Parameter Card	(16I4)

Card Cols	Program Label	Description of Field Contents
1- 5	NITEM	Number of unique items.
6-10	ITPF	Number of items per form.
11-15	ITPL	Number of items per link.
16-20	INFLE	The unit number from which the item file is to be read. b, Default = 5

SAMPLE FORM JCL

```
//job card with Region=129K

//  EXEC PGM=FORM
    where FORM refers to the name of the load module where
    the FORM source deck is compiled

//STEPLIB DD DSN=$account.sub.dataset,DISP=SHR
    where dataset refers to the dataset where FORM
    is stored

//FT03F001 DD DSN=$account.sub.INPUTITEMFILE,DISP=SHR
    where INPUTITEMFILE refers to the name of the
    original item file

//FT10F001 DD DSN=$account.sub.OUTPUTFORMFILE,DISP=(NEW,CATLG),
//  UNIT=SYSDA,DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000),
//  SPACE=(TRK,(1,1),RLSE)
    where OUTPUTFORMFILE refers to the name of the
    form file output from FORM

//FT06F001 DD SYSOUT=A,DCB=BUFNO=1

//FT05F001 DD *
```

Sample JCL and Control Cards
For KCT FORM Run

```
//job card with Region=129K
//  EXEC PGM=FORM
//STEPLIB DD DSN=$account.sub.BANKPROG,DISP=SHR
//FT03F001 DD DSN=$account.sub.I0KCT,DISP=SHR
//FT10F001 DD DSN=$account.sub.F0KCT,DISP=(NEW,CATLG),
//  UNIT=SYSDA,DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000),
//  SPACE=(TRK,(1,1),RLSE)
//FT06F001 DD SYSOUT=A,DCB=BUFNO=1
//FT05F001 DD *
FORMS ANALYSIS FOR KCT5 BANK - FIVE FORMS - 40 ITEMS
   40   16    4    5
```

FORCAL DOCUMENTATION

Card Number	Card Name	Format and Description
1	Title Card	(20A4) Description of analysis
2	Parameter Card	(16I4)

Card Cols	Program Label	Description of Field Contents
1- 4	LRC	Length (in bytes) of the input person record. b, Default = 80
5- 8	ITPOS	Position in the person record of the first item score. b, Default = 1
9-12	IDLEN	Length of the identification field in the person record. b, Default = LRC
13-16	IDPOS	Start of the identification field in the person record. b, Default = 1
17-20	IFLEN	Length of the form number in the person record. b, Default = 1
21-24	IFPOS	Start of the form number in the person record. b, Default = 1
25-28	MINS1	The minimum percent correct necessary for a person to be included in the calibration sample. b, Default = 40
29-32	MAXS1	The maximum percent correct allowed for a person to be included in the calibration sample. b, Default = 80
33-36	MGROP	The minimum average group size for the fit analysis; can be used to control the number of score groups used in the item fit. b, Default = 25.
37-40	INFLE	The unit number from which the person file is to be read. b, Default = 5 (cards)

41-44	KCAB	Item calibration code: 1 = PROX 2 = UCON b, Default = UCON
45-48	KPFIT	The maximum total fit-t allowed for a person to be included in the recalibration sample. Read as F5.1. All persons with a total fit-t above KPFIT/10 will be eliminated from recalibration. b, Default = 99999
49-52	KCFIT	Controls the item status code. Read as F5.2. If the total fit-t of an item is greater than KCFIT/100, the item status code of that item will be changed to 0 and not used in the linking structure. b, Default = 99999
53-56	KKEY	Item scoring code: 1 = Score items against key as read from form file 2 = Items already scored 0/1 3 = Items coded alpha/integer with alpha = correct; integer = incorrect b, Default = 1
57-60	KSTAT	The minimum item status code (col. 80 of the form file) to be included in the analysis. b, Default = 1
61-64	IPFILE	The unit number to which the output person file will be written. If left blank, no person file will be written. b, Default = 0

3	Format Card	(20A4) Must be of the form: (nnA4) where nn = LRC/4
4	Output File Format Card	(20A4) Format to describe output person file. Must be of the form: (nA1,I4,4F8.3,mmI1) where n = IDLEN mm = the number of items in the longest form

SAMPLE FORCAL JCL

```
//job card with Region=257K

//  EXEC PGM=FORCAL
    where FORCAL refers to the name of the load module where
    the FORCAL source deck is compiled

//STEPLIB DD DSN=$account.sub.dataset,DISP=SHR
    where dataset refers to the dataset where FORCAL
    is stored

//FT20F001 DD SYSOUT=A,DCB=RECFM=FA

//FT01F001 DD UNIT=SYSDA,DISP=NEW,SPACE=(CYL,(1,1)),
//  DCB=(RECFM=FB,LRECL=xxx,BLKSIZE=yyyy)
    where  xxx = 36 + IDLEN + the number of items in the
                 longest form
           yyyy = some multiple of xxx

//FT02F001 DD UNIT=SYSDA,DISP=NEW,SPACE=(CYL,(1,1)),
//  DCB=(RECFM=FB,LRECL=xxx,BLKSIZE=yyyy)
    where  xxx = 36 + IDLEN + the number of items in the
                 longest form
           yyyy = some multiple of xxx

//FT03F001 DD DSN=$account.sub.INPUTFORMFILE,DISP=SHR
    where INPUTFORMFILE refers to the name of the
    original form file

//FT04F001 DD DSN=$account.sub.OUTPUTFORMFILE,DISP=(NEW,CATLG),
//  UNIT=SYSDA,DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000),
//  SPACE=(TRK,(5,5),RLSE)
    where OUTPUTFORMFILE refers to the name of the
    form file output from FORCAL

//FT06F001 DD SYSOUT=A,DCB=BUFNO=1

//FT17F001 DD DSN=$account.sub.INPUTPERSONFILE,DISP=SHR
    where INPUTPERSONFILE refers to the name of the
    original person file

//FT18F001 DD DSN=$account.sub.OUTPUTPERSONFILE,UNIT=SYSDA,
//  DISP=(NEW,CATLG),DCB=(RECFM=FB,LRECL=xxx,BLKSIZE=yyyy),
//  SPACE=(TRK,(10,5),RLSE)
    where OUTPUTPERSONFILE refers to the name of the
    person file output from FORCAL
     xxx = 66 + the number of items in the longest form
    yyyy = some multiple of xxx

//FT08F001 DD UNIT=SYSCR,DISP=NEW,SPACE=(TRK,(10,10))

//FT05F001 DD *
```

Sample JCL and Control Cards
For KCT FORCAL Run

```
//job card with Region=257K
//  EXEC PGM=FORCAL
//STEPLIB DD DSN=$account.sub.BANKPROG,DISP=SHR
//FT20F001 DD SYSOUT=A,DCB=RECFM=FA
//FT01F001 DD UNIT=SYSDA,DISP=NEW,SPACE=(CYL,(1,1)),
//  DCB=(RECFM=FB,LRECL=82,BLKSIZE=4100)
//FT02F001 DD UNIT=SYSDA,DISP=NEW,SPACE=(CYL,(1,1)),
//  DCB=(RECFM=FB,LRECL=82,BLKSIZE=4100)
//FT03F001 DD DSN=$account.sub.F1KCT,DISP=SHR
//FT04F001 DD DSN=$account.sub.F2KCT,DISP=(NEW,CATLG),
//  UNIT=SYSDA,DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000),
//  SPACE=(TRK,(5,5),RLSE)
//FT06F001 DD SYSOUT=A,DCB=BUFNO=1
//FT17F001 DD DSN=$account.sub.P0KCT,DISP=SHR
//FT18F001 DD DSN=$account.sub.P9KCT,UNIT=SYSDA,
//  DISP=(NEW,CATLG),DCB=(RECFM=FB,LRECL=104,BLKSIZE=4160),
//  SPACE=(TRK,(10,5),RLSE)
//FT08F001 DD UNIT=SYSCR,DISP=NEW,SPACE=(TRK,(10,10))
//FT05F001 DD *
KCT BANK - USING FORMS KCTA, KCTB, KCTC, KCTD - FEBRUARY 1, 1980
  80  13   8   1   3   9   5  95  10  17   0   0   0   1   1  18
(20A4)
(8A1,I4,4F8.3,38I1)
//
```

UPDATE DOCUMENTATION

Card Number	Card Name	Format and Description
1	Parameter Card	(5I5)

Card Cols	Program Label	Description of Field Contents
1- 5	NITEM	Number of unique items.
6-10	NFI	Number of times each item is used.
11-15	NF	Number of forms.
16-20	NIS	The lowest item number. b, Default = 1
21-25	NFS	The lowest form number. b, Default = 1

SAMPLE UPDATE JCL

```
//job card with Region=129K

//  EXEC PGM=UPDATE
    where UPDATE refers to the name of the load module where
    the UPDATE source deck is compiled

//STEPLIB DD DSN=$account.sub.dataset,DISP=SHR
    where dataset refers to the dataset where UPDATE
    is stored

//FT01F001 DD DSN=$account.sub.INPUTFORMFILE,DISP=SHR
    where INPUTFORMFILE refers to the name of the
    form file output from FORCAL

//FT02F001 DD DSN=$account.sub.INPUTITEMFILE,DISP=SHR
    where INPUTITEMFILE refers to the name of the
    original item file

//FT03F001 DD DSN=$account.sub.OUTPUTITEMFILE,DISP=(NEW,CATLG),
//  UNIT=SYSDA,DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000),
//  SPACE=(TRK,(1,1),RLSE)
    where OUTPUTITEMFILE refers to the name of the
    item file output from UPDATE

//FT06F001 DD SYSOUT=A,DCB=BUFNO=1

//FT05F001 DD *
```

Sample JCL and Control Cards
For KCT UPDATE Run

```
//job card with Region=129K
//  EXEC PGM=UPDATE
//STEPLIB DD DSN=$account.sub.BANKPROG,DISP=SHR
//FT01F001 DD DSN=$account.sub.F2KCT,DISP=SHR
//FT02F001 DD DSN=$account.sub.I1KCT,DISP=SHR
//FT03F001 DD DSN=$account.sub.I2KCT,DISP=(NEW,CATLG),
//  UNIT=SYSDA,DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000),
//  SPACE=(TRK,(1,1),RLSE)
//FT06F001 DD SYSOUT=A,DCB=BUFNO=1
//FT05F001 DD *
   40    4    4    1    1
```

SHIFT DOCUMENTATION

Card Number	Card Name	Format and Description
1	Title Card	(20A4) Description of analysis
2	Cut-Off Value	(F5.0) All items with a between form residual greater than this cut-off value are not used in the linking.
3	Parameter Card	(10I5)

Card Cols	Program Label	Description of Field Contents
1- 5	LREC	Length (in bytes) of the input item record. b, Default = 80
6-10	INFLE	The unit number from which the item file is to be read. b, Default = 1
11-15	IFPOS	Start of the form name in the item record. b, Default = 15
16-20	IFLEN	Length (in bytes) of the form name in the item record. b, Default = 12
21-25	NFPOS	Start of the form number in the item record. b, Default = 6
26-30	NFLEN	Length (in bytes) of the form number in the item record. b, Default = 3
31-35	MSTRT	The lowest form number in this run. b, Default = 1
36-40	KSTAT	The minimum item status code (col. 80 of the item file) to be included in the analysis. b, Default = 1

41-45	KOUT	Controls the writing of a new item and a new form file. b,0 = No item or form files will be written GT 0 = New item and form files will be written
46-50	KBUG	Controls the printing of intermediate results. b,0 = No intermediate results will be printed GT 0 = Intermediate results will be printed

SAMPLE SHIFT JCL

```
//job card with Region=129K

//  EXEC PGM=SHIFT
    where SHIFT refers to the name of the load module where
    the SHIFT source deck is compiled

//STEPLIB DD DSN=$account.sub.dataset,DISP=SHR
    where dataset refers to the dataset where SHIFT
    is stored

//FT06F001 DD SYSOUT=A,DCB=RECFM=FA

//FT10F001 DD SYSOUT=A,DCB=RECFM=FA

//FT08F001 DD SYSOUT=A,DCB=RECFM=FA

//FT09F001 DD SYSOUT=A,DCB=RECFM=FA

//FT01F001 DD DSN=$account.sub.INPUTITEMFILE,DISP=SHR
    where INPUTITEMFILE refers to the name of the
    item file output from UPDATE

//FT02F001 DD DSN=$account.sub.INPUTFORMFILE,DISP=SHR
    where INPUTFORMFILE refers to the name of the
    form file output from FORCAL

//FT03F001 DD DSN=$account.sub.OUTPUTFORMFILE,DISP=(NEW,CATLG),
//  UNIT=SYSDA,DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000),
//  SPACE=(TRK,(5,5),RLSE)
    where OUTPUTFORMFILE refers to the name of the
    form file output from SHIFT

//FT04F001 DD DSN=$account.sub.OUTPUTITEMFILE,DISP=(NEW,CATLG),
//  UNIT=SYSDA,DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000),
//  SPACE=(TRK,(5,5),RLSE)
    where OUTPUTITEMFILE refers to the name of the
    item file output from SHIFT

//FT05F001 DD *
```

Sample JCL and Control Cards
For KCT SHIFT Run

```
//job card with Region=129K
//  EXEC PGM=SHIFT
//STEPLIB DD DSN=$account.sub.BANKPROG,DISP=SHR
//FT06F001 DD SYSOUT=A,DCB=BUFNO=1
//FT10F001 DD SYSOUT=A,DCB=BUFNO=1
//FT08F001 DD SYSOUT=A,DCB=BUFNO=1
//FT09F001 DD SYSOUT=A,DCB=BUFNO=1
//FT01F001 DD DSN=$account.sub.I2KCT,DISP=SHR
//FT02F001 DD DSN=$account.sub.F2KCT,DISP=SHR
//FT03F001 DD DSN=$account.sub.F3KCT50,DISP=(NEW,CATLG),
//  UNIT=SYSDA,DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000),
//  SPACE=(TRK,(5,5),RLSE)
//FT04F001 DD DSN=$account.sub.I3KCT50,DISP=(NEW,CATLG),
//  UNIT=SYSDA,DCB=(RECFM=FB,LRECL=80,BLKSIZE=4000),
//  SPACE=(TRK,(5,5),RLSE)
//FT05F001 DD *
SHIFT ON KCT BANK - 0.50 LOGIT CUT-OFF - FEBRUARY 3, 1980
 0.50
   80    1   15    4    6    3    1    1    1    0
```

ITEMMAP DOCUMENTATION

Card Number	Card Name	Format and Description
1	Title Card	(20A4) Description of analysis
2	Parameter Card	(2F5.2,I5,F5.2)

Card Cols	Program Label	Description of Field Contents
1- 5	AMIN	Smallest difficulty in item file.
6-10	AMAX	Largest difficulty in item file.
11-15	NIT	Number of items in item file.
16-20	CNST	Constant added to all difficulties.

SAMPLE ITEMMAP JCL

```
//job card

//  EXEC PGM=ITEMMAP
    where ITEMMAP refers to the name of the load module where
    the ITEMMAP source deck is compiled

//STEPLIB DD DSN=$account.sub.dataset,DISP=SHR
    where dataset refers to the dataset where ITEMMAP
    is stored

//FT01F001 DD DSN=$account.sub.INPUTITEMFILE,DISP=SHR
    where INPUTITEMFILE refers to the name of the
    item file output from SHIFT and sorted by difficulty

//FT06F001 DD SYSOUT=A,DCB=BUFNO=1

//FT05F001 DD *
```

Sample JCL and Control Cards For KCT ITEMMAP Run

```
//job card
//  EXEC PGM=ITEMMAP
//STEPLIB DD DSN=$account.sub.BANKPROG,DISP=SHR
//FT01F001 DD DSN=$account.sub.I6KCT,DISP=SHR
//FT06F001 DD SYSOUT=A,DCB=BUFNO=1
//FT05F001 DD *
MAP OF KCT BANK VARIABLE - FEBRUARY 3, 1980
  2.0 17.00  41  8.0
/*
```

ITEMLIST DOCUMENTATION

Card Number	Card Name	Format and Description
1	Title Card	(20A4) Description of analysis

SAMPLE ITEMLIST JCL

```
//job card
//  EXEC PGM=ITEMLIST
    where ITEMLIST refers to the name of the load module where
    the ITEMLIST source deck is compiled

//STEPLIB DD DSN=$account.sub.dataset,DISP=SHR
    where dataset refers to the dataset where ITEMLIST
    is stored

//FT03F001 DD DSN=$account.sub.INPUTITEMFILE,DISP=SHR
    where INPUTITEMFILE refers to the name of the
    item file output from SHIFT

//FT06F001 DD SYSOUT=A,DCB=BUFNO=1

//FT05F001 DD *
```

Sample JCL and Control Cards For KCT ITEMLIST Run

```
//job card
//  EXEC PGM=ITEMLIST
//STEPLIB DD DSN=$account.sub.BANKPROG,DISP=SHR
//FT03F001 DD DSN=$account.sub.I3KCT50,DISP=SHR
//FT06F001 DD SYSOUT=A,DCB=BUFNO=1
//FT05F001 DD *
```

ITEM FILE KCT VARIABLE - FEBRUARY 3, 1980

FORMLIST DOCUMENTATION

Card Number	Card Name	Format and Description
1	Title Card	(20A4) Description of analysis

SAMPLE FORMLIST JCL

```
//job card
//  EXEC PGM=FORMLIST
    where FORMLIST refers to the name of the load module where
    the FORMLIST source deck is compiled

//STEPLIB DD DSN=$account.sub.dataset,DISP=SHR
    where dataset refers to the dataset where FORMLIST
    is stored

//FT03F001 DD DSN=$account.sub.INPUTFORMFILE,DISP=SHR
    where INPUTFORMFILE refers to the name of the
    form file output from SHIFT

//FT06F001 DD SYSOUT=A,DCB=BUFNO=1

//FT05F001 DD *
```

Sample JCL and Control Cards For KCT FORMLIST Run

```
//job card
//  EXEC PGM=FORMLIST
//STEPLIB DD DSN=$account.sub.BANKPROG,DISP=SHR
//FT03F001 DD DSN=$account.sub.F3KCT50,DISP=SHR
//FT06F001 DD SYSOUT=A,DCB=BUFNO=1
//FT05F001 DD *
```

FORM FILE KCT VARIABLE - FEBRUARY 3, 1980

FILE PREPARATION

Step	File Type	Operation	Input File	Output File
1	Person	Sort by form number	Original data file	P0
2	Form	Sort by: Form number (col 2-4) Card type (col 1) Item position within form (col 5-7)	Original data file	F0
3	Form	For all records, put "9" in column 80 Insert card at end of file with a "4" in column 1	F0	F1
4	Item	Sort by: Item number (col 2-5) Card type (col 1)	Original data file	I0
5	Item	For all records, put "9" in column 80 Insert card at end of file with a "4" in column 1	I0	I1

CREATING THE ITEM BANK

Step	Program	Input File	Output File
6	FORCAL	F1	F2
7	UPDATE	F2 I1	I2
8	SHIFT	F2 I2	F3 I3

BANK OUTPUT PREPARATION

Step	File Type	Operation	Input File	Output File
9	Item	Delete card type "3"	I3	I4
10	Item	Add constant (+5) (make all positives)	I5	I6
11	Item	Sort by item difficulty (col 65-70)	I5	I6

Running the Bank Output Programs

Step	Program	Input File
12	ITEMMAP	I6
13	ITEMLIST	I3
14	FORMLIST	F3

JCL for Sorting Original Form File

```
//job card
//   EXEC PGM=SYNCSORT
//SYSOUT DD SYSOUT=A,DCB=RECFM=FA
//SORTIN DD DSN=$account.sub.ORIGINALFORMFILE,DISP=SHR
//SORTOUT DD DSN=$account.sub.F0KCT,UNIT=SYSDA,
//   DCB=(RECFM=FB,LRECL=80,BLKSIZE=4560),
//   SPACE=(TRK,(5,1)),DISP=(NEW,CATLG)
//SORTWK01 DD UNIT=SYSCR,SPACE=(CYL,4)
//SORTWK02 DD UNIT=SYSCR,SPACE=(CYL,4)
//SORTWK03 DD UNIT=SYSCR,SPACE=(CYL,4)
//SORTWK04 DD UNIT=SYSCR,SPACE=(CYL,4)
//SORTWK05 DD UNIT=SYSCR,SPACE=(CYL,4)
//SYSIN DD *
 SORT FIELDS=(2,3,CH,A,1,1,CH,A,5,3,CH,A),SIZE=NIT
  where NIT is the number of cards in input file SORTIN
 END
```

JCL for Sorting Original Item File

```
//job card
//   EXEC PGM=SYNCSORT
//SYSOUT DD SYSOUT=A,DCB=RECFM=FA
//SORTIN DD DSN=$account.sub.ORIGINALITEMFILE,DISP=SHR
//SORTOUT DD DSN=$account.sub.I0KCT,UNIT=SYSDA,
//   DCB=(RECFM=FB,LRECL=80,BLKSIZE=4560),
//   SPACE=(TRK,(5,1)),DISP=(NEW,CATLG)
//SORTWK01 DD UNIT=SYSCR,SPACE=(CYL,4)
//SORTWK02 DD UNIT=SYSCR,SPACE=(CYL,4)
//SORTWK03 DD UNIT=SYSCR,SPACE=(CYL,4)
//SORTWK04 DD UNIT=SYSCR,SPACE=(CYL,4)
//SORTWK05 DD UNIT=SYSCR,SPACE=(CYL,4)
//SYSIN DD *
 SORT FIELDS=(2,4,CH,A,1,1,CH,A),SIZE=NIT
  where NIT is the number of cards in input file SORTIN
 END
```

BIBLIOGRAPHY

Allen, R. and Fredrick, V. Testing Services Manual. Wisconsin State Department of Public Instruction, Madison. Division for Management and Planning Services, 1979.

Australian Council for Educational Research. Mathematics Item Bank. Books 1 and 2 and Teachers Handbook. Hawthorn, Victoria: ACER, 1978.

Australian Council for Educational Research. Science Item Bank, Books 1, 2 and 3. Science Item Bank Handbook. Hawthorn, Victoria: ACER, 1978.

Australian Council for Educational Research. Social Science Item Bank (and Item Bank Handbook). Hawthorn, Victoria: ACER, 1978.

Bejar, I. I. et. al. Calibration of an Item Pool for the Adaptive Measurement of Achievement. Minnesota University, Minneapolis. Department of Psychology, 1977.

Buckley-Sharp, M. D. and Harris, F. T. C. The banking of multiple-choice questions. The British Journal of Medical Education, 1970, 4, 42-52.

Chelu, C. J. and Elton, L. R. B. An item bank for multiple-choice questions. Physics Education, 1977, 12, 4, 263-267.

Childs, R. Item Banking. Slough, England: National Foundation for Educational Research.

Choppin, B. An item bank using sample-free calibration. Nature, 1968, 219, 870-872.

Choppin, B. Recent developments in item banking. Advances in Psychological and Educational Measurement. New York: Wiley, 1976.

Choppin, B. Item Banking and the Monitoring of Achievement. Research in Progress Series No. 1. Slough, England: National Foundation for Educational Research, 1978.

Choppin, B. The national monitoring of academic standards. Paper presented at the Annual Meeting of the National Council on Measurement in Education. Toronto, 1978.

Choppin, B. Testing the questions - the Rasch model and item banking. In M. St. Raggett, C. Tutt and P. Raggett (Eds.), Assessment and Testing of Reading: Problems and Practices. Ward Lock Educational, 1979.

Connes, B. A dynamic computer system for item banking. Journal of Structural Learning, 1973, 4, 1, 33-40.

Connolly, A. J., Nachtman, W. and Pritchett, E. M. Key Math: Diagnostic Arithmetic Test. Circle Pines, Minn.: American Guidance Service, 1971.

Cornish, G. and Wines, R. ACER Mathematics Profile Series. Operations Test and Teachers Handbook. Hawthorn, Victoria: Australian Council for Educational Research, 1977.

Duckworth, D. and Hoste, R. Question banking: An approach through biology. Schools Council Examination Bulletin 35. London: Evans/Methuen Educational, 1976.

Elliott, C. D., Murray, D. J. and Pearson, L. S. The British Ability Scales. Slough, England: National Foundation for Educational Research, 1977.

Forster, F. Everything You Always Wanted to Know about the Rasch [Model](But Were Afraid to Ask). Portland Public Schools Occasional Papers in Measurement No. 17. Portland Public Schools, Oregon, 1978.

Forster, F., Ascher, G. and Carr, C. Research on the Rasch Measurement Model. Papers presented at the Annual Meeting of the American Educational Research Association. Toronto, 1978.

Forster, F. and Doherty, V. Using the Rasch Approach to Measurement to Solve Practical School Testing Problems. Paper presented at the Annual Meeting of the American Educational Research Association. Toronto, 1978.

Forster, F. and Ingebo, G. Linking Groups of Items. Portland Public Schools Occasional Papers in Measurement No. 19. Northwest Evaluation Association, Oregon; Portland Public Schools, Oregon, 1978.

Forster, F., Ingebo, G. and Wolmut, P. Rasch Model Monograph Series. Portland Public Schools Occasional Papers in Measurement No. 20. Northwest Evaluation Association, Oregon, 1978.

Foyster, J., Allanson, P. and Carter, V. The Australian Item Banking Program: A Planning Study. Hawthorn, Victoria: ACER, 1976.

Goldstein, H. and Blinkhorn, S. Monitoring educational standards - an inappropriate model. Bulletin of the British Psychological Society, 1977, 30, 309-311.

Gartland, M. Illinois School Research and Development, 1980, 16, 3, 99-103.

Gorth, W. P. Comprehensive achievement monitoring. Symposium presented at the Annual Meeting of the American Educational Research Association. Los Angeles, 1969.

Gorth, W. D., Allen, D. W. and Grayson, A. Computer programs for test objective and item banking. Educational and Psychological Measurement, 1971, 31, 1, 245-250.

Gullikson, H. Theory of Mental Tests. New York: Wiley, 1950.

Hiscox, M. D. Item Banks - Where Are They?. Northwest Regional Educational Laboratory, 1979.

Holmes, S. E. Item Banking: An Alternate Approach to Title I Evaluation?. Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, Mass., April 1980).

Holmes, S. E. ESEA Title I Linking Project. Final Report. Oregon State Department of Education, Salem, 1980.

New Zealand. Item Bank: Mathematics Level 2 and Level 6. Wellington, Department of Education, 1973.

Newbould, C. A. and Massey, A. J. A computerised item banking system. British Journal of Educational Technology, 1977, 8, 2, 114-123.

Pollitt, A. B. Item Banking. Issues in Educational Assessment, S.E.D. Occasional Paper, H.M.S.O., 1979.

Popham, W. J. The Instructional Objectives Exchange: Progress and Prospects. Paper presented at the Annual Meeting of the American Educational Research Association. Minneapolis, 1970.

Popham, W. J. Customized criterion-referenced tests. Educational Leadership, 1977, 34, 4, 258-259.

Preece, P. F. W. Objective measurement in education: the Rasch model. Science Eduaction Notes, 770-773.

Purushothaman, M. Secondary Mathematics Item Bank. Slough, England: National Foundation for Educational research, 1975.

Rasch, G. Statistical investigation of current and former students in classes for slow pupils (with M. Ellehammer and C. A. Larsen). Arbejds- og socialministeriet. Kobenhavn, 1952.

Rasch, G. On Applying a General Measuring Theory of Bridgebuilding Between Similar Psychological Tests. Kobenhavn: Danmarks Paedagogiske Institut, 1958.

Rasch, G. Probabilistic Models for Some Intelligence and Attainment Tests. Kobenhavn: Danmarks Paedagogiske Institut, 1960 (Reprinted by University of Chicago Press, 1980).

Rasch, G. An individualistic approach to item analysis. In P. F. Lazarsfeld and N. W. Henry (Eds.), Readings in Mathematical Social Science. Chicago: Science Research Associates, 1966, 89-108.

Rasch, G. On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. Danish Yearbook of Philosophy, 1977, 14, 58-94.

Reckase, M. D. Item Pool Construction for Use with Latent Trait Models. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, 1979.

Rentz, R. R. Monitoring the quality of an item-pool calibrated by the Rasch model. Paper presented at the Annual Meeting of the National Council on Measurement in Education. Toronto, 1978.

Rentz, R. R. and Bashaw, W. L. The national reference scale for reading: an application of the Rasch model. Journal of Educational Measurement, 1977, 14, 161-179.

Rentz, R. R. and Rentz, C. C. Does the Rasch Model Really Work? A Discussion for Practitioners. ERIC/TM Report 67. ERIC Clearinghouse on Tests, Measurement, and Evaluation. Princeton, N. J., 1978.

Robitaille, D. F. et. al. Mathematics Achievement Test Project. Technical Manual. British Columbia Department of Education, Victoria, 1980.

Rutherford, R. J. D. and Trickey, D. S. Objective items - analysis and banking. The Technical Journal, 1973, 11, 4, 14-16.

Scriven, M. The methodology of evaluation. In Tyler, R. W. et. al. Perspectives in Curriculum Evaluation. AERA Monograph 1, Chicago: Rand McNally, 1967.

Shoemaker, D. M. Toward a framework for achievement testing. Review of Educational Research, 1975, 27, 1, 127-147.

Smetherham, D. Banking school knowledge. British Journal of Educational Studies, 1979, 27, 1, 57-68.

Taylor, J. S. Improve your classroom testing skills. Clearing House, 1977, 50, 9, 381-385.

Whitely, S. and Davies, R. The nature of objectivity with the Rasch model. Journal of Educational Measurement, 1974, 2, 2, 163-178.

Willmott, A. S. The place of item banks in local research. Research Intelligence. British Educational Research Association, 1976, 2, 2, 40-42.

Willmott, A. S. The Rasch model - an operational item bank. Paper presented at European Contact Workshop (Council of Europe) on Research into Monitoring National Standards of Educational Achievement in Schools. Windsor, 1976.

Willmott, A. S. and Fowles, D. E. The Objective Interpretation of Test Performance: The Rasch Model Applied. Slough, England: National Foundation for Educational Research, 1974.

Wood, R. The place and value of item banking. Educational Research, 1968, 10, 2, 114-125.

Wood, R. and Skurnik, L. S. Item Banking. Slough, England: National Foundation for Educational Research, 1969.

Wright, B. D. Sample-free test calibration and person measurement in Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1968.

Wright, B. D. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977, 14, 97-116.

Wright, B. D. and Panchepakesan, N. A procedure for sample-free item analysis. Educational Psychological Measurement, 1969, 29, 1, 23-48.

Wright, B. D. and Stone, M. H. Best Test Design. Chicago: MESA Press, 1979.