DOCUMENT RESUME

ED 263 128 TM 850 586

AUTHOR Bejar, Isaac I.
TITLE Test Speededness under Number-Right Scoring: An Analysis of the Test of English as a Foreign Language.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-85-11
PUB DATE Feb 85
NOTE 62p.
PUB TYPE Reports - Research/Technical (143) -- Statistical Data (110)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS College Entrance Examinations; *English (Second Language); Estimation (Mathematics); Evaluation Methods; Higher Education; *Item Analysis; *Language Tests; Latent Trait Theory; *Scoring Formulas; Second Language Learning; Simulation; Test Length; *Timed Tests
IDENTIFIERS Number Right Scoring; *Speededness (Tests); *Test of English as a Foreign Language

ABSTRACT

The Test of English as a Foreign Language (TOEFL) was used in this study, which attempted to develop a new methodology for assessing the speededness of right-scored tests. Traditional procedures of assessing speededness have assumed that the test is scored under formula-scoring instructions; this approach is not always appropriate. In this study, the approach used capitalized on the use of item response theory in equating TOEFL. Two indices were investigated and validated against artificial data and real data. Both indices were able to identity data that contained the speed component. The application of these indices to several test forms suggested that speed did not play a significant role in TOEFL as a whole. However, the second section appeared to be affected by speed in administrations which included pretest items, resulting in a slightly shorter time allowed per item. It was recommended that the item-based index be used on an operational basis for a trial period. It was also concluded that the time limit for Section 2 be increased when new items are being pretested, or that the number of items be reduced accordingly. (GDC)

RESEARCH REPORT

TEST SPEEDEDNESS UNDER NUMBER-RIGHT SCORING: AN ANALYSIS OF THE TEST OF ENGLISH AS A FOREIGN LANGUAGE

Isaac I. Bejar

Educational Testing Service
Princeton, New Jersey
February 1985

Test Speededness Under Number-Right Scoring: An Analysis of the Test of English as a Foreign Language

Isaac I. Bejar

Educational Testing Service

Princeton, New Jersey

February 1985

Abstract

This investigation aimed to develop a new methodology for assessing the speededness of right-scored tests such as the Test of English as a Foreign Language (TOEFL). A new methodology is needed because traditional procedures of assessing speededness assume that the test is scored under formula-scoring instructions. The approach to the problem used in this study capitalizes on the fact that Item Response Theory is used in equating TOEFL. Two indices were investigated and validated against artificial data and real data in a way that was consistent with the ETS criteria for declaring forms speeded. Both indices were able to identify data that contained the speed component. The applications of these indices to several forms suggest that speed does not play a significant role in the TOEFL as a whole. However, Section 2 appeared to be affected by speed in administrations where pretest items are included. This may have been due to the slightly shorter time per item allowed when pretest items are included. It was recommended that one of the indices studied be used on an operational basis for a trial period and that the time limit for Section 2 be increased when pretest items are present or that the number of items be reduced to achieve the same effect.

Test Speededness Under Number-Right Scoring: An Analysis of the Test of English as a Foreign Language

Isaac I. Bejar

A test is speeded when some portion of the test-taking population does not have sufficient time to attempt every item in the test within the allocated time. The effect of speededness depends on, among other things, the scoring procedures used with the test. In particular, in a multiple-choice test where the total score is based on the number of items answered correctly, one effect of speededness could be that students who run out of time answer the remaining items in a more or less random fashion. They can hardly be blamed for doing so since it would be to their disadvantage to leave items unanswered. Unanswered items are scored in the same manner as items answered incorrectly, therefore, the expected score on those unanswered items is simply zero. By contrast, the expected score under random responding is 1/4 (for four-choice items) of the items answered randomly.

Clearly, to the extent that a test is speeded and students engage in random responding, responses to some items will not depend on the students level of knowledge, which of necessity tends to reduce the validity and reliability of total scores. Hence, it is important to closely monitor the speededness of a test. In the case of the Test of English as a Foreign Language (TOEFL) where multiple forms of the test are issued, this requires a fresh analysis for each form. The analysis that is currently performed, however, is appropriate only for tests that are scored with formula scores. Under typical directions for formula-scored tests, a series of unanswered items toward the end of the test can be taken as evidence that those items were not reached. By analyzing the position and number of unreached items it becomes possible to evaluate the extent of a text's speededness.

The speededness analysis based on items not reached is clearly not appropriate for tests such as TOEFL on which scores are based on the number of right answers (ie., without penalty for wrong responses) because, as indicated above it is in the student's interest to answer all items, even if it requires responding randomly. As a result, we would expect very few "unreached" items, and those unreached items that are found would not necessarily be valid indicators of speededness. Unreached items may be more an indication that some students did not comprehend the instructions. (The students' booklets emphasize that the test is scored according to number right and that it is in the student's interest to answer all questions.) The present study looks at speededness from a different perspective, one which does not rely on unreached items.

Purpose of the study. The purpose of this study is to develop new speededness indices that are valid when a test score is based on the total number of correct responses. The approach taken here is

based on Item Response Theory (Lord, 1981) since Item Response Theory (IRT) provides the psychometric infrastructure of the TOEFL (Hicks, 1983). The next section indicates a "theory" of how students are likely to approach a test that is in fact speeded. This theory provides the motivation for the various indices that are studied. We will evaluate the utility of each of these indices for the TOEFL.

Assumptions

The most fundamental assumption we make when speed is not a factor is that students' behavior on the test can be modeled by the three parameter logistic model. That model expresses the student's behavior as the probability of answering an item correctly given the student's ability and the item's psychometric characteristics. The item characteristics incorporated by this model are difficulty (denoted by b) and discrimination (denoted by a). A third parameter, c, takes into consideration that in a multiple-choice item even examinees at the lowest ability levels have a probability greater than zero of answering the item correctly. The model that expresses the probability of answering an item correctly is depicted by Equation 1.

$$P_i(\theta) = c + \frac{1 - c}{1 + \exp[-1.7a_i(\theta - b_i)]} \qquad (1$$

where the subscript i denote the ith item and θ represents ability level.

The validity of Equation 1 depends crucially on the assumption of unidimensionality. According to this assumption the response to a given item only depends on that item's parameters and the students ability, and nothing else. This assumption specifically precludes the possibility that an item's position in the test affects the probability of answering correctly the item. For the most part this model provides a good description of student behavior on the TOEFL. Figure 1 plots the performance of a sample of over 10,000 students on one item. The solid line shows the estimated item characteristic curve (ICC), whereas the squares represent the number of students correctly answering that item in that interval of ability. The size of the square is proportional to the number of students in that interval. As can be seen, for these items the fit of the model is good.

Insert Figure 1 here

When speed is a factor, we assume that within a section of the test students cycle through the section. As they encounter items they are not sure of, they leave them for last. The more able student would probably attempt all items in one or two cycles but the less able student may require more cycles. Since there is a time limit,

the less able student may be left with a few unanswered items when the time is called, and at that point proceeds to answer the remaining items in a fashion that does not depend entirely on ability.

It is reasonable to expect that the more difficult items would be left for last and would thus be subject to random responses. This would have several implications both at the item level and more globally, which provides a rationale for the indices that will be examined. In particular, to the extent that examinees behave as indicated above, we may find that the 3-parameter logistic model is not an accurate model for low ability students on difficult items. Specifically examinees of very low ability, by virtue of their behavior on the more difficult items, may actually have a higher probability of answering the item correctly than examinees of slightly higher ability. Figure 2 plots performance on an item that would be consistent with this form of examinee behavior. Notice that while the solid curve increases monotonically with θ, as required by the model, the data shows a distinct lack of monotonicity. For example, students which according to the model have an ability around -2.0 have a higher chance of answering the item correctly than examinees of somewhat higher ability.

Insert Figure 2 here

We make a similar assumption for section 1. Although section 1 is paced, that is, each student is presented each item for a fixed period of time and there is no opportunity to recycle through the test, for the more difficult items the less able students may not have sufficient time to respond to the item and rather than leaving the item unanswered they simply do not respond as a function of their ability.

To summarize, the speededness theory proposed here suggests that for items above a certain difficulty level examinees of lowest ability actually have a higher probability of answering those items correctly. An empirical implication of the theory is the presence of non-monotonicity of probability of a correct response as a function of ability. Although it is not pretended that the theory describes different strategies students may use in going through the test, it is one possible explanation for the lack of monotonicity that are observed sometimes with difficult items.

Approach

The approach taken in this investigation is to generate data that meets our assumptions when speed is present and when it is not. How this is done will be described shortly. Then we study two types of indices, what may be called an item-level index and a examinee-level

index. Each of these indices is validated, first against artificial data, and then with data from several real TOEFL forms. These analyses are based on Item Response Theory which, as previously indicated, is the psychometric foundation of TOEFL.

Description of the indices

Item-based index. The assumptions outlined earlier regarding speededness suggest that a reasonable item-level index is simply to compute for each item a discrepancy measure between the estimated item characteristic curve, that is Equation 1, and the data. The estimation of Equation 1 is a complex process. Fortunately, a program, LOGIST, exists for performing the estimation. (Wingersky, Barton and Lord, 1982).

A model-free estimate of Equation 1 can be obtained by simply grouping examinees in intervals of ability and computing the proportion of students within each interval that answer the item correctly. This second estimate is not truly model-free since we group the examinees based on their estimated ability and the estimation of ability is based on the model. Nevertheless, the second estimate would not be constrained in the same fashion that the estimate resulting from the LOGIST program is. In particular, estimates obtained from LOGIST are subject to the constraint that the probability of a correct response increases with ability, whereas the "model free" estimate is not constrained in that fashion. Therefore, a comparison of the LOGIST estimate and the model-free estimate can at least detect items where the probability of a correct response does not increase with θ. We are especially interested in this kind of discrepancy since, as outlined earlier, it is the kind of discrepancy expected when speed is present.

We will focus our attention on an index proposed by Yen (1981). The index is generic in the sense that it can be used to detect lack of fit at the item level regardless of the source. The index, called Q, has been modified slightly by Cook and Douglass (1982). It is computed as follows. Examinees are rank ordered in terms of their estimated ability and grouped into 15 equally spaced intervals of θ from -3.00 to 3.00. If any cell contains fewer than 5 examinees, it is collapsed with the neighboring cell closest to $\theta = 0$.

The formula for the index is given by Equation 2,

$$Q = \sum_{j=1}^{15} \frac{[N_j(O_{ij} - E_{ij})]^2}{E_{ij}(1 - E_{ij})} \qquad (2$$

where

N_j is the number of examinees in cell j

O_{ij} is the observed proportion of examinees in cell j that answers the ith item correctly

E_{ij} is the predicted proportion of examinees in cell j that answer the ith item correctly That is,

$$E_{ij} = (1/N_j) \sum_{k \epsilon j}^{N_j} \hat{P}_i(\hat{\theta}_k) \qquad (3$$

where

$\hat{P}_i(\hat{\theta}_k)$ is given by Equation 1.

An examinee-based index. The above index focuses on the item as the unit of analysis. That is, we look for items that may not be performing according to the model. A complementary approach to detecting speededness is to look for examinees that fail to perform as predicted by the model. There is a fair amount of research on this topic (Levine and Drasgow, 1982, for a review), although no applications to the speededness problem have been made. The purpose of this index is to detect students who get a spuriously high or low score. In the present application our focus is on low ability students who get spuriously high scores. By spuriously high I mean that they get a score higher than would be predicted if they were responding as the model expects them to.

Clearly, there is a certain amount of circularity in these indices for if we find such examinees then the model needs to be corrected. For our purposes however, this circularity is reduced because in practice we can not continue to reestimate the model indefinitely. The model here is understood to mean the 3-parameter logistic model since that is the model used by the program to maintain a common scale across forms.

Just as the item-level index involves a comparison between the prediction of the model and the actual data, the examinee-level index also involves a comparison of the prediction of the model and the actual data. Perhaps, the most meaningful index is one that compares the predicted number of responses correct and the actual number of correct responses. However, since the focus of the present investigation is on speededness the index adopted for investigation is the discrepancy between observed and predicted performance on the most difficult items. The choice of this index is in line with the assumption made earlier that students cycle through the test and leave the most difficult items for last. Thus, it is the most difficult items that would be vulnerable to speeded testing conditions.

Computation of index. The computation of the test-level index is as follows:

For each separately timed section of length 1 reorder the items so that they are in increasing order of difficulty, regardless of

item types. (It is often the case that a test contains more than one item type and that within an item type items are ordered in difficulty. For the purposes of this analysis we ignored the item types and simply ordered the item in increasing order of difficulty.)

Next, we divide the section into an "easy" and a "hard" part. The division point is obviously arbitrary but we will adopt a division that is compatible with the standard used at ETS for declaring a form as speeded. According to that standard, a section is speeded if fewer than 80% of the examinees reach the last item on the section and fewer than 100% complete 75% of the section. Thus, we will divide the section into two with the easy part containing the 75% easiest items and the hard part containing the 25% hardest items.

Now we compute the expected and observed performance on the hard part. Again, the rationale for this is that it is the performance on the most difficult items that is vulnerable to speed. The expected performance is computed by

$$E(\theta_e) = \sum_i P_i(\theta_e) \qquad (4$$

where P_i is given by Equation 1, the sum is taken over the hard items and θ_e is the estimated ability based on the easy items. In other words, we compute the expected performance on the hard items based on the ability estimate computed on the easy items.

The observed performance on the hard items is simply

$$O = \sum_i u_i \qquad (5$$

where u_i is 1 if the ith item is answered correctly and 0 otherwise, and the sum is taken over the hard items only.

Evaluation of indices with artificial data

Since there is no prior information on the indices under study a significant portion of this investigation was devoted to the evaluation of the IRT indices with artificial data. This approach allows us to examine the behavior of the indices when speed is a factor and when it is not. Put briefly, the evaluation consists of generating data that fits the 3-parameter logistic model but having statistical characteristics similar to those found in a typical TOEFL form. The performance of the indices with these data will let us know what to expect when in

fact speed is not a factor. We then introduce a speed effect into these data. Clearly, in order to do this we must make several assumptions that will be outlined below.

Generation of data. The procedure for generating data according to the 3-parameter model requires that we specify the item parameters and the distribution of ability. Since we are concerned with simulating TOEFL data we will choose item parameters from a typical TOEFL form (November 1982). For concreteness, we will assume that the distribution of ability is normal. The a, b, and c parameters for simulating data for sections 2 and 3 are found in Appendix A. (Although the validation with real data was based on all three sections, the validation with simulated data was based only on sections 2 and 3.) These are simply the "operational" parameter estimates for the November 1982 administration. By using these parameters, we insure that the simulated data have the characteristics of data from typical TOEFL form. With the item parameters and distribution of ability in hand, the procedure to generate the data is as follows:

Sample an ability value from a normal distribution with mean 0 and standard deviation of 1.0

For each item Compute the probability of a correct response by evaluating Equation 1 substituting the sampled ability and the parameters for this item. Call the obtained probability p.

Then: Generate a random number from a rectangular distribution in the interval 0 to 1.0. Call this number p*.

If $p > p^*$, then the response to this item is correct, otherwise it is incorrect.

We record the outcome of the simulation in a data matrix U, such that u_{ji} is the jth subject response to the ith item.

Repeat for each item.

Repeat for each sampled ability value.

The outcome of this procedure is a data matrix with as many rows as there are ability values and as many columns as there are items. We have chosen 3,000 as the sample size. For the simulation of section 2 we have 38 items, and 58 for the simulation of section 3. We will call this matrix P, for power, since it contains data that fits Equation 1 and it is uncontaminated by speed. We now need to modify the P matrix in such a way that the effect of speed is simulated in a form that is consistent with the ETS criteria for speededness.

The procedure to introduce speed in the artificial data consists of changing the responses to the 25% most difficult items of examinees in the lower 20% of the ability distribution. This does not parallel the conventional criterion since neither difficulty nor ability come into the definition but it does follow the spirit of the traditional criterion.

Insert Figure 3 here

Figure 3 shows the structure of an S matrix where the ability levels have been ordered from high to low and the items have been ordered in difficulty. We assume that responses in all but the "southeast" region of this matrix are modeled by the 3-parameter logistic model. Responses in that region are subject to speededness because they result from the interaction of students with the lowest ability attempting the most difficult items. To match the spirit of ETS criteria for speededness, that region is defined by the lowest 20% of the ability values and the 25% most difficult items. According to our "theory" of speededness, responses in the "southeast" follow one of two models: the usual 3-parameter logistic model and a random response one. As ability gets lower and the items get more difficult the probability that the usual model applies diminishes while the probability of a random responses increases. Within an interval of ability, the probability of a random response increases in a gradual fashion. Figure 3 illustrates the scheme for a 40-item test. In a 40-item test there would be 10 items subject to speed. Then the interval from the 20th percentile to -3.0 is divided into 10 subintervals, or in general as many subintervals as there are items subject to speed. The width of the subintervals is given by

$$w = (\theta^* + 3)/n_s \qquad (6$$

where n_s is the number of items subject to speed and θ^* is the ability value corresponding to the 20th percentile. Let us call the n_s intervals i_k where k goes from 1 to n_s and $i_k = k$. For example, if $\theta^* = -2.0$, then $w = .10$. Similarly, let us refer to the 25% most difficult items by j where j_1 is the least difficult among those items and $j_{ns} = 10$ refers to most difficult item. Then the probability of a random response is given by

$$r_{ij} = \begin{cases} (i_k j_l)/n_s \quad k,l = 1 \ldots n_s & \text{if } r_{ij} \leq 1.0 \\ 1.0 & \text{if } r_{ij} > 1.0 \end{cases} \qquad (7$$

With these definitions in hand we can now describe the S data matrix as follows

If the ability level is above the 20th percentile the probability of a correct response is given by the 3-parameter logistic model.

For ability levels below the 20th percentile correct responses to the easiest 75% of the items are also given by the 3-parameter logistic model.

For ability levels below the 20th percentile correct, responses to the 25% most difficult items are given by

.25 with probability r_{ij}

$P(\theta)$ with probability $1 - r_{ij}$

Insert Figure 3 here

Results for artificial data

Examinee-index. In this section we present results relevant to the evaluation of the proposed indices with artificial data. Table 1 shows the result based on one sample of approximately 2000 subjects with normally distributed ability for both section 2 and section 3. The first step in the computation of this table is to classify response vectors in intervals based on estimated theta for all the items. Next, we compute the mean number right and predicted number right score in each interval. Finally, the difference between the number right and predicted number right at each interval is an index of the effect of speed.

The columns labeled "P" are for the data under power conditions, while the column labeled "S" is for the speeded condition. The difference was computed by substracting PNR, the expected number-right from the observed number-right, NR. Thus, a negative difference indicates that the expected number-right exceeded the observed score.

We note that as θ increases the differences become positive and increase in magnitude. In other words, regardless of the presence of a speed factor the estimate of ability based on the easiest items appears to be an underestimate for high levels of ability. This was not expected. However, since we are interested in the other extreme of the scale, it is not pertinent to our results.

Insert Table 1 here

The effect of the presence of speed can best be examined by plotting the differences between observed and expected performance.

Figure 4 shows the index for section 2 data. The curve labeled "S2" is for the data for section 2 with a speed factor. The curve labeled "P2" is for the same data without a speed factor. As can be seen "S2" bulges towards the low values of ability. This is an indication that when a speed factor is present low ability "subjects" get spuriously high scores.

Insert Figure 4 here

Figure 5 shows corresponding results for section 3. The curve labeled "P3" is for data simulating section 3; the curve "S3" is for the same data with a speed factor added. We see again that there is a bulging toward the lower ability values.

Insert Figure 5 here

The above results are not entirely applicable to real TOEFL data. The reason for this lies in the way the TOEFL is equated. The procedure in current use is called "fixed b's" equating (Hicks, 1983). As its name implies, that method is based on fixing the difficulty estimate of some of the items to the value obtained using a different sample. However, only the difficulty or b parameter is held fixed; the other parameters, c and a are reestimated. As a result, there is an opportunity for improving the fit of the items at equating time. That opportunity is missing from the results just presented.

In order to approximate the effect of the opportunity to adjust the parameter estimates, the a and c parameters of all the items were reestimated and the index recomputed. The results are shown in Table 2.

Insert Table 2 here

The differences between the expected number right and the observed number right for sections 2 and 3 are plotted in Figures 6 and 7 respectively. As can be seen, the effect of speed is still present after reestimating the parameters.

Insert Figures 6 and 7 here

Item-based index. These results suggest that the index proposed here is in fact capable of detecting situations when a speed factor,

thus meeting the spirit of ETS criteria. We now would like to examine the performance of the item-based index discussed earlier, Q.

We first compute the index for each item for the P and S datasets for sections 2 and 3. Our interest is to see if the value of the index is higher for the most difficult items when they are subject to speed. Rather than present the raw results, a curve was computed containing the running average of Q beginning with the most difficult item. Thus, the right-most point in this curve is just the value of Q for the most difficult item, the second point is the average of Q for the two most difficult items, the last point in the curve is just the average of Q across all the items. These mean values are then plotted as a function of the difficulty of the most difficult item for a given point. For example, the first point, which is the value of Q for the most difficult item, is plotted against the difficulty of that item; the second value is plotted against the difficulty of the second most difficult item.

The results for sections 2 and 3 appear in Figures 8 and 9 respectively. For section 2 under power conditions the running average remains fairly constant although there is a slight tendency for it to increase as a function of difficulty. However, the contrast with the curve for data under the speed condition is obvious. The mean is almost three times as high for the most difficult items, which of course are the items subject to speed, and then decreases sharply. The pattern is much the same for section 3: under power conditions the running average is fairly constant but it increases as a function of difficulty when speed is present.

Insert Figures 8 and 9 here

Conclusions based on the artificial data. Based on the results presented for the examinee-based and item-based index, it can be concluded that both indices are able to identify the presence of speed. We can be reasonably confident that the contrast between the indices under the power and speed condition are due to speed because of the nature of the artificial data. Thus, other things being equal, it is appropriate to conclude that if both indices behave as they did here we may be justified in concluding that a form is speeded. In practice, however, there may be other deviations from the model present in the data due to factors other that speed, which naturally will make the interpretation of these indices less certain. Nevertheless, the above results gives us an indication of what would happen under idealized circumstances and should be valuable in the interpretation of these indices when applied to real data.

Results with real data

We now turn to the results based on actual TOEFL data. We will present results with the two indices for the November 1982

administration. Additional results using the item level index will be presented for four recent administrations.

Results for the examinee-based index. First we present the results of the examinee-based index. Table 3 contains the results of applying the procedure to a spaced sample of approximately 2,000 examinees from the November 1982 administration. We notice a similar increase in the difference between NR and PNR for high values of ability. But, as before, our interest lies in the other direction. To better see the results, Figure 10 shows the index for sections 1, 2, and 3.

Insert Table 3 here

Insert Figure 10 here

Figure 10 suggests that sections 2 and 3 are not affected by guessing but section 3 appears to show a similar pattern to the simulated data with a speeded factor. This sugggests that a speed factor may be present in section 3. To add credence to this suggestion, a different spaced sample from the same administration was examined. (Only the results for section 2 and 3 are presented because section 1 did not converge after spending a considerable amount of time). The results appear in Table 6 and a corresponding plot in Figure 11.

Insert Table 4 here

As can be seen a similar pattern emerges: there is no "bulging" of the curve at low ability levels for section 2 but there is for section 3. As with the first sample the frequency of the lowest three ability intervals is 0 for section 2, which precludes the possibility of bulging in that range of ability.

Insert Figure 11 here

Results for the item-index with real data. Tables 5, 6, and 7 display the item parameter estimates, for sections 1, 2, 3 respectively, for the November 1982 administration. The tables also contain the value of the Q index associated with each item. Figure 12 shows the average Q value as a function of difficulty, computed in the same manner as with the artificial data.

Insert Tables 5, 6, and 7 here

Insert Figure 12 here

For section 1 the index rises slowly from a value of about 30 to about 45; for section 2 the index also raises slowly but from 20 to about 40; finally for section 3 the index remains fairly constant at 20 except that it decreases somewhat for high values of difficulty. Several observations are in order. The best indication we have of what the value of Q should be when the data fits the underlying model is from the P data sets. For those datasets the value is approximately 15 throughout the range of difficulty. It is interesting to note that for section 1 with real data the value is approximately 30 and rises somewhat as difficulty increases. However, the increase is nowhere in magnitude of what it was for the S data sets. Thus, although the fit is not as good as it is for the artificial data it seems that the lack of fit is not necessarily due to the presence of speed. For example, item 14 of Section 3 had an extremely high value of Q. Examination of this item indicates that the reason is a shift in the difficulty of the item. As stated earlier, in equating, only the a and c parameters are reestimated. The b parameter, difficulty, is held fixed from pretest administration to final administration. Such shifts in difficulty are reflected in high Q values.

For sections 2 and 3, the fit at low difficulty values is somewhat higher than it is for artificial data that fits the model. For section 3 the fit does not seem to depend on difficulty but for section 2 there is a tendency for lack of fit to increase with difficulty although not at the same rate as it does with artificial data with a speed component.

Further analysis. The results just presented suggest that the November 1982 form is not speeded, using as criteria the expectation provided by artificial data subject to speededness to an extent consistent with ETS conventions. Although Section 3 appeared speeded on the examinee-level index, it did not appear speeded on the item-level index. Similarly, Section 2 appeared speeded on the item-based index, although it did not reach the level found with the simulated data, and it did not appear at all on the examinee-based index. It became apparent, however, that the computation of the examinee-based index is considerably more difficult and expensive. Indeed, in one instance the process ran out of computer time. Therefore, it appears that for operational purposes the item-based index is to be preferred.

In this section we present additional results for the item-based index on additional administrations. Specifically, the item-based

index was applied to all sections of the September, October, November, and December 1983 administrations. Tables showing the item parameters and Q index can be found in Appendix A. The results are summarized in Figures 13 through 16.

An examination of the figures suggests the following: for section 1 the item-based index remains flat as difficulty increases. However, as with the results presented earlier, the general elevation of the plot is higher for Section 1 than for Sections 2 and 3. For Section 3, again as before there is no evidence that speed plays a significant role and also as before the general elevation of the plot is the lowest. For Section 2, however, there is a tendency for the plot to increase with difficulty. For the September administration there is a marked increase of the plot beyond a difficulty value of 1.0. A similar increase is noticed for the November administration but not for the October administration.

For the December administration the index for Section 2 appears similar to the plot that we obtained with artificial data in that the plot begins to rise at around -.5 and actually reaches the level comparable to the artificial data with a speed factor. Thus, it would seem that Section 2 for the December administration is speeded. Because of the interrelationship between ability and item difficulty in the present approach to assessing speededness, the apparent speededness of the December 1983 administration may have arisen because in that administration the examinees were of lower ability or the items were unusually difficult. Table 8 reports the mean and standard deviations for the September 1983 through December 1983 administration on the three TOEFL sections' scores as well as the total score. The sample sizes on which these statistics are computed are shown as well. As can be seen, it does not appear that the December examinees are of lower ability.

Insert Table 8 here

Next we looked at the mean difficulty of the September to December 1983 administrations to see if in the December form Section 2 was unusually difficult. The results are shown in Table 9. It does not appear that Section 2 was harder in the December form than in September, October, or November. However, the variability of the difficulty and the discrimination present is highest for December.

Insert Table 9 here

Summary and Conclusions

This investigation aimed to develop new methodology for assessing the speededness of TOEFL test forms. This methodology is needed

because the traditional procedures of assessing speededness assume the test is scored under formula-scoring instructions whereas TOEFL is scored under number-right scoring instructions. The approach to the problem capitalizes on the fact that for TOEFL Item Response Theory is the equating methodology used.

Two complimentary indices of speededness were investigated. Both of these indices are based on the assumption that if a test is speeded, performance on the most difficult items will not be entirely a function of ability. One index contrasts the observed performance on the most difficult items in the test against the performance by the model for those items. The second index is based on an index of fit of each item.

Both indices were validated by creating data that fit the model and then introducing a speed component into the data. This was done in a way that was consistent with the ETS criteria for declaring forms speeded. Both indices were able to identify data that contained the speed component. The indices were then applied to real data from the November 1982 TOEFL administration. The results suggest that none of the sections is speeded. The application of the examinee-based index suggested that Section 3 might be speeded, but the item-based index did not indicate this. For Sections 1 and 2, the examinee-based index did not suggest either section was speeded. The item-based index increased with item difficulty, which would be consistent with the presence of speed, but the increases were not as large as would be expected if speed were present to the extent necessary to meet the conventionally-inspired criteria for declaring a test speeded.

It became apparent during the investigation that the examinee-level index was expensive to compute. Therefore, the item-level index was applied to data from four recent administrations. There was no indication that Sections 1 or 3 were speeded, but Section 2 of the December form tended to appear speeded. The reason for this apparent effect was not that the December form was harder or the population less proficient. Rather the effect seems to be due to the slightly different time allocation for each section. The time allowed for Section 2 is 25 minutes for 40 items or .62 minute per item. If there are pretest items, however, the allowance of time per item is 35/60 = .58 minutes per item. To maintain an identical time allowance, 37.5 minutes should be allowed for the entire section instead of the current 35 minutes. Therefore, when there are pretest items in Section 2, 2 1/2 more minutes should be allowed to maintain identical timing constraints. For Section 3 without pretest items, the time allowance is 35 for 60 items of .58 minute per item. If there are pretest items, the time allowance is 65 minutes for 90 items or .72 minute per item.

The November and December administrations were pretest; and it was these two administrations, especially December's administration, that appeared more speeded according to the item-level index. Based on

these results it is recommended that the item-based index proposed here be implemented by the program, at least on a trial basis, to monitor the speededness, or lack of it, of future TOEFL forms. If further application of the index continues to suggest that Section 2 may be somewhat speeded, it may be advisable to increase the time allocation for Section 2 when pretest items are available by 2 1/2 minutes.

References

Cook, L. L. & Douglass, J. B. Analysis of fit and vertical equating with the three-parameter model. Paper presented at the Annual Meeting of The American Educational Research Association, New York, 1982.

Hicks, M. M. True score equating by fixed b's scaling: A flexible and stable equating alternative. Applied Psychological Measurement, 1983, 7, 255-266.

Hicks, M. M. and Sperling, J. V. Test analysis report. Test of English as a foreign language, Form 3ETF5, May 18 administration. (Unpublished statistical report) Princeton, NJ: Educational Testing Service, 1982.

Levine, M. V. & Drasgow, F. Appropriateness measurement: Review, critique and validating studies. British Journal of Mathematical and Statistical Psychology, 1982, 35, 42-56.

Lord, F. M. Application of Item Response Theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

Wingersky, M. S., Barton, M. A. and Lord, F. M. LOGIST users guide. Princeton, NJ: Educational Testing Service, 1982.

Figure	Description
1	Empirical and theoretical item characteristic function based on a sample of over 10,000 examinees for an item that fits the model well.
2	Empirical and theoretical item characteristic function based on a sample of over 10,000 examinees for an item that exhibits lack of fit consistent with the presence of speed.
3	Data matrix for data with a speed factor.
4	Plot of the difference between NR and PNR for Section 2 data under power and speed conditions before adjusting item parameter estimates.
5	Plot of the difference between NR and PNR for Section 3 under power and speed conditions before adjusting item parameter estimates.
6	Plot of the difference between NR and PNR for Section 2 data under power and speed conditions after adjusting item parameter estimates.
7	Plot of the difference between NR and PNR for Section 3 under power and speed conditions after adjusting item parameter estimates.
8	Average Q as a function of difficulty under speed and power condition for Section 2 data.
9	Average Q as a function of difficulty under speed and power condition for Section 3 data.
10	Plot of the difference between NR and PNR for Sections 1, 2, and 3 based on data from approximately 2,000 examinees from the November 1982 administration.
11	Plot of the difference between NR and PNR for Sections 1, 2, and 3 based on a replication sample of approximately 2,000 examinees from the November 1982 administration sample from Sections 2 and 3.

12 Plot of item-based index for Sections 1, 2, and 3. The top panel is for Section 1, the middle panel for Section 2, and the bottom panel for Section 3.

13 Plot of item-based index for Section 1 (top panel), 2 (middle panel), and 3 (bottom panel) for the September 1983 administration based on approximately 2,000 examinees.

14 Plot of item-based index for Section 1 (top panel), 2 (middle panel), and 3 (bottom panel) for the October 1983 administration based on approximately 2,000 examinees.

15 Plot of item-based index for Section 1 (top panel), 2 (middle panel), and 3 (bottom panel) for the November 1983 administration based on approximately 2,000 examinees.

16 Plot of item-based index for Section 1 (top panel), 2 (middle panel), and 3 (bottom panel) for the December 1983 administration based on approximately 2,000 examinees.

Table 1

NR, PNR and the difference between the two for data sets with and without speed for sections 2 and 3 as a function of estimated ability without adjusting item parameters.

*	P 2			P 3			S 2			S 3			P2	S2	P3	S3
TH	nr	pnr	diff	nr	pnr	diff	nr	pnr	diff	nr	pnr	diff	N	N	N	N
-2.6	*	*	*	2.50	2.97	-.47	*	*	*	3.00	2.96	.04	0	0	2	1
-2.2	1.50	1.55	-.05	2.67	3.01	-.34	2.50	1.55	.95	4.22	3.00	1.22	2	2	9	9
-1.8	1.68	1.62	.06	2.79	3.07	-.28	3.00	1.97	1.03	3.28	3.08	.20	25	21	19	18
-1.4	1.24	1.66	-.42	3.50	3.21	.29	2.40	1.65	.75	3.59	3.21	.38	55	47	48	41
-1.0	1.75	1.79	-.04	3.30	3.43	-.13	2.07	1.78	.29	3.69	3.42	.27	115	114	94	103
-.6	2.10	2.04	.06	3.86	3.83	.03	2.28	2.00	.28	3.93	3.83	.10	203	195	229	235
-.2	2.33	2.54	-.22	4.64	4.52	.12	2.45	2.54	-.08	4.64	4.55	.09	333	332	322	316
.2	3.17	3.44	-.27	5.76	5.72	.04	3.16	3.27	-.11	5.77	5.72	.05	416	414	413	413
.6	4.49	4.67	-.18	7.25	7.38	-.13	4.49	4.63	-.14	7.25	7.38	-.13	340	339	385	385
1.0	5.88	6.27	-.39	9.03	9.28	-.25	5.87	6.27	-.40	9.03	9.28	-.25	236	235	248	248
1.4	7.55	7.07	.48	10.70	11.01	-.31	7.55	7.07	.48	10.70	11.01	-.31	111	111	137	137
1.8	8.68	7.68	1.00	11.97	12.40	-.44	8.68	7.68	1.00	11.97	12.40	-.44	50	50	29	29
2.2	9.58	7.84	1.74	13.17	12.62	.55	9.55	7.92	1.62	13.17	12.62	.55	12	11	23	23
2.6	10.00	8.23	1.77	13.87	13.61	.26	10.00	8.23	1.77	13.87	13.61	.26	5	5	15	15
3.0	10.00	8.87	1.13	14.60	13.36	1.24	10.00	8.87	1.13	14.60	13.36	1.24	2	2	5	5

Table 2

NR, PNR and the differnce between the two for data sets with and without speed for setions 2 and 3 as a function of estimated ability after adjusting item parameters.

*	P 2			P 3			S 2			S 3			P2	S2	P3	S3
TH	nr	pnr	diff	nr	pnr	diff	nr	pnr	diff	nr	pnr	diff	N	N	N	N
-2.6	*	*	*	3.25	2.98	.27	*	*	*	3.50	3.05	.45	0	0	4	2
-2.2	1.67	2.19	-.52	2.60	3.02	-.42	2.25	1.65	.60	4.00	3.09	.91	6	4	5	8
-1.8	1.70	1.48	.22	2.86	3.08	-.22	3.18	1.69	1.49	3.13	3.14	-.01	20	17	21	23
-1.4	1.33	1.74	-.41	3.36	3.23	.12	2.43	1.77	.67	3.48	3.28	.20	48	44	45	50
-1.0	1.72	1.77	-.05	3.41	3.45	-.04	2.04	1.90	.15	3.66	3.49	.18	109	117	95	109
-.6	2.11	1.90	.21	3.90	3.86	.05	2.39	2.10	.28	3.91	3.89	.02	198	199	223	235
-.2	2.31	2.40	-.09	4.63	4.52	.11	2.46	2.60	-.14	4.61	4.57	.04	327	341	317	327
.2	3.16	3.19	-.03	5.78	5.70	.08	3.17	3.38	-.21	5.78	5.74	.04	414	415	419	424
.6	4.47	4.52	-.05	7.25	7.37	-.12	4.54	4.71	-.17	7.25	7.41	-.17	345	348	374	378
1.0	5.86	6.15	-.29	8.98	9.17	-.19	5.96	6.31	-.36	9.01	9.27	-.26	225	224	242	246
1.4	7.48	7.00	.48	10.59	10.91	-.31	7.55	7.13	.42	10.66	11.05	-.39	121	106	138	136
1.8	8.65	7.73	.93	11.95	12.20	-.25	8.53	7.67	.86	12.03	12.46	-.43	46	49	38	35
2.2	9.38	7.68	1.71	13.21	12.77	.44	9.50	7.74	1.76	13.35	12.85	.50	13	10	24	23
2.6	10.00	8.27	1.73	13.92	13.90	.01	10.00	8.46	1.54	13.92	13.99	-.07	4	6	12	12
3.0	10.00	8.84	1.16	14.40	13.38	1.02	*	*	*	14.50	13.74	.76	2	0	5	5

Table 3

NR, PNR and difference for real data set from the November 1982 administration

TH	Section 1 nr	Section 1 pnr	Section 1 diff	Section 2 nr	Section 2 pnr	Section 2 diff	Section 3 nr	Section 3 pnr	Section 3 diff	Section 3 N	Section 3 N	Section 3 N
-2.6	*	*	*	*	*	*	*	*	*	0	0	0
-2.2	3.22	2.89	.32	*	*	*	4.14	3.01	1.13	23	0	7
-1.8	3.17	2.97	.20	*	*	*	3.27	3.08	.18	36	0	15
-1.4	3.20	3.15	.05	1.83	1.65	.19	3.20	3.20	.00	83	65	45
-1.0	3.26	3.46	-.20	1.71	1.79	-.08	3.70	3.40	.29	203	118	102
-.6	3.92	4.03	-.11	2.05	2.02	.03	3.89	3.82	.08	279	198	194
-.2	4.99	4.90	.09	2.36	2.53	-.17	4.64	4.54	.10	391	332	360
.2	6.45	6.33	.12	3.27	3.27	-.01	5.82	5.66	.16	382	446	476
.6	8.26	8.02	.24	4.65	4.45	.21	7.51	7.28	.24	292	342	372
1.0	9.83	9.53	.30	6.17	5.62	.56	9.16	9.03	.13	171	198	230
1.4	11.33	10.44	.89	7.74	6.32	1.42	10.92	10.54	.38	81	111	97
1.8	12.68	10.92	1.76	9.14	6.74	2.40	12.71	11.07	1.65	31	35	42
2.2	*	*	*	10.00	6.86	3.14	13.52	11.81	1.71	0	9	21
2.6	*	*	*	*	*	*	14.75	12.12	2.63	0	0	12
3.0	*	*	*	*	*	*	*	*	*	0	0	0

Table 4

NR, PNR and difference for a second spaced sample from the November 1982 administration.

	Section 2			Section 3			Section 2	Section 3
TH	nr	pnr	diff	nr	pnr	diff	N	N
-2.6	*	*	*	*	*	*	0	0
-2.2	*	*	*	3.89	3.01	.88	0	9
-1.8	*	*	*	3.43	3.08	.35	0	14
-1.4	1.86	1.64	.22	3.46	3.2	.26	59	41
-1.0	1.64	1.80	-.16	3.65	3.41	.24	124	114
-.6	2.06	2.02	.04	3.73	3.82	-.09	216	236
-.2	2.45	2.54	-.09	4.62	4.54	.08	351	315
.2	3.27	3.31	-.04	5.63	5.70	-.07	412	449
.6	4.72	4.44	.28	7.27	7.33	-.06	358	389
1.0	5.95	5.79	.16	9.18	9.06	.12	232	252
1.4	7.79	6.37	1.42	10.79	10.69	.09	94	103
1.8	9.13	6.75	2.38	12.79	10.99	1.81	31	39
2.2	9.38	7.16	2.22	13.81	11.67	2.14	13	26
2.6	*	*	*	14.38	12.25	2.14	0	13
3.0	*	*	*	*	*	*	0	0

Table 5

Item parameter estimates, and Q indices for section 1 of the November 1892 administration

Pos	eta	a	b	c	Q
1	.18	.53	-2.68	.16	9.48
2	.33	1.23	-1.42	.16	11.55
3	.28	1.04	-1.94	.16	13.15
4	.25	.99	-1.27	.16	29.02
5	.29	1.16	-1.29	.16	11.88
6	.13	.83	-.82	.16	30.32
7	.26	.79	-1.06	.16	14.37
8	.26	.77	-.61	.16	65.32
9	.28	1.48	-.77	.05	10.05
10	.22	.94	-.76	.06	12.26
11	.11	.77	.46	.16	34.24
12	.23	1.20	-.65	.11	31.57
13	.18	1.27	-.22	.10	19.88
14	.15	1.50	.60	.32	28.29
15	.18	.46	-.34	.16	273.56
16	.18	1.50	-.07	.32	17.06
17	.19	.97	.09	.04	24.93
18	.09	1.50	-.13	.06	52.64
19	.12	1.50	.42	.20	14.34
20	.18	1.50	.54	.11	65.93
21	.28	1.20	-1.70	.16	30.96
22	.28	.94	-1.18	.16	12.97
23	.24	.67	-1.12	.16	23.54
24	.22	1.32	-.31	.29	10.47
25	.28	.57	-1.45	.16	48.23
26	.15	.87	-.77	.16	54.98
27	.22	.88	-1.53	.16	28.44
28	.25	.87	-.23	.28	11.60
29	.17	1.29	-.23	.15	10.57
30	.14	.43	-.18	.16	46.17
31	.12	.61	-.28	.16	25.94
32	.18	1.00	-.08	.27	11.01
33	.28	1.25	-.14	.18	8.88
34	.18	1.41	.15	.16	17.38
35	.11	1.50	1.20	.19	44.75
36	.30	.86	-1.49	.16	25.75
37	.22	.83	.20	.13	8.20
38	.27	.70	-2.27	.16	6.75
39	.16	1.03	.00	.20	13.50
40	.24	.74	-.78	.16	7.04
41	.28	1.18	-.40	.12	15.22
42	.19	.97	-.62	.23	11.23
43	.26	1.03	-.87	.16	47.86
44	.24	1.50	-.44	.30	21.47
45	.22	1.06	-.88	.31	11.23
46	.14	.95	-.53	.24	4.66
47	.14	.92	.26	.26	18.96
48	.21	1.18	.16	.26	14.31
49	.13	1.50	-.01	.38	41.01
50	.20	.88	.33	.38	24.98

TABLE 6

ITEM PARAMETER ESTIMATES, ETA AND Q INDICES FOR SECTION 2 OF THE NOVEMBER 1892 ADMINISTRATION

POS	ETA	A	B	C	Q
1	.24	.60	-2.41	.19	18.86
2	.33	1.44	-1.20	.22	94.82
3	.33	1.46	-.87	.37	8.27
4	.33	1.50	-.80	.22	8.48
5	.26	1.07	-.61	.19	9.53
6	.26	1.50	-.48	.23	14.81
7	.16	.69	-.63	.19	33.62
8	.30	1.47	-.12	.25	13.11
9	.28	1.50	-.15	.19	7.42
10	.20	1.08	.01	.13	11.29
11	.09	1.29	.31	.10	7.55
12	.19	1.34	.48	.13	13.88
13	.14	1.35	.47	.08	39.07
14	.19	1.13	1.69	.05	34.44
16	.35	1.04	-.78	.32	12.51
17	.38	1.07	-1.36	.19	12.90
18	.30	1.50	-.80	.27	11.85
19	.23	1.14	-.64	.26	13.32
20	.23	1.39	-.71	.12	21.99
21	.26	1.50	-.59	.19	10.40
22	.24	1.03	-.90	.09	8.92
23	.31	.86	-.55	.14	7.13
24	.19	.90	-.62	.04	10.35
25	.26	.69	-.64	.19	22.76
26	.20	.92	-.23	.22	46.15
27	.29	.92	-.45	.32	9.71
28	.12	1.50	1.02	.13	47.40
29	.26	1.06	.02	.29	13.08
30	.23	.48	-.61	.19	20.00
31	.16	.91	.13	.29	18.91
32	.22	1.14	-.14	.15	15.88
33	.32	.73	.05	.18	10.22
34	.14	1.50	.13	.28	14.04
35	.28	.79	.56	.32	12.77
36	.20	1.17	.64	.25	22.20
37	.13	.92	.43	.05	37.00
38	.15	.80	.89	.20	42.17
39	.03	1.50	.75	.18	27.36

Table 7

Item parameter estimates, eta and Q indices for section 3 of the November 1892 administration

Pos	eta	a	b	c	Q
1	.29	1.34	-1.37	.24	6.79
2	.21	1.50	-.77	.30	10.69
3	.24	1.49	-.52	.23	8.78
4	.17	1.50	-.25	.45	16.81
5	.20	.87	-.10	.44	21.29
6	.20	.87	-.63	.12	11.94
7	.27	.72	-.97	.10	33.94
8	.17	.68	-.84	.10	30.03
9	.28	.92	-.36	.26	12.05
10	.23	1.23	-.33	.23	5.55
11	.15	1.25	-.02	.36	13.76
12	.13	.93	.06	.24	8.41
13	.22	1.18	-.13	.23	18.16
14	.16	.45	.35	.10	42.42
15	.19	1.13	-.20	.18	14.10
16	.19	.96	-.31	.11	18.89
17	.14	1.05	.33	.28	9.08
18	.21	1.06	.22	.26	13.68
19	.15	.88	.15	.13	19.11
20	.14	1.01	-.12	.05	19.80
21	.23	.35	.05	.10	59.92
22	.17	1.50	.41	.25	13.75
23	.20	.76	.08	.16	13.21
24	.14	1.17	.60	.21	36.49
25	.14	.99	.62	.23	14.55
26	.09	1.10	.32	.13	11.07
27	.03	.98	1.73	.17	13.19
28	.12	1.24	1.20	.29	13.64
29	.20	1.45	.49	.17	49.12
30	.19	1.15	-.85	.31	9.62
31	.22	1.04	-.39	.17	23.80
32	.38	1.50	-.88	.06	23.30
33	.18	.48	.21	.10	61.59
34	.31	.95	-1.61	.10	28.49
35	.27	1.50	-.61	.10	17.13
36	.12	1.07	-.31	.17	21.29
37	.30	.73	-.68	.10	22.34
38	.28	.81	-.56	.30	44.96
39	.26	1.50	-1.02	.21	9.86
40	.21	1.44	.00	.30	33.90
41	.25	.81	-.92	.10	8.14
42	.19	.62	-2.51	.10	13.98
43	.09	1.01	.40	.11	19.80
44	.07	1.19	1.59	.05	11.71
45	.19	1.40	.08	.32	14.79
46	.12	.64	-.84	.10	22.28
47	.14	1.25	.27	.13	15.68
48	.21	.65	-1.43	.10	25.64
49	.13	.93	1.79	.17	7.94
50	.17	1.29	.26	.22	44.51
51	.16	.97	.32	.29	8.45
52	.17	.50	.47	.10	29.29
53	.08	1.10	.65	.15	14.59
54	.20	1.30	-.21	.28	15.25
55	.27	1.50	-.16	.10	17.81
56	.12	.20	.86	.10	14.02
57	.15	1.50	-.16	.24	12.36
58	.13	1.50	-.29	.20	24.72

Table 8

Mean and standard deviations of TOEFL scores for
the September, October, November, and December administrations

	September				October				November				December			
	Section				Section				Section				Section			
	1	2	3	Total	1	2	3	Total	1	2	3	Total	1	2	3	Total
Mean	51	49	50	500	52	49	49	502	53	49	50	507	52	50	50	508
S.D.	6.0	7.1	7.2	60	6.6	7.0	7.3	63	6.7	7.3	6.9	63	6.6	7.2	6.9	62
N		5332				10793				21359				6167		
Form		3FTF9				3FTF10				3FTF11				3FTF12		

Table 9
Mean and standard deviations of item parameter on the September, October, November, and December 1983 administrations on Section 2

	Mean			Standard Deviation		
	a	b	c	a	b	c
September	1.02	-.24	.19	.33	.93	.07
October	1.02	-.22	.17	.33	.91	.08
November	1.05	-.32	.16	.33	.88	.09
December	1.03	-.24	.18	.36	.98	.08

Figure 1

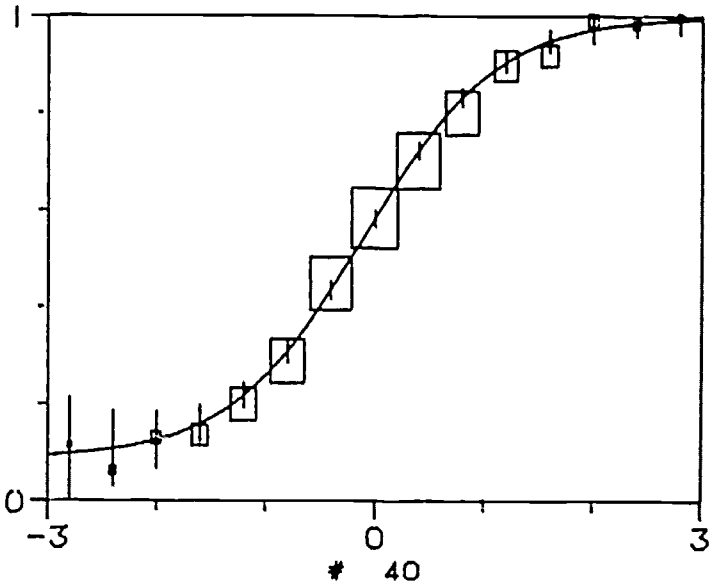


Figure 2

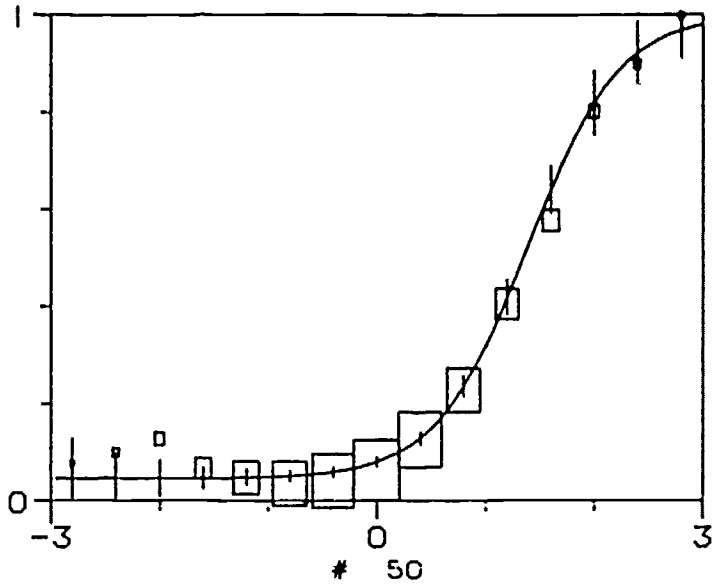


Figure 3

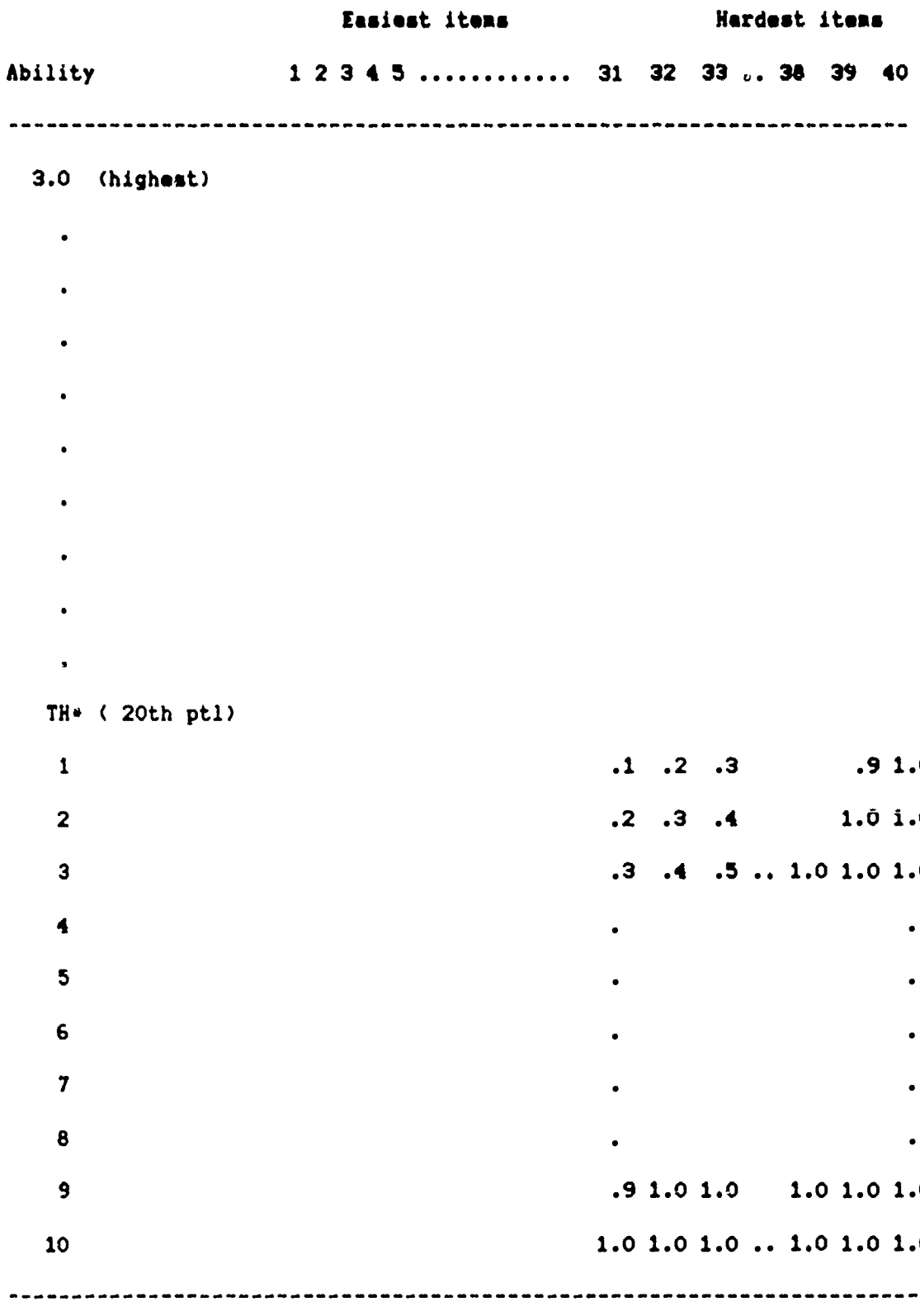

Easiest items
Hardest items
Ability
1 2 3 4 5 ..... 31 32 33 .. 38 39 40
3.0 (highest)
TH* (20th ptl)
1 .1 .2 .3 .9 1.0
2 .2 .3 .4 1.0 1.0
3 .3 .4 .5 .. 1.0 1.0 1.0
4 .
5 .
6 .
7 .
8 .
9 .9 1.0 1.0 1.0 1.0 1.0
10 1.0 1.0 1.0 .. 1.0 1.0 1.0

Figure 4

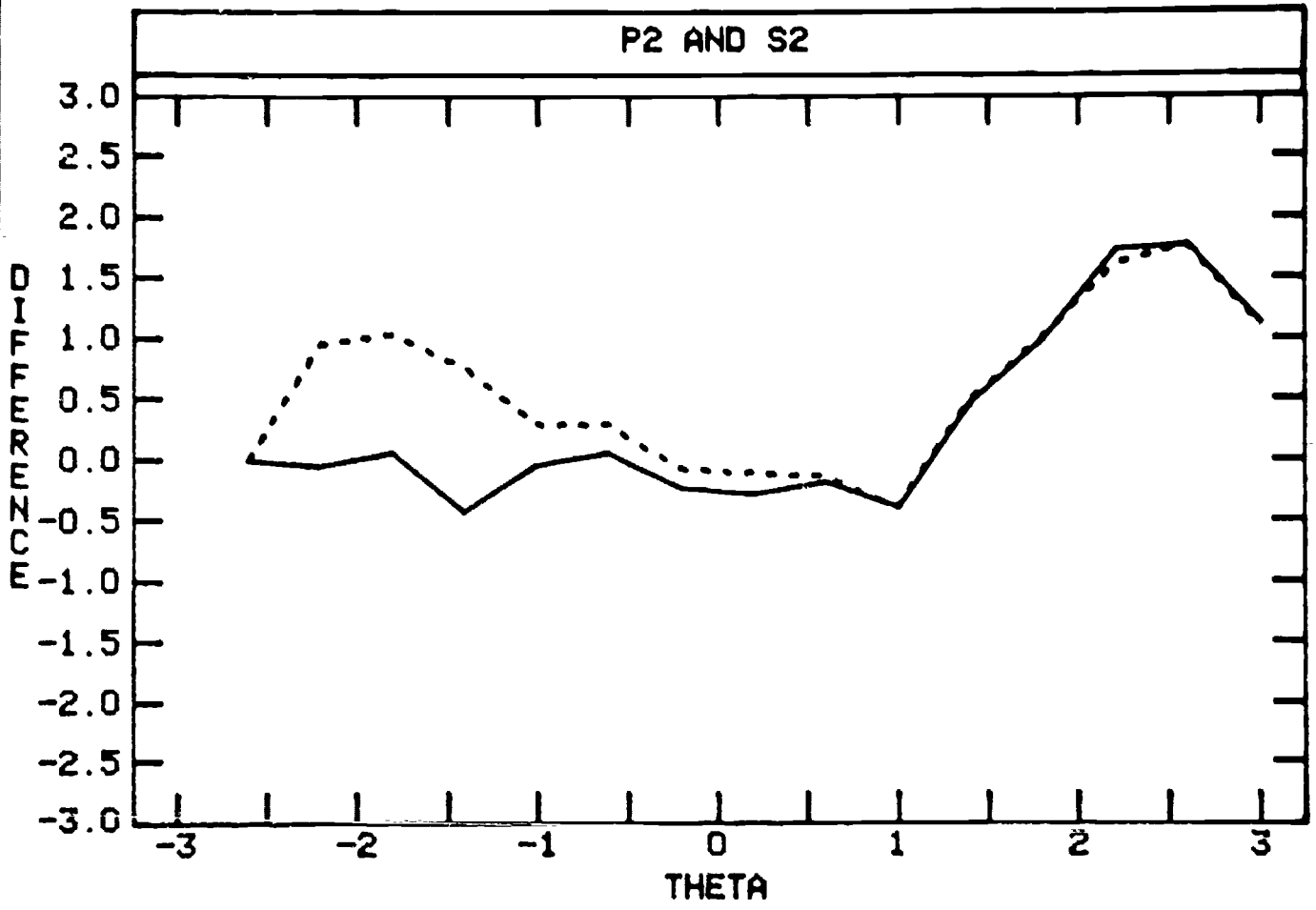

P2 AND S2
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
DIFFERENCE
-3
-2
-1
0
1
2
3
THETA

Figure 5

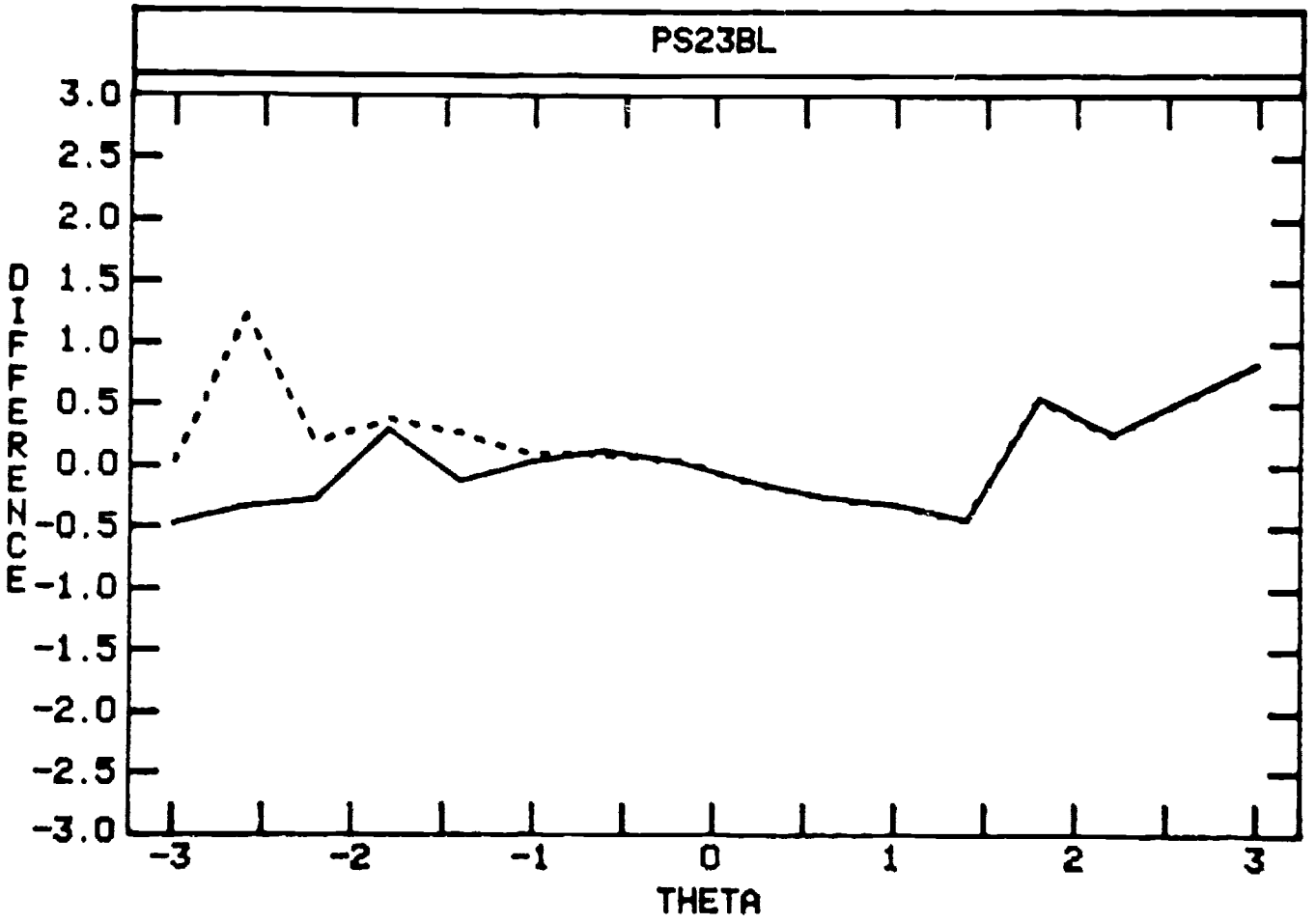

PS23BL
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
DIFFERENCE
-3
-2
-1
0
1
2
3
THETA

Figure 6

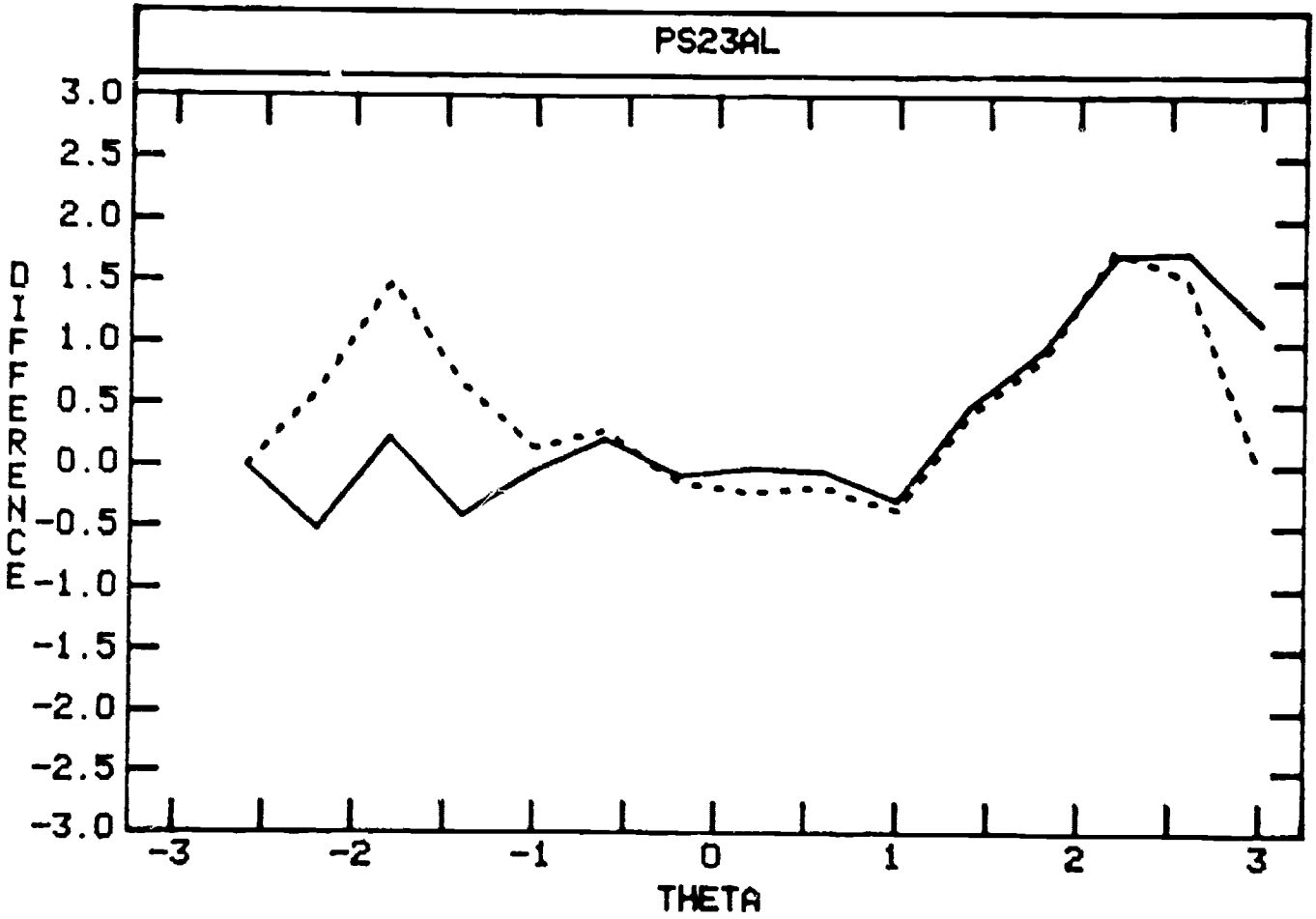

PS23AL
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
DIFFERENCE
-3
-2
-1
0
1
2
3
THETA

Figure 7

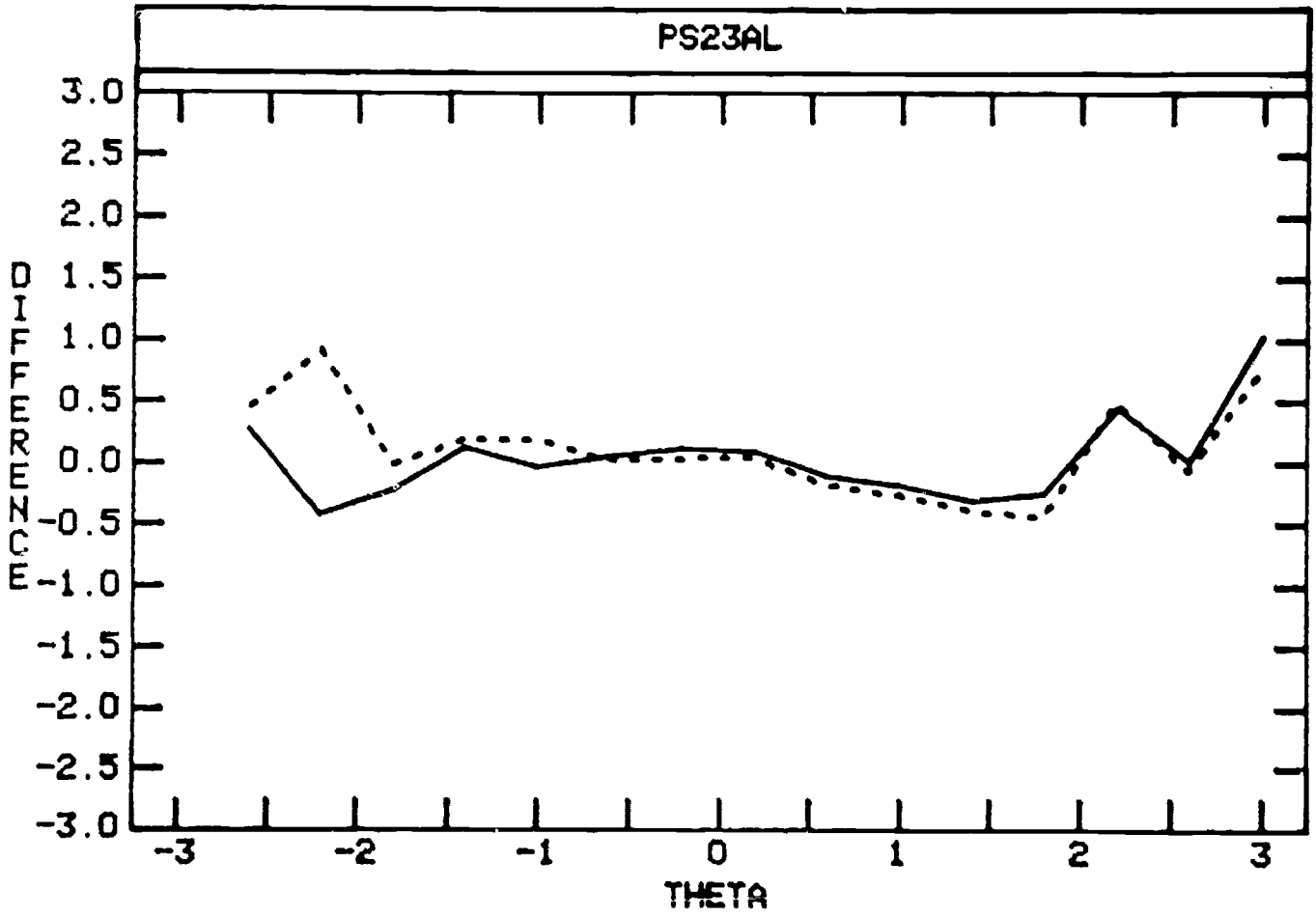

PS23AL
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
DIFFERENCE
-3
-2
-1
0
1
2
3
THETA

Figure 8

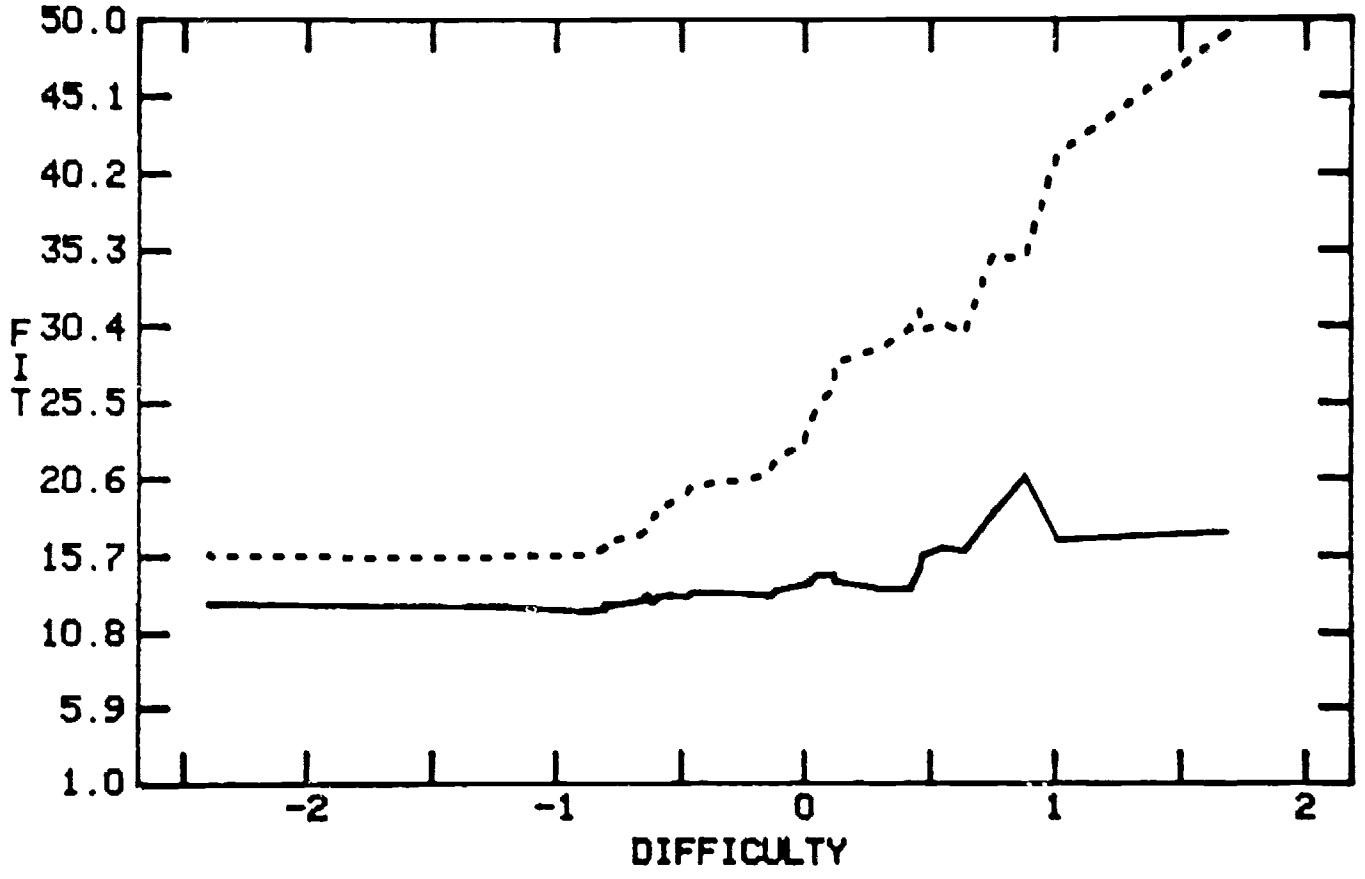


Figure 9

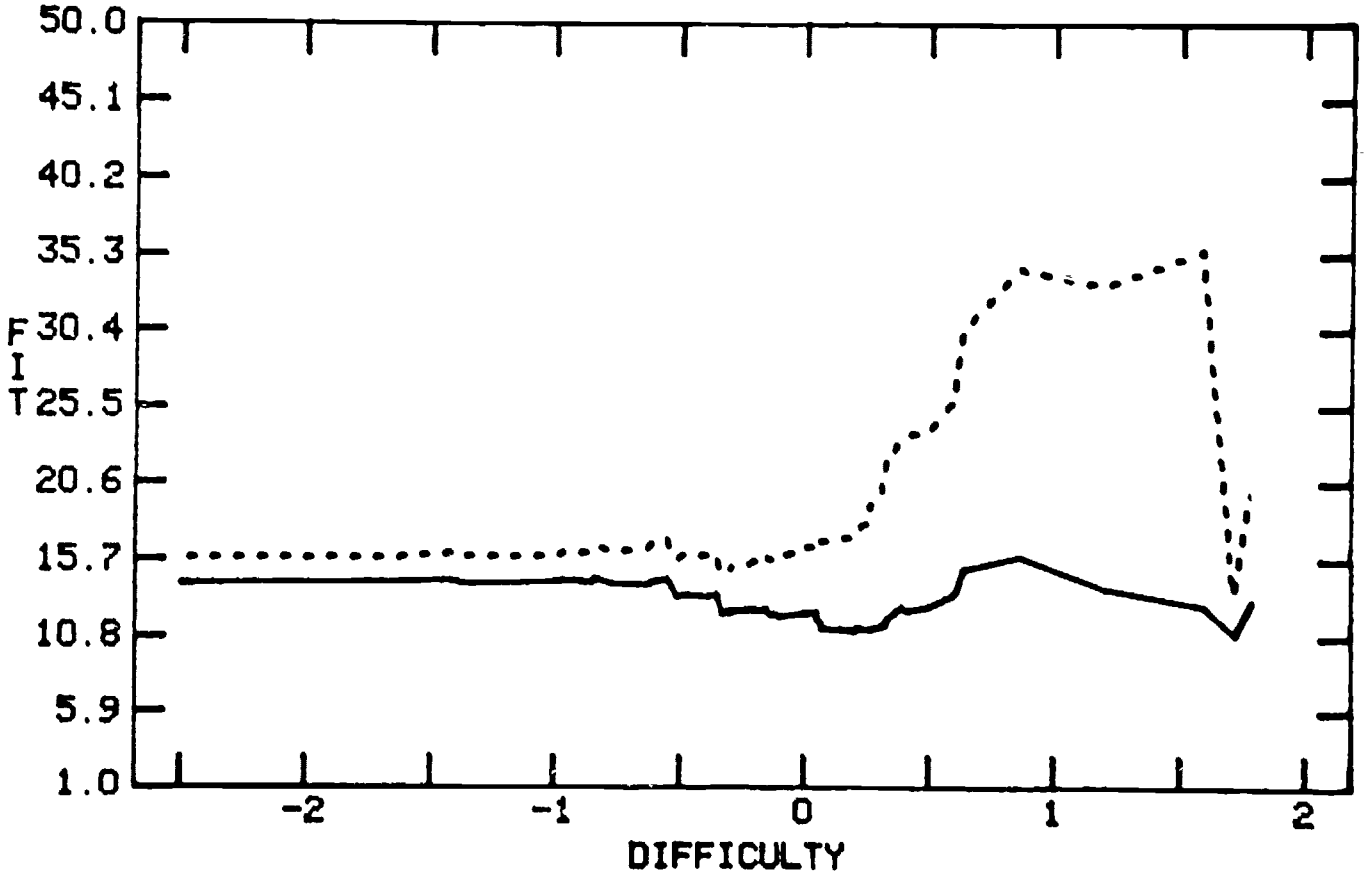


Figure 10

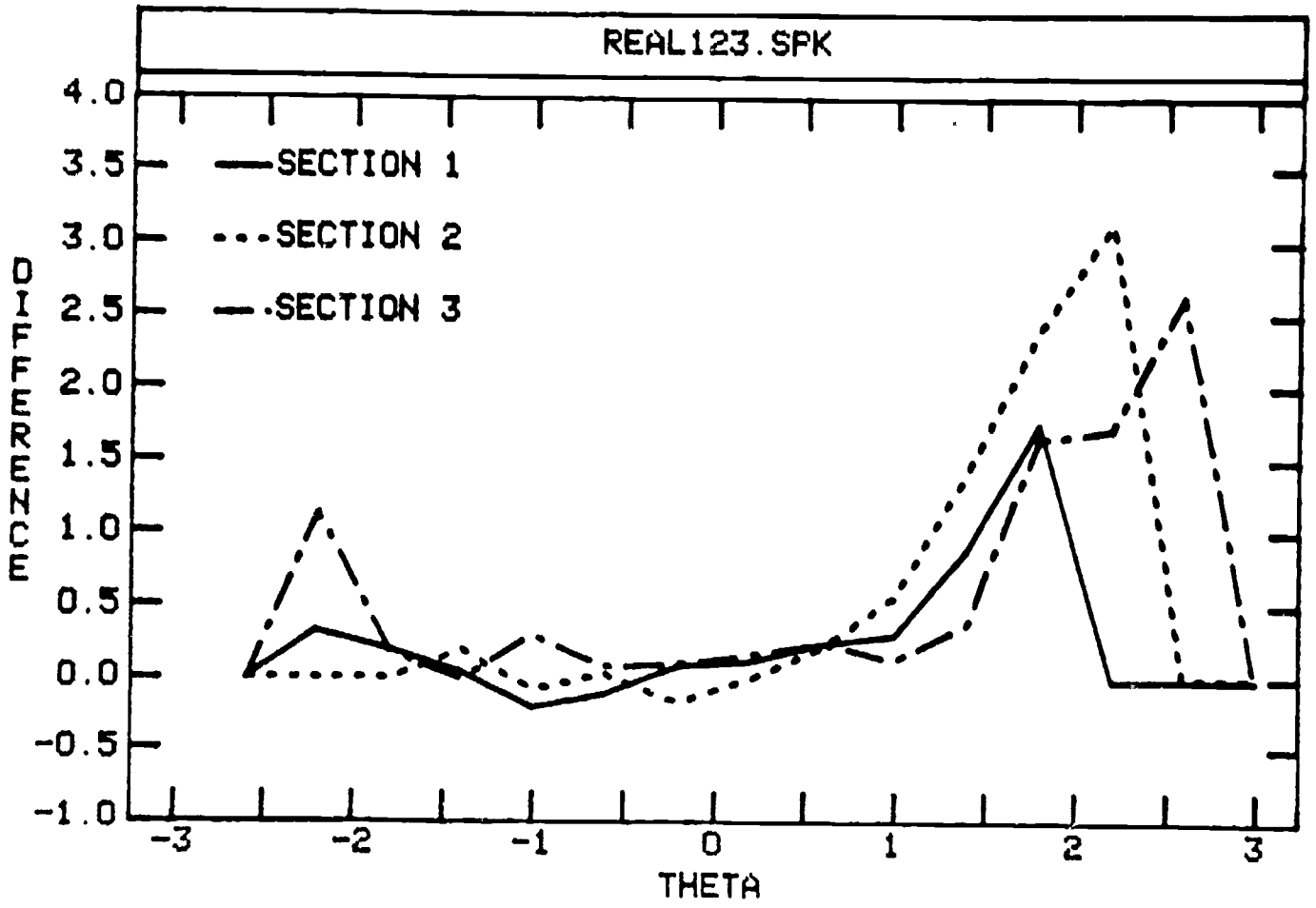

REAL123.SPK
SECTION 1
SECTION 2
SECTION 3
DIFFERENCE
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-3
-2
-1
0
1
2
3
THETA

Figure 11

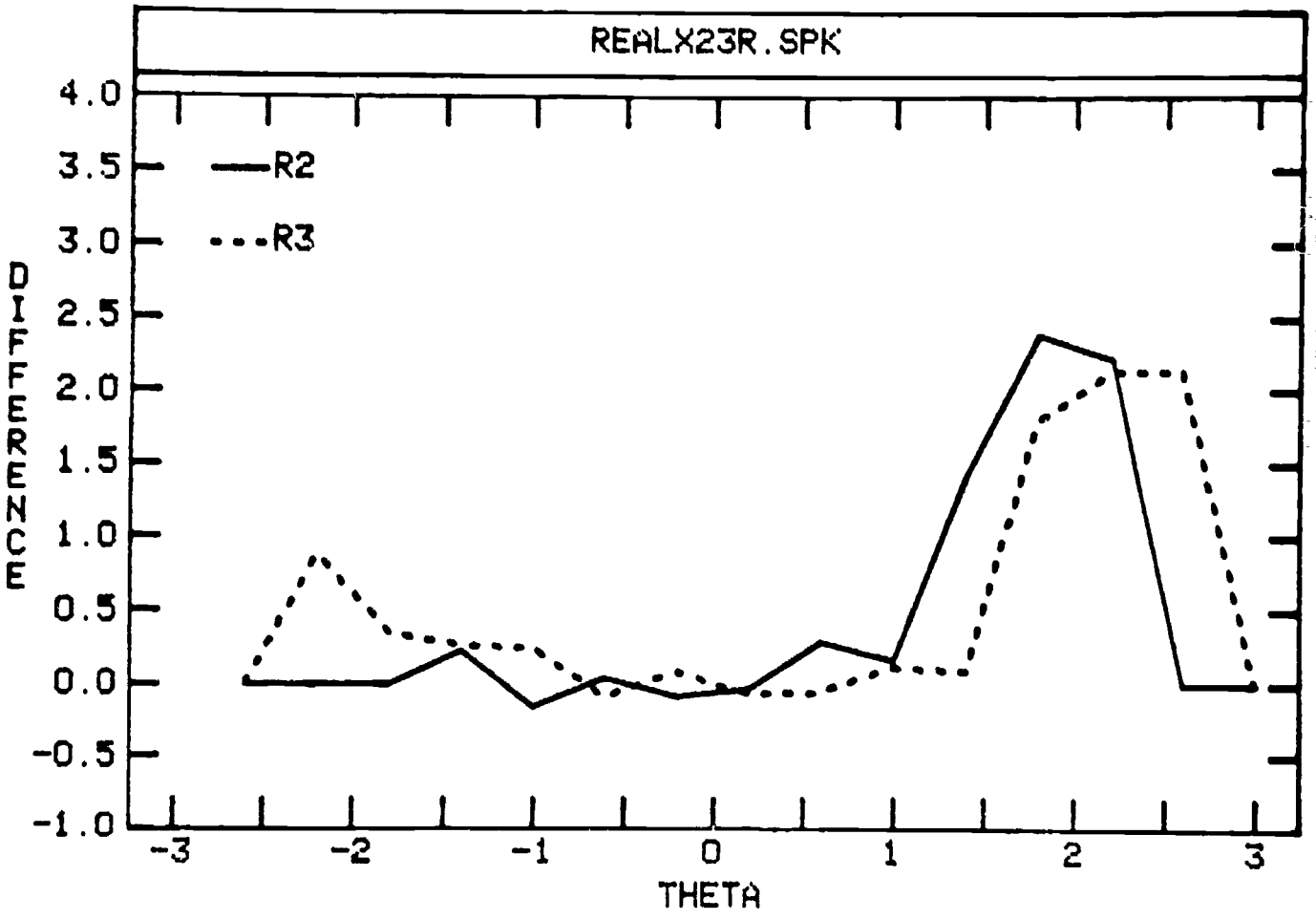
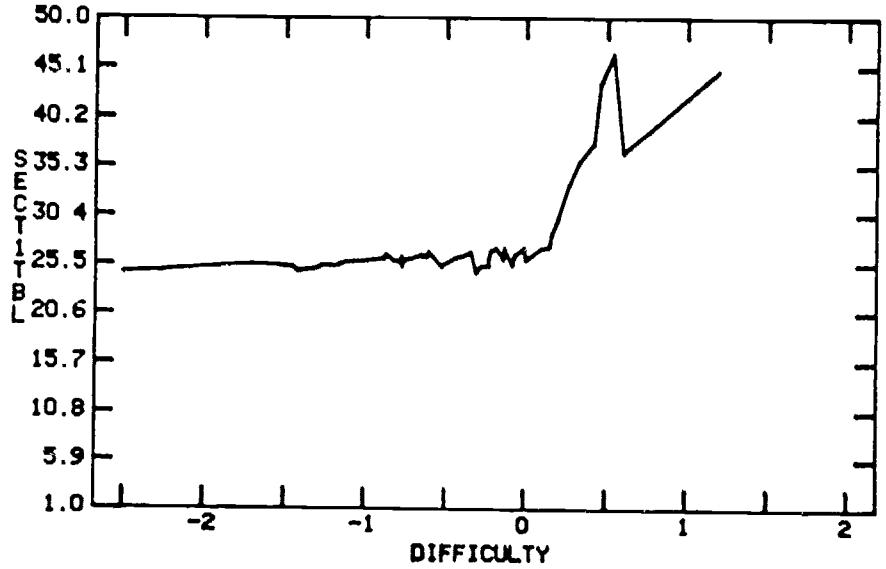
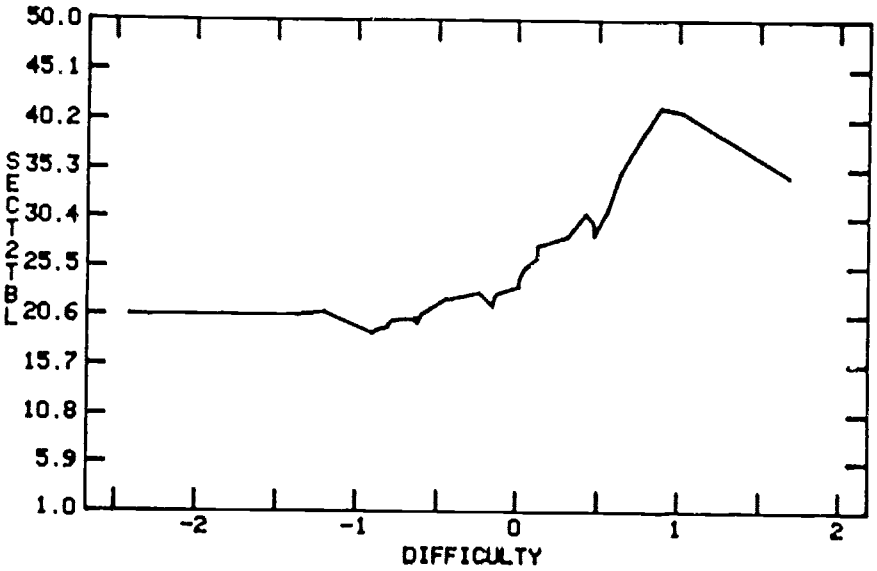

REALX23R.SPK
R2
R3
DIFFERENCE
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-3
-2
-1
0
1
2
3
THETA

Figure 12


50.0
45.1
40.2
SECT1TBL 35.3
30 4
25.5
20.6
15.7
10.8
5.9
1.0
-2
-1
0
1
2
DIFFICULTY


50.0
45.1
40.2
SECT2TBL 35.3
30.4
25.5
20.6
15.7
10.8
5.9
1.0
-2
-1
0
1
2
DIFFICULTY

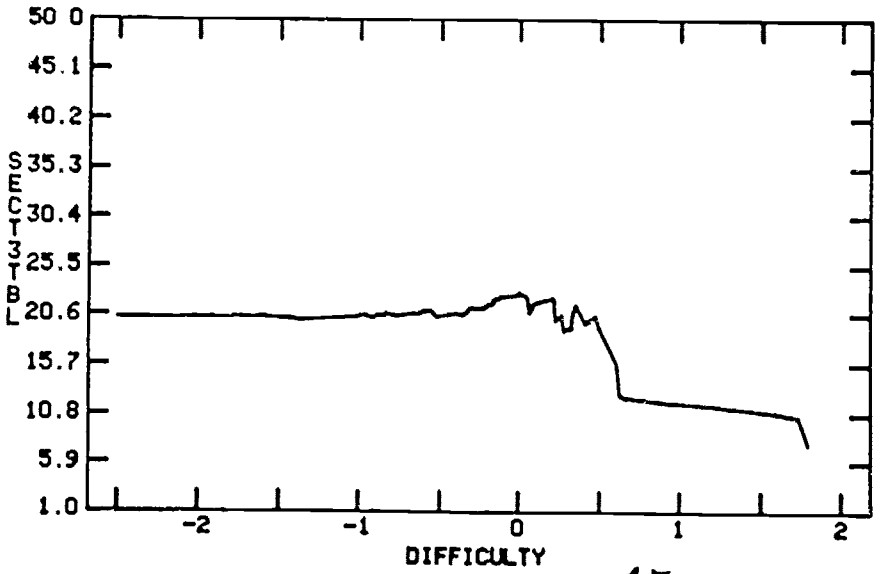

50 0
45.1
40.2
SECT3TBL 35.3
30.4
25.5
20.6
15.7
10.8
5.9
1.0
-2
-1
0
1
2
DIFFICULTY

Figure 13

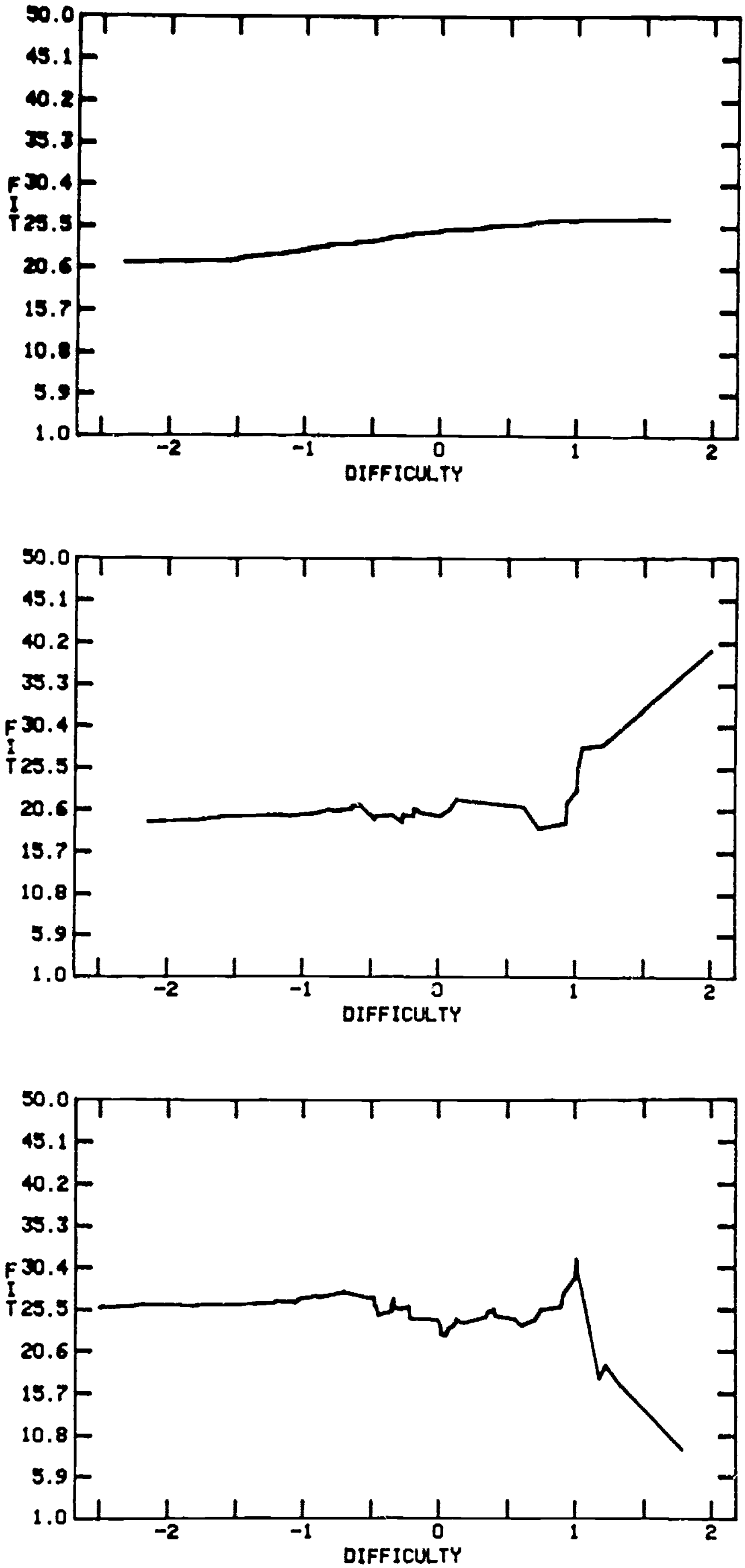


Figure 14

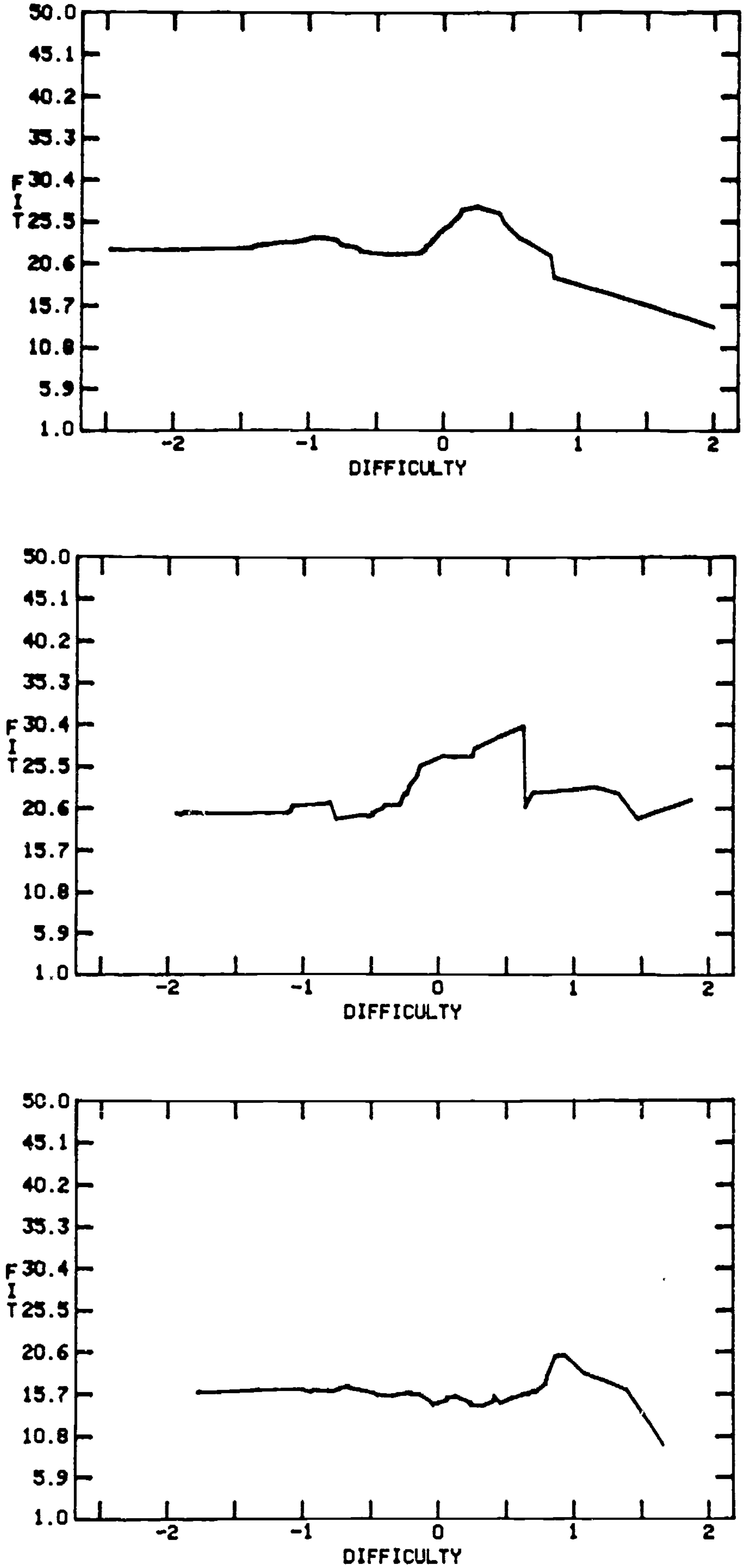

50.0
45.1
40.2
35.3
30.4
25.5
20.6
15.7
10.8
5.9
1.0
FIT
-2
-1
0
1
2
DIFFICULTY
50.0
45.1
40.2
35.3
30.4
25.5
20.6
15.7
10.8
5.9
1.0
FIT
-2
-1
0
1
2
DIFFICULTY
50.0
45.1
40.2
35.3
30.4
25.5
20.6
15.7
10.8
5.9
1.0
FIT
-2
-1
0
1
2
DIFFICULTY

Figure 15

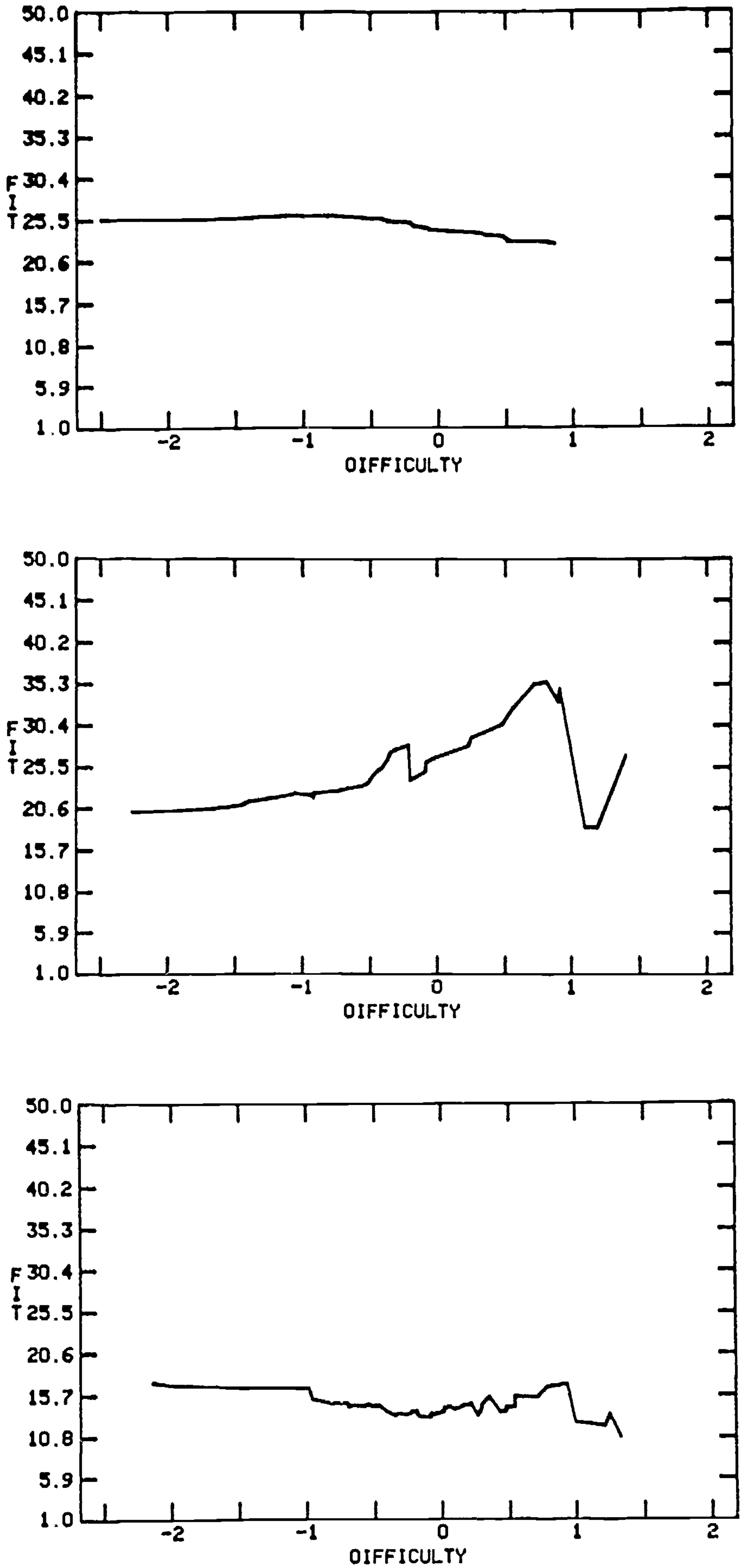
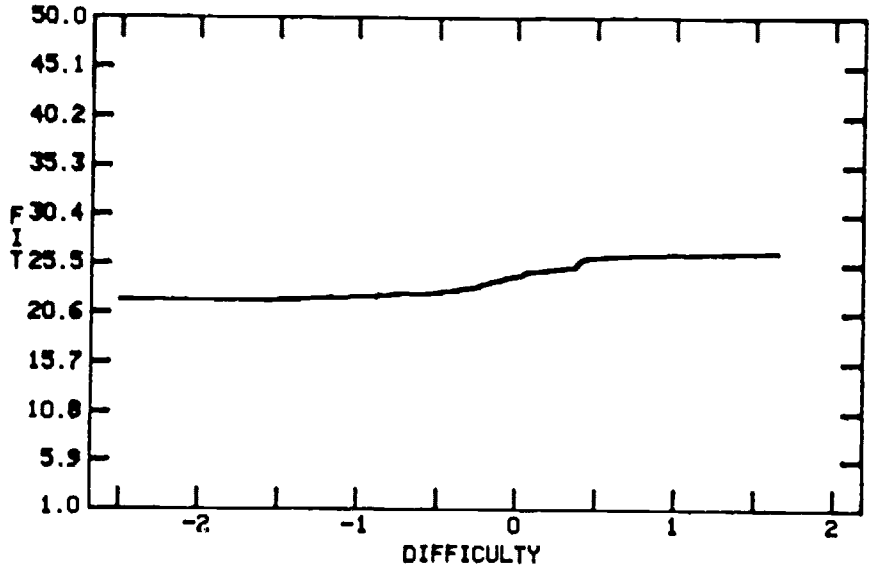
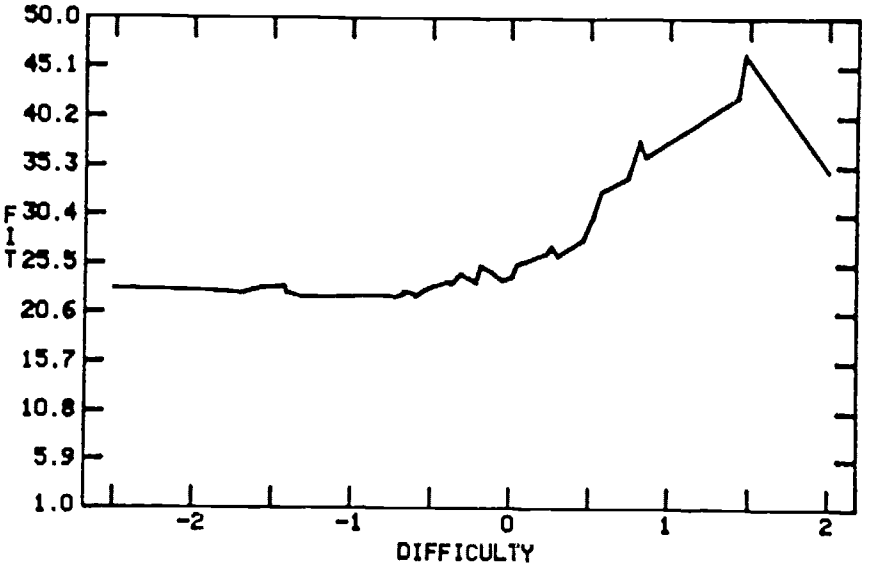
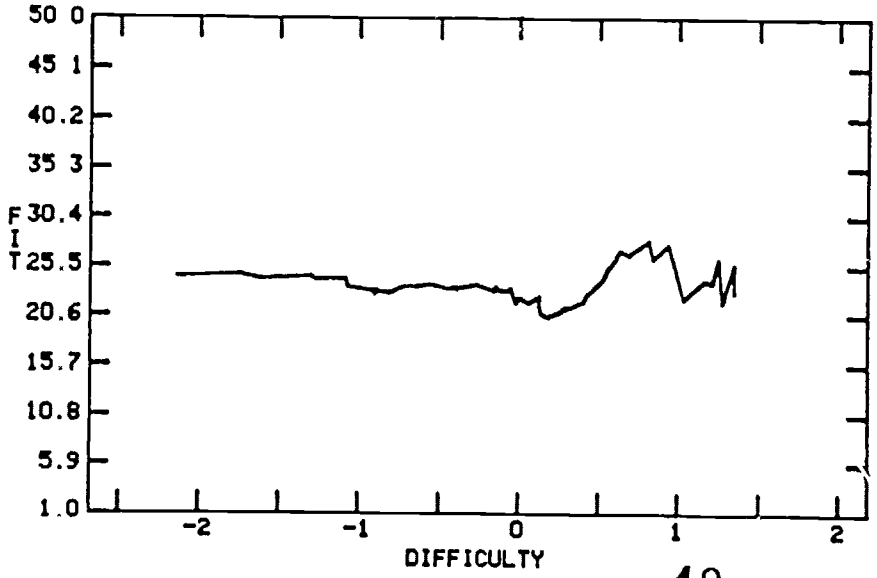


Figure 16


50.0
45.1
40.2
35.3
F 30.4
I
T 25.5
20.6
15.7
10.8
5.9
1.0
-2
-1
0
1
2
DIFFICULTY


50.0
45.1
40.2
35.3
F 30.4
I
T 25.5
20.6
15.7
10.8
5.9
1.0
-2
-1
0
1
2
DIFFICULTY


50 0
45 1
40.2
35 3
F 30.4
I
T 25.5
20.6
15.7
10.8
5.9
1.0
-2
-1
0
1
2
DIFFICULTY

Appendix A

Table A1
Item parameters estimates and Q for section 1
of the September 1983 administration

POS	A	B	C	Q
1	.88	-1.51	.08	37.75
2	.46	-1.90	.08	30.80
3	.89	-1.27	.08	30.62
4	1.24	-1.05	.08	20.50
5	1.10	-1.06	.14	13.50
6	.78	-.40	.08	65.89
7	.76	-.84	.08	61.64
8	.99	.17	.13	30.79
9	.79	.80	.24	49.41
10	.81	1.67	.18	36.63
11	.51	-2.33	.08	8.01
12	1.26	-1.31	.18	8.94
13	1.14	-.26	.06	11.60
14	1.02	-.64	.11	21.18
15	.89	-.43	.17	34.94
16	1.45	-.38	.22	53.24
17	.46	.47	.08	47.19
18	.92	-.01	.01	85.45
19	.67	-.98	.08	19.73
20	.87	.65	.26	15.81
21	.40	-.90	.08	19.34
22	1.15	.65	.23	17.99
23	.90	-.63	.37	11.92
24	1.14	-.98	.18	20.02
25	1.20	-1.47	.08	12.65
26	.95	-.05	.29	7.74
27	1.29	-1.01	.12	4.47
28	1.13	-.84	.07	9.80
29	1.50	-.52	.13	6.20
30	1.37	.38	.19	22.81
31	.36	.03	.08	22.64
32	1.12	-.83	.11	8.50
33	.94	.70	.29	6.18
34	1.03	.29	.14	10.64
35	1.50	-1.16	.13	23.19
36	.85	-1.17	.08	7.78
37	1.29	-.62	.12	8.49
38	1.13	-.45	.14	4.26
39	1.50	-.35	.28	5.36
40	1.17	-.21	.21	6.94
41	1.50	.31	.11	13.75
42	.46	-1.45	.08	5.66
43	.47	-1.50	.08	7.49
44	.89	-1.58	.08	7.39
45	.27	-.21	.08	23.24
46	.94	1.05	.16	20.21
47	1.07	-.26	.09	12.45
48	.56	.71	.19	15.49
49	.89	-.94	.05	29.47
50	1.13	.34	.36	7.64

Table A2
Item parameters estimates and Q for section 2 of the September 1983 administration

POS	A	B	C	Q
1	.82	-1.62	.17	16.50
2	.42	-.18	.17	21.76
3	1.08	-.91	.17	8.89
4	.96	-.46	.21	17.40
5	1.26	-.49	.03	31.75
6	1.50	.31	.32	26.18
7	.58	.62	.17	41.18
8	1.00	1.02	.08	17.86
9	1.32	-.95	.26	17.22
10	1.48	-.81	.15	24.39
11	.76	-.51	.17	20.24
12	.54	-.95	.17	18.21
13	.55	-.58	.17	49.10
14	.97	.01	.39	11.87
15	.50	-.34	.17	35.89
16	1.02	-.27	.20	5.55
17	1.17	.94	.18	13.94
18	1.50	2.10	.22	40.65
19	1.23	.08	.34	7.74
20	.55	-1.17	.17	22.56
21	1.26	-.14	.13	24.36
22	.86	-2.14	.17	10.98
23	1.35	1.20	.12	15.85
24	1.03	-1.75	.17	8.41
25	1.26	-.26	.13	21.35
26	1.15	-.76	.19	15.51
27	.89	1.01	.19	12.98
28	.66	-1.60	.17	13.96
29	.93	-1.08	.17	13.33
30	.35	-.15	.17	25.34
31	1.25	-.64	.24	15.86
32	1.50	-.64	.18	12.20
33	.75	.13	.15	24.71
34	1.38	-.47	.10	11.32
35	1.19	-.18	.31	7.53
36	1.50	.73	.10	14.37
37	.97	1.05	.14	27.28
38	.97	.93	.20	5.01

Table A3

Item parameters estimates and Q for section 3 of the September 1983 administration

POS	A	B	C	Q
1	.40	-2.30	.25	16.66
2	.47	-2.76	.25	6.76
3	1.50	-1.06	.14	26.04
4	1.23	-1.22	.10	16.41
5	.36	-2.19	.25	27.78
6	1.38	-.52	.23	28.24
7	.78	-.46	.25	65.00
8	1.50	.13	.34	30.44
9	1.29	.02	.19	29.03
10	1.10	.41	.27	31.50
11	1.50	1.22	.21	23.43
12	.99	.89	.07	21.81
13	.92	1.05	.35	68.81
14	.89	1.01	.18	40.35
15	1.50	1.33	.28	24.28
16	.83	-.13	.47	27.39
17	.91	-.48	.50	64.92
18	1.50	-.33	.40	27.28
19	.31	.55	.25	34.35
20	1.36	.60	.08	22.00
21	.74	.02	.10	42.26
22	1.49	-.34	.06	24.63
23	1.05	-.90	.06	31.12
24	.86	-.47	.10	30.12
25	1.50	.40	.09	37.66
26	1.30	-.01	.18	49.61
27	1.47	-.22	.11	67.39
28	1.17	-.33	.05	60.18
29	1.50	.34	.27	13.70
30	.79	.11	.09	10.51
31	.91	-1.36	.25	19.63
32	.65	-.35	.25	7.35
33	1.08	-.87	.19	4.54
34	.84	-1.05	.06	7.74
35	.96	-.45	.12	11.93
36	1.38	-.28	.15	20.05
37	.41	.06	.25	12.18
38	.73	-.34	.25	11.72
39	1.13	.63	.18	17.61
40	1.11	1.17	.37	13.18
41	.84	.70	.25	10.24
42	1.50	1.78	.30	9.12
43	.96	.75	.12	21.64
44	1.08	1.00	.20	14.49
45	.80	.07	.15	12.60
46	.53	-.88	.25	26.02
47	.11	.17	.20	11.07
48	.98	.36	.18	19.35
49	1.11	-1.03	.21	10.57
50	.80	-1.20	.25	33.11
51	.71	1.02	.16	40.08
52	.57	-1.84	.25	17.85
53	.68	.91	.15	11/91
54	1.25	.05	.11	13.63
55	.91	-.70	.12	52.37
56	1.37	-.35	.19	8.31
57	1.12	-.21	.22	26.94
58	.63	.90	.16	16.66

Table A4
Item parameters estimates and Q for section 1
of the October 1983 administation

POS	A	B	C	Q
1	.54	-2.47	.08	8.82
2	1.03	-1.40	.08	32.08
3	.86	-1.36	.08	15.16
4	.87	-.13	.24	5.68
5	.78	-1.10	.08	19.57
6	.98	-1.03	.23	11.05
7	1.22	.41	.15	24.27
8	1.28	2.07	.19	10.08
9	.53	.00	.08	17.33
10	.68	-2.02	.08	20.95
11	.99	-.78	.16	21.84
12	.55	-.16	.08	20.00
13	.61	-.66	.08	12.19
14	.61	.27	.0	44.24
15	.75	-.96	.08	27.81
16	.91	.44	.24	22.46
17	.95	.55	.22	29.64
18	1.19	-1.21	.22	11.30
19	.48	.25	.08	10.18
20	.96	.79	.42	71.95
21	1.08	-.04	.27	20.03
22	1.37	-.51	.35	17.21
23	.85	-.62	.38	41.84
24	.77	1.44	.33	31.70
25	1.09	-.10	.37	9.50
26	.91	-1.42	.08	11.57
27	.84	-1.74	.08	27.45
28	.99	-.76	.08	5.76
29	1.13	-.62	.17	4.53
30	.71	-.87	.08	11.44
31	.77	-.18	.23	10.79
32	.93	-.09	.09	14.17
33	1.32	.12	.20	8.48
34	.49	-.98	.08	21.47
35	.31	-1.42	.08	24.57
36	1.50	.81	.15	13.90
37	1.06	-1.00	.13	22.45
38	1.49	.05	.30	17.23
39	.69	-1.33	.08	6.07
40	.67	-.82	.08	16.75
41	1.50	.23	.25	12.23
42	1.50	.13	.05	12.84
43	1.43	.08	.14	20.06
44	1.50	-.42	.11	20.61
45	.46	-.93	.08	28.42
46	1.03	-.13	.12	24.50
47	1.29	-.91	.11	23.25
48	1.19	-.06	.11	53.72
49	1.23	-.79	.15	27.30
50	1.25	-.35	.26	4.77

Table A5
Item parameters estimates and Q for section 2
of the October 1983 administration

POS	A	B	C	Q
1	.88	-1.88	.14	22.84
2	.72	.25	.09	19.21
3	1.40	-.76	.13	8.09
4	1.42	-1.11	.12	10.15
5	1.11	-.41	.38	15.04
6	1.50	1.00	.14	21.86
7	.89	.69	.13	20.50
8	1.30	.10	.10	25.52
9	.96	-1.09	.14	6.78
10	1.32	-.22	.20	6.08
11	.55	-1.60	.14	15.86
12	.87	-.84	.14	14.05
13	.38	-1.90	.14	11.68
14	1.50	-.80	.26	78.73
15	.62	-.58	.14	20.90
16	.91	-.16	.22	6.98
17	.81	1.15	.20	25.70
18	.88	.62	.03	95.64
19	.30	1.87	.0	21.63
20	.80	-.49	.12	29.91
21	1.09	-1.94	.14	23.41
22	.88	-.49	.09	8.79
23	1.10	-.14	.17	10.81
24	1.48	.26	.18	14.09
25	1.50	1.33	.07	28.00
26	1.37	-.51	.21	9.42
27	1.41	-.24	.24	6.82
28	.48	-.25	.14	26.39
29	1.50	-.29	.16	15.83
30	1.05	-.40	.27	20.84
31	.76	-.47	.16	7.90
32	.99	-.28	.22	9.64
33	1.01	-1.08	.07	12.84
34	.89	-.27	.15	11.66
35	1.21	.03	.16	27.39
36	1.09	1.47	.27	17.30
37	1.01	.63	.29	11.13
38	.76	.45	.29	19.94

Table A6
Item parameters estimates and Q for section 3
of the October 1983 administration

POS	A	B	C	Q
1	1.26	-1.37	.18	9.32
2	.58	-1.78	.13	6.64
3	1.28	-.92	.33	19.33
4	1.08	-.82	.29	22.31
5	.61	-.66	.13	24.09
6	1.40	.06	.23	10.94
7	.62	-.81	.13	19.39
8	1.32	.13	.21	17.86
9	.77	.86	.19	19.76
10	.79	.65	.26	17.41
11	.58	1.39	.22	22.43
12	1.04	.93	.21	26.21
13	1.25	.80	.24	6.18
14	1.49	1.07	.20	22.38
15	.70	.23	.30	9.58
16	1.50	-.74	.47	12.58
17	1.34	-.46	.46	21.28
18	.34	.41	.13	26.97
19	.56	.66	.11	12.32
20	.81	-1.00	.46	30.99
21	.76	.60	.19	13.37
22	1.18	-.00	.29	11.67
23	1.20	-.20	.28	16.63
24	1.50	-.03	.19	14.81
25	1.50	.14	.10	27.00
26	.53	-.47	.13	23.41
27	1.40	-.20	.24	20.81
28	1.50	-.04	.26	8.58
29	.28	.50	.13	12.35
30	1.50	-.33	.39	8.62
31	.86	-.66	.18	32.20
32	.56	-.94	.13	8.23
33	1.05	-.78	.24	3.24
34	1.44	-.73	.33	2.52
35	.60	-.82	.13	8.48
36	.80	-.25	.16	13.36
37	1.31	-.14	.14	39.41
38	.79	-.57	.09	28.23
39	1.13	.24	.30	17.51
40	.85	-.06	.12	31.18
41	.84	.62	.25	14.51
42	1.23	.53	.23	10.88
43	.86	.24	.06	20.70
44	1.39	.33	.21	6.05
45	1.50	1.66	.35	9.87
46	1.06	-1.63	.13	6.05
47	.68	-1.77	.13	18.87
48	1.27	.06	.14	6.45
49	1.07	-.22	.22	16.74
50	.40	-.34	.13	13.22
51	1.30	-.00	.17	7.88
52	1.03	.45	.25	10.38
53	.67	.79	.20	11.68
54	1.20	.70	.40	17.15
55	1.30	.71	.27	8.87
56	1.50	.32	.32	13.03
57	1.50	.22	.31	22.76
58	1.29	.40	.26	5.19

Table A7

Item parameters estimates and Q for section 1 of the November 1983 administration

POS	A	B	C	Q
1	1.25	-1.40	.34	8.82
2	1.05	-1.14	.43	7.40
3	.90	-1.32	.19	8.06
4	1.07	-1.14	.27	12.50
5	.96	-.38	.01	129.04
6	.85	-.79	.19	36.49
7	.95	-.33	.00	80.35
8	1.19	.26	.11	11.31
9	1.50	-.77	.27	32.05
10	1.49	.48	.18	76.52
11	.94	-1.40	.19	13.74
12	.68	-.49	.19	84.78
13	1.06	.07	.24	33.50
14	.94	-1.78	.19	33.63
15	.63	-1.35	.19	69.69
16	.75	.86	.17	25.85
17	.62	.51	.05	25.54
18	1.27	.50	.12	28.16
19	.25	-.19	.19	53.06
20	1.29	-.70	.16	12.09
21	1.24	-.41	.14	14.75
22	.47	-2.77	.19	20.90
23	1.09	.51	.29	6.78
24	.69	.35	.28	26.56
25	.86	-1.13	.0	5.55
26	.90	-1.10	.19	7.70
27	1.24	-.81	.17	14.63
28	1.19	-.20	.33	12.12
29	.98	-1.52	.19	27.26
30	.94	-.12	.15	15.21
31	.69	-.42	.19	19.40
32	1.50	.49	.18	15.76
33	1.23	-.17	.10	11.49
34	1.25	-.04	.18	6.11
35	.77	-1.82	.19	18.44
36	1.25	-1.22	.17	10.91
37	1.40	-.38	.16	10.42
38	1.24	-.65	.17	12.41
39	1.32	-.20	.12	12.57
40	.92	.81	.12	18.09
41	1.30	-.06	.14	13.55
42	.57	-.08	.19	25.34
43	.67	-1.60	.19	20.19
44	.87	-1.10	.19	17.29
45	.91	-1.37	.19	12.23
46	1.13	-.57	.0	3.32
47	.84	-1.29	.19	16.08
48	.47	-.87	.19	14.75
49	1.38	.33	.13	12.41
50	1.12	-.65	.14	17.28

Table A8

Item parameters estimates and Q for section 2 of the November 1983 administration

POS	A	B	C	Q
1	.82	-1.87	.14	10.27
2	.60	-.21	.14	89.37
3	.69	-1.45	.14	7.50
4	1.49	-.37	.12	11.32
5	1.15	-.71	.01	12.41
6	1.17	.23	.09	18.44
7	.47	.91	.14	83.40
8	1.47	.72	.05	33.27
9	1.27	-1.09	.05	13.62
10	.91	-.73	.32	20.56
11	.87	-1.05	.14	30.12
12	1.36	-1.40	.36	12.35
13	.85	-.08	.34	21.18
14	.95	-.51	.10	8.36
15	.67	-.93	.14	30.58
16	1.20	1.19	.19	9.91
17	.59	-.32	.14	17.39
18	1.29	-.20	.03	10.94
19	1.40	-.35	.07	22.62
20	1.50	.81	.37	47.41
21	1.50	1.40	.22	26.94
22	1.10	-1.57	.14	9.98
23	.64	-2.26	.14	11.04
24	1.04	1.10	.27	18.46
25	1.50	.25	.19	14.77
26	1.12	-.45	.11	16.69
27	1.50	.90	.18	27.21
28	.69	-.92	.14	14.57
29	1.33	-.92	.10	7.74
30	1.05	-1.27	.14	6.41
31	1.20	-.52	.09	19.29
32	1.37	-.48	.17	13.56
33	1.02	-.41	.10	6.28
34	.69	-.03	.08	10.14
35	.77	-.57	.0	15.95
36	.74	.56	.25	12.61
37	.58	-.09	.14	10.49
38	1.50	.48	.32	15.04

Table A9
Item parameters estimates and Q for section 3
of the November 1983 administration

POS	A	B	C	Q
1	.94	-2.01	.11	31.41
2	1.23	-.96	.32	39.82
3	1.13	-.98	.20	90.23
4	.82	-.42	.21	21.85
5	1.04	-.52	.0	12.60
6	.36	-.55	.11	18.19
7	.81	.06	.31	21.49
8	.74	.54	.08	16.75
9	.74	-.33	.11	20.78
10	.48	-.39	.11	24.28
11	1.50	.29	.10	11.13
12	.73	.43	.15	13.17
13	1.06	.93	.32	34.48
14	1.41	1.33	.12	10.90
15	1.33	.78	.21	15.24
16	1.30	-.16	.29	14.98
17	1.09	-.69	.28	11.12
18	1.06	-.24	.14	5.94
19	.84	-.70	.11	39.06
20	.66	.15	.26	13.17
21	1.07	-.35	.35	5.81
22	.87	.30	.32	7.13
23	1.50	.28	.27	5.11
24	1.44	-.05	.18	11.30
25	1.23	-.46	.22	41.68
26	1.50	-.08	.12	3.19
27	1.39	.47	.23	9.46
28	1.23	.35	.07	36.47
29	1.15	-.03	.25	8.93
30	1.31	-.82	.24	13.45
31	.65	-2.14	.11	32.24
32	1.41	-.79	.20	13.88
33	1.30	-.22	.25	9.19
34	1.50	-.18	.11	14.95
35	.62	-.57	.11	8.35
36	.51	-.18	.11	27.98
37	.58	-.53	.11	19.99
38	.76	.29	.20	12.07
39	.98	.47	.17	13.03
40	.61	.21	.04	41.58
41	1.03	.02	.20	10.19
42	.75	.71	.27	8.03
43	.77	1.21	.16	9.53
44	1.50	.26	.12	6.34
45	1.24	-.69	.25	9.12
46	1.16	.21	.13	9.30
47	.88	-1.53	.11	18.43
48	1.50	.09	.22	6.56
49	.96	1.25	.14	16.07
50	.80	.15	.26	13.24
51	1.10	-.26	.17	9.42
52	.90	.33	.26	10.69
53	1.16	.02	.27	11.23
54	.89	-.16	.15	18.25
55	1.16	-.08	.17	14.19
56	.80	.01	.29	4.26
57	1.43	.99	.18	13.87
58	1.50	.54	.25	4.96

Table A10

Item parameters estimates and Q for section 1 of the December 1983 administration

POS	A	B	C	Q
1	1.31	-1.69	.19	19.20
2	.86	-1.28	.51	23.87
3	1.17	-1.24	.16	7.56
4	.98	-.38	.25	9.55
5	1.17	-.23	.16	15.24
6	1.12	-.57	.13	19.13
7	.40	.38	.19	15.54
8	1.47	.04	.27	15.30
9	.84	-.47	.06	19.85
10	1.18	-.21	.07	42.13
11	1.26	.37	.14	13.57
12	.60	-2.66	.19	14.24
13	1.18	-.36	.20	33.29
14	1.04	-1.49	.19	25.11
15	.50	-.10	.21	45.11
16	.72	-.33	.28	11.38
17	1.50	.06	.24	45.20
18	.79	.64	.09	27.42
19	1.46	.45	.23	19.64
20	.88	-.49	.11	27.69
21	.61	-1.45	.19	29.50
22	.30	-.88	.19	38.38
23	1.06	-1.40	.19	30.41
24	.85	-.87	.19	16.95
25	1.15	.40	.33	8.75
26	1.01	-.24	.12	49.92
27	.78	-1.11	.19	51.07
28	1.07	-.77	.30	42.34
29	.91	-.26	.14	27.03
30	.95	-.74	.19	19.20
31	1.25	-.03	.14	17.46
32	.87	-.76	.14	10.11
33	1.41	-.18	.22	12.62
34	1.34	-.07	.17	18.44
35	1.29	.16	.11	14.71
36	.98	.03	.30	17.46
37	.48	-1.01	.19	19.49
38	.92	-1.50	.19	17.51
39	1.37	-.17	.17	12.06
40	.44	-.47	.19	18.00
41	1.07	.24	.13	14.52
42	.88	1.09	.12	26.32
43	.49	1.65	.19	46.29
44	1.04	.39	.19	31.69
45	.80	-.15	.25	20.46
46	.92	-.40	.20	8.54
47	.61	-1.18	.19	13.79
48	.75	.41	.15	20.48
49	.87	-.82	.19	4.69
50	.56	-1.01	.19	5.05

Table A11
Item parameters estimates and Q for section 2
of the December 1983 administration

POS	A	B	C	Q
1	.71	-1.41	.12	39.64
2	1.11	-1.42	.12	43.36
3	1.29	.30	.32	11.96
4	.63	-.55	.12	19.25
5	1.47	-.72	.25	14.20
6	1.50	-.21	.22	13.10
7	1.30	.04	.14	12.57
8	1.50	1.43	.13	33.66
9	.98	-1.31	.12	16.42
10	1.14	-.60	.37	9.12
11	.72	-1.57	.12	14.49
12	.45	-.40	.12	25.78
13	1.50	1.47	.29	65.70
14	.67	-.61	.12	29.89
15	.28	-2.54	.12	32.97
16	1.50	-.31	.27	38.85
17	.51	.85	.20	18.37
18	.94	.01	.21	8.51
19	.97	.81	.15	44.10
20	.56	2.35	.16	27.09
21	.89	-1.87	.12	29.33
22	1.50	-.67	.18	25.69
23	1.15	-.53	.22	5.68
24	1.43	.74	.07	15.66
25	1.50	-.18	.22	35.31
26	1.50	.57	.16	24.08
27	.87	-1.69	.12	5.79
28	.84	-.78	.12	28.39
29	1.14	-.67	.10	15.45
30	1.44	-.20	.17	7.14
31	.77	.26	.21	36.51
32	.98	-.38	.10	23.53
33	1.44	-.05	.09	19.34
34	.93	-.37	.19	7.09
35	.47	-.11	.12	36.46
36	.64	.52	.30	12.31
37	1.28	.23	.41	18.20
38	.85	.46	.28	9.80

Table A12
Item parameters estimates and Q for section 3
of the December 1983 administration

POS	A	B	C	Q
1	1.03	-1.74	.20	48.58
2	.63	-1.30	.20	37.86
3	1.50	-.72	.42	20.64
4	1.02	-.40	.10	27.45
5	1.09	-.44	.12	19.46
6	.96	-.57	.03	24.07
7	.92	-.06	.23	15.99
8	1.34	-.14	.21	33.34
9	.65	.40	.28	12.31
10	.93	.81	.42	44.63
11	.96	-.54	.10	41.44
12	1.14	.63	.10	31.77
13	1.14	.94	.20	65.00
14	1.37	1.03	.13	11.46
15	1.17	1.27	.13	14.61
16	1.34	-1.08	.48	61.95
17	.71	-2.14	.20	15.43
18	.69	-.89	.20	33.32
19	1.27	.48	.17	12.46
20	.78	.26	.35	15.53
21	.60	.69	.27	12.76
22	.73	-.34	.20	23.87
23	1.32	-.02	.32	20.17
24	1.50	-.31	.21	17.16
25	1.16	.60	.18	15.18
26	.65	-.27	.20	45.18
27	1.50	1.25	.23	38.96
28	.87	.18	.16	13.87
29	1.50	-.15	.25	13.07
30	.75	-.02	.10	13.28
31	1.05	-1.07	.14	44.37
32	.98	-.75	.25	14.94
33	.70	-1.62	.20	11.66
34	.68	-.81	.20	5.52
35	1.23	-.64	.08	15.24
36	1.27	-.14	.11	27.76
37	.94	.28	.27	20.09
38	.60	.06	.20	5.18
39	.53	.84	.21	14.67
40	1.30	.29	.21	8.46
41	1.50	.13	.09	38.30
42	.87	1.35	.17	22.82
43	1.50	1.16	.22	24.72
44	1.33	.53	.17	8.82
45	1.15	-.91	.07	37.94
46	1.10	.23	.37	16.18
47	1.12	-1.27	.20	23.20
48	1.50	1.35	.12	28.31
49	1.40	1.21	.19	14.70
50	1.22	-.90	.22	7.92
51	.56	-.14	.20	18.76
52	1.41	.14	.19	30.39
53	.87	.01	.07	33.69
54	1.50	.13	.11	45.33
55	1.10	.42	.25	9.22
56	1.01	.56	.14	13.54
57	.65	-.05	.20	64.13
58	.66	-.39	.20	14.06