DOCUMENT RESUME

ED 054 659 FL 002 381

AUTHOR Berger, Louis S.; And Others

TITLE A Methodology to Achieve Secure Administration of English Comprehension Level Tests--Phase 1. Final Report.

INSTITUTION Southwest Research Inst., San Antonio, Tex.
SPONS AGENCY Defense Language Inst., Lackland AFB, Tex. English Language Branch.
PUB DATE 10 May 71
NOTE 59p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Achievement Tests; *Computer Programs; Computers; Digital Computers; *English (Second Language); Estimated Costs; Item Analysis; *Language Tests; Native Speakers; Psychometrics; Second Language Learning; *Tables (Data); *Test Construction

ABSTRACT

This report analyzes a two-step program designed to achieve security in the administration of the English Comprehension Level (ECL) test given by the Defense Language Institute. Since the ECL test score is the basis for major administrative and academic decisions, there is great motivation for performing well, and student test compromise is prevalent, especially on tests given in the students' own country. The best way to combat compromise is to have a large number of test forms. This report first presents an analysis of the estimated cost of test compromise. There is a discussion of how the problem was handled, and a formula for estimating the cost of compromise is given. The second part of the study describes the development of conceptual tools and computer programs to enable a digital computer to generate valid ECL test-item lists in quantity. Details and statistics are provided along with a discussion of the computer methodology. (VM)

A METHODOLOGY TO ACHIEVE SECURE ADMINISTRATION OF ENGLISH COMPREHENSION LEVEL TESTS - PHASE I

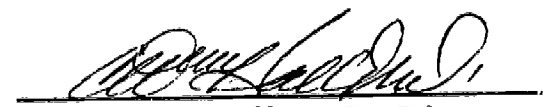
by
Louis S. Berger
Thomas R. Jackson
Thomas E. Hawkins

FINAL REPORT
SwRI Project 13-2825
Contract No. F 41609-70-C-0030

for
Defense Language Institute
English Language Branch
Lackland Air Force Base, Texas

10 May 1971

Approved:



C. William Hall, M.D., Director
Department of Bioengineering

FOREWORD

This is the Final Report of work performed at Southwest Research Institute, 8500 Culebra Road, San Antonio, Texas, under Contract No. F41609-70-C-0030, for Defense Language Institute, English Language Branch, Lackland Air Force Base, Texas. The contract period was 1 May 1970 to 11 May 1971.

The Technical Monitor was Francis A. Cartier, Ph.D., Chief, Development Division, English Language Branch; his supervision and technical advice was most helpful throughout the project. At Southwest Research Institute, the Behavioral Sciences Section, Department of Bioengineering, had responsibility for the project. The computer programs were developed in the Computer Laboratory by Mr. Thomas R. Jackson, manager of that facility, while the cost/benefits study was performed within the Operations Research Section, Department of Electronic Systems Research, by Messrs. Thomas E. Hawkins and Richard A. McCoy; the early conceptual development of the approach to the cost/benefits study benefitted significantly from the contributions of Dr. W. R. Brian Caruth, then manager of the Operations Research Section. Mr. Louis S. Berger was Principal Investigator and Project Manager.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

I. INTRODUCTION

To support the Military Assistance and Foreign Military Sales programs of the United States, selected foreign military personnel are sent to the United States for technical training in a wide variety of skills. Because knowledge and comprehension of English are essential to successful completion of this technical training, Defense Language Institute (DLI) has developed a comprehensive system for the training of foreign nationals in English, both "in-country" and in the United States at DLI English Language Branch, Lackland Air Force Base, San Antonio, Texas. Within this system, the English Comprehension Level test (ECL) is the basic tool for measuring a student's proficiency in the language; broadly speaking, a potential student's initial command of the language, as well as his academic progress during language training, are measured by ECL tests.

A number of major administrative and academic decisions are based on a student's ECL score. The initial evaluation, usually performed in the applicant's home country, determines whether or not a candidate is ready for training in the United States. Then, should he qualify, the following far-reaching programming and scheduling decisions are made on the basis of his ECL score: if he is scheduled for language training at English Language Branch, the ECL is used to place him within the curriculum, to measure his academic progress, to predict the duration of his English-language training, and to provide a criterion for his graduation from language training; in other instances, his ECL score may be high enough to enable him to bypass language training in the United States entirely. Thus, considerable resource decisions are made on the basis of ECL test score results.

Viewed from the student's vantage point, the motivation for performing well on an ECL test, be it a screening admissions test overseas or an academic evaluation test at English Language Branch, is high because of socioeconomic and other personal rewards which derive from satisfactory language performance as measured by the ECL test. Consequently, student test compromise is prevalent, particularly in screening tests in a student's home country. The two major avenues to compromise are through previous knowledge of test questions, and through exchanging information during test administration. Both of these compromise techniques are effectively countered by the use of a large number of test forms. Unfortunately, because of the rather complex categorial and statistical specifications, assembling an ECL test form is an intricate procedure, and the time required for assembling alternate ECL forms by hand has, in the past, limited the number of operational ECL forms available to DLI.

The program described in this report was designed to meet the need for ready availability of a large number of valid ECL forms; we proposed to develop a computer methodology that would construct a very large number of ECL form lists to specifications from a basic test item pool provided by the sponsor. Two tasks were proposed: Task I would develop the desired ECL test generation methodology; Task II would determine the magnitude of the economic penalties for compromise of the ECL tests, and thereby provide a realistic basis for evaluating further applications of computer technology to test generation.

The goal of Phase I, Task I, was to develop a computer methodology which would assemble appropriate sets of test items (120 items each), from which the Sponsor could prepare ECL test forms. The methodology to be developed under Task I should select ECL test item sets which would perform at least as well as the sets constituting the operational forms now in use at DLI, would be compatible with standard computer systems, and would ultimately be expandable in a straightforward way to wider applications in other DLI programs.

The methodology that was developed in the course of this project was thoroughly evaluated in two validation studies at English Language Branch. These studies led the Sponsor to conclude that the methodology is generating valid ECL test item lists and that computer-selected test forms can be put into operational use at English Language Branch. We feel that two important conditions existed at English Language Branch which contributed vitally to the success of Task I: first, a proven set of test items was at hand, together with a well-defined operational taxonomy; second, the insight of the staff into test construction science enabled the staff to provide valuable guidance to us during development of the computer methodology.

The Task II effort to estimate the economic penalties of compromise was based on a very conservative analytical approach. Assumptions were conservative, and where the complexities of a situation or the availability of data precluded sound analysis, penalty estimates were not incorporated in the computations. The basic data were acquired largely from interaction with English Language Branch faculty, supervisory and management staff, supplemented by available source records. Despite the conservative approach, the estimated economic penalties proved to be quite large. The calculations indicated that the cost of compromise was approximately $76,000 for every 1000 students entering English Language Branch, and approximately $117,000 for every 1000 direct entry students (those bypassing language training in the United States). Thus, the annual costs of ECL test compromise for the period analyzed were very conservatively estimated to be on the order of 1.25 million dollars.

In the main body of this report, the Task II effort is reported first (Section II) in order to provide a general framework within which the Task I technical discussion (Section III) is then presented.

II. TASK II EFFORT–COST/BENEFITS ANALYSIS

A. Introduction

1. Purpose of Task II Research

The purpose of the research under Task II was to determine the magnitude of the economic penalty of in-country compromise of the ECL tests, and thereby provide a realistic basis for evaluating the cost effectiveness of wider applications of computer technology to ECL test generation. The research had to rely on a number of approximations, since the budget for Task II was one-fourth of the total project budget. The task was completed within the limitations of this modest budget.

2. Approach to Task II Research

There are three major subsystems within the overall system of training foreign nationals: (1) an English Language Instructor Program which trains foreign nationals at Lackland Air Force Base for return to their own country as instructors in the in-country English Language Training Program; (2) the in-country English Language Training Program, followed by direct entry into the technical training; and (3) the in-country English Language Training Program followed by intermediate training in General and/or Specialized English at Lackland Air Force Base prior to graduation to technical training. Two of these subsystems were considered in this research effort. The English Language Instructor Program was considered by the senior personnel of English Language Branch (P1)* to be relatively free of ECL test compromise, and its consideration was excluded from the analysis of the magnitude of the economic penalty from compromise. Both of the other major subsystems (intermediate training and direct entry) were considered in the analysis.

There were two other factors which were excluded from the analysis (P1). First, it is accepted by the senior personnel of English Language Branch that there occurs some testing compromise while students undergo intermediate English language training at Lackland Air Force Base, but it was considered to be minimal and under effective control. Therefore, the research of Task II was limited to an analysis of the economic penalty stemming from test compromise of ECL's given in-country. Second, if cheating on ECL tests given in-country were reduced and/or eliminated, one anticipated result would be to increase the scope and cost of the in-country training programs. However, as the cost of this training is borne by the host countries, except for certain countries in Southeast Asia, it was considered by the senior personnel of English Language Branch, Defense Language Institute, to be outside the scope of this research effort.

It became clear early in the research effort that there was no direct way to measure the economic penalty of ECL test compromise, that there was a paucity of historical source data concerning the direct entry program, and that there was a large amount of historical source data concerning the intermediate training program conducted at Lackland Air Force Base. Based upon these early findings, three guidelines were developed at SwRI and approved by the senior personnel of English Language Branch for the research effort (P1). First, because of the necessity for indirect measurement of economic penalties, the techniques used for the definition of test compromise and for the translation of the penalties into economic

*Codes in parentheses refer to Contact Lists and Bibliography.

terms should be conservative. Second, because of the availability of source data, emphasis should be given to the analysis of the intermediate English Language Training Program conducted by English Language Branch at Lackland Air Force Base. And, third, sampling techniques should be used for the analysis of large numbers of historical records. These three guidelines were followed in the research of Task II.

The results of the analysis of the intermediate English Language Training Program are contained in Part B, while Part C contains the results of the analysis of the direct entry program. Part D contains a summary of the magnitude of the economic penalties being paid by the U.S. Government and several of its agencies for in-country compromise of the ECL tests. Parts E and F contain a list of personal interviews and telephone conversations, and a listing of source documents, respectively.

B. Analysis of the Cost of In-Country ECL Test Compromise for the Intermediate English Language Training Program

1. General

There are four training programs presently being conducted by English Language Branch at Lackland Air Force Base: (1) General English Training; (2) Specialized English Training; (3) General and Specialized English Training; and (4) I & M Training specifically designed for personnel from Viet Nam. Based upon personal interviews with the Section Chiefs responsible for each of these programs (P2, P3, P4), it was concluded that the training programs were similar enough to permit the development of a generalized scheme for the assessment of the economic penalty stemming from in-country ECL test compromise. Each of the training programs must accommodate itself to the fact that many students arriving in the United States score significantly lower on their entry ECL tests than they did on their final ECL test in-country. A certain amount of this drop in scores can be attributed to the time factor between tests, but experience has shown that these time-factor related drops are generally compensated for by a rapid rise in ECL scores once a student resumes course work at ELB. Therefore, significant drops without a rapid increase can logically be attributed to in-country test compromise.

Compromise results in significant reprogramming of course duration at Lackland, remedial efforts to attempt to graduate the students on schedule, and administrative burdens of considerable magnitude. It also was determined that for each program, the decision-making process concerning training duration and graduation is controlled externally to English Language Branch, and consideration is given to factors other than final ECL test scores meeting prescribed criteria. In other words, the training program at Lackland is conducted in a flexible, rather than a rigid, environment.

Based upon these considerations, it was concluded that the generalized scheme for the assessment of the economic penalty must include criteria for judging who did and who did not cheat, algorithms for translating the operational penalties of additional course time and remedial and administrative burdens into monetary terms, and should take into account the fact that ECL test scores were not the only factors considered in the decision-making process. It also was concluded that the generalized scheme must, of necessity, be designed upon the availability of historical source data.

2. Data Availability

There were available three sources of historical data: (1) "Quarterly Training Statistics" for fiscal years 1967, 1968, 1969, and 1970; (2) "Student Performance Records" for calendar

ears 1969 and 1970; and (3) an "I & M Summary" for six training classes (Groups 12 to 17) of Vietnamese students.

a. *Quarterly Training Statistics.* The Quarterly Training Statistics (D3) provide information on the number of students by country and sponsoring service, average course length, number of failures, extensions in course length, and reductions in course length. Preliminary review of this information indicated that it was summarized for a different purpose and in a manner that made it of limited use to the research effort; it did provide the only available data on the total number of students passing through the system and was therefore valuable in providing a data base for extrapolation of the analysis of selected samples.

b. *Student Performance Records.* The Student Performance Records (D4) provide comprehensive data on each student passing through the system. Specifically, for each student, there is recorded data on country of origin, date of entry, date of graduation, scheduled training period, actual training period program, in-country ECL test score, entry ECL test score, biweekly ECL test scores, final ECL test score, required ECL test score, remedial action, and disposition. Additionally, the original orders, which are filed with the Student Performance Record, provide data on the sponsoring service and type of contract (sales or grant). Preliminary review of these forms indicated that, in addition to providing a basis for country and date of entry matrixes, they could be used to develop the extent of reprogramming of training (either reductions or extensions), percent graduated/failed, percent meeting required ECL at graduation, and percent within x points of in-country ECL at 2-week intervals. Based upon this preliminary review, it was concluded that the Student Performance Records were an excellent and the best available source of historical data.

c. *I & M Summary.* This summary, which was a computer printout, provided data on in-country, entry, entry plus 1 week, and entry plus 2 weeks ECL test scores for the Vietnamese students in six groups. It also contained historical data on the ECL test forms which had been used for the in-country test. Preliminary review of this summary indicated that it would be extremely valuable in the development of criteria for in-country cheating. Fortunately, the six groups used both old and new forms of the ECL test, and the time span extended to periods before and after there were changes made in Viet Nam of the administration of ECL tests. Several personnel of English Language Branch had indicated that the periods before and after October 1969 should give an indication of the extent of the in-country cheating (P2, P3). On the basis of this review, it was concluded that there could be developed criteria to decide who had and who had not cheated on the in-country ECL. The rationale for the development of these criteria is described in the following section.

3. Compromise Criteria

In order to establish a realistic and conservative criterion for who did and who did not cheat on the in-country ECL tests, consideration was given both to the views of selected personnel at English Language Branch, Lackland Air Force Base, as well as to the analysis of the available data from the I & M students from Viet Nam.

All three Section Chiefs (P2, P3, P4) at English Language Branch, Lackland Air Force Base, indicated that there is a time delay between the final in-country ECL test (which determines whether or not a particular student is qualified to move from the in-country training program to the intermediate language training program at Lackland) and the entry ECL test. This time may be as great as several months and typically is accompanied by a decrease in ECL test scores at entry, but that decrease is narrowed within 2 weeks by students who had not cheated on their in-country ECL

test. One Section Chief (P4) stated that he was convinced that the following would be appropriate assumptions for a definition of cheating:

(1) If the ECL test score 2 weeks after entry into the intermediate program was more than 10 points below the in-country ECL, there was a 100-percent probability that cheating had occurred on the in-country ECL;

(2) If the ECL test score 2 weeks after entry into the intermediate program was more than 7 points, but less than 10 points, below the in-country ECL, there was some probability that cheating had occurred on the in-country ECL; and

(3) If the ECL test score 2 weeks after entry into the intermediate program was less than 7 points below the in-country ECL, there was a zero probability that cheating had occurred on the in-country ECL.

There were available within the data concerning the I & M Groups (D2) the records of forty students who were given final in-country ECL tests on versions of the ECL tests so new as to make compromise difficult. There were also available the records of several hundred students who had taken their final in-country ECL tests on versions of the ECL test which had been in use for about a year. It was hypothesized that there would be a significant difference between these two groups of students. To prove or disprove this hypothesis, there was made a comparison of the forty students from Groups 15-16-17 who had used the new versions of the ECL test form and forty students selected at random from Groups 12-13-14 who had used old versions of the ECL test. The results of this comparison are shown graphically in Figure 1.

On the basis of this analysis, a criterion for classifying students into cheaters or non-cheater groups was developed. The criterion would make use of two measures of a student's ECL performance history: the difference between his in-country and entry ECL score and the difference between his in-country and entry-plus-2 weeks score. A student would, for the purpose of our study, be classified as a cheater if (1) the difference between a student's in-country ECL score and his entry score was greater than 15, and also (2) the difference between a student's in-country ECL score and his entry plus 2 weeks ECL score had been greater than 10. Unless both these conditions were satisfied, a student would be classified as a non-cheater. This criterion was considered conservative and was recommended by the research team and approved for use in the analysis of the economic penalties associated with English Language Branch Intermediate training program (P7).

4. Penalty Measures

Logically, it appeared that the penalties of compromise were:

(1) Ultimate failure or poor performance in subsequent technical training;

(2) Failure at English Language Branch;

(3) Additional training at English Language Branch;

(4) Required remedial help while at English Language Branch; and

(5) Administrative burden, including the penalties of missed quotas and deviations from sequential schedules.

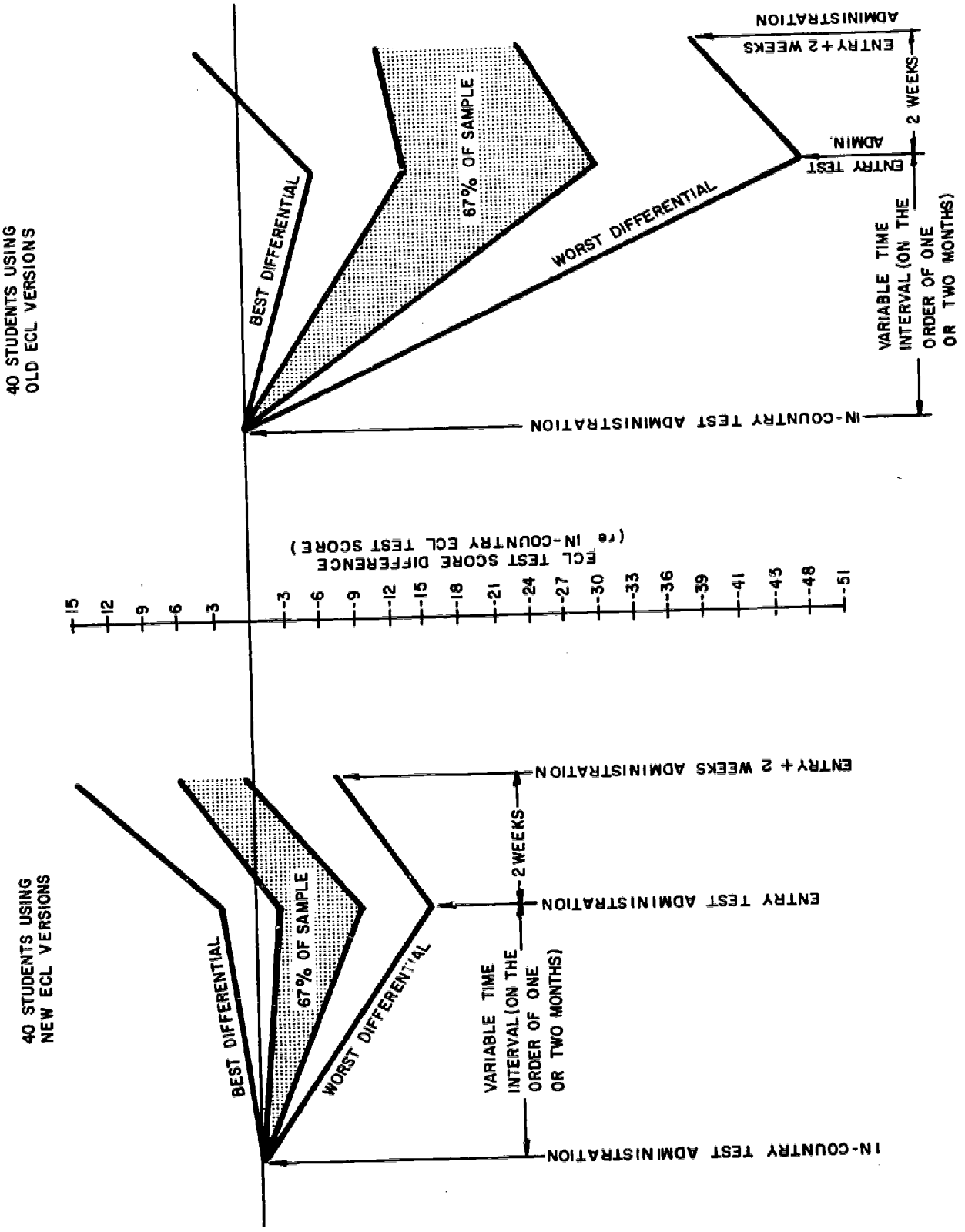


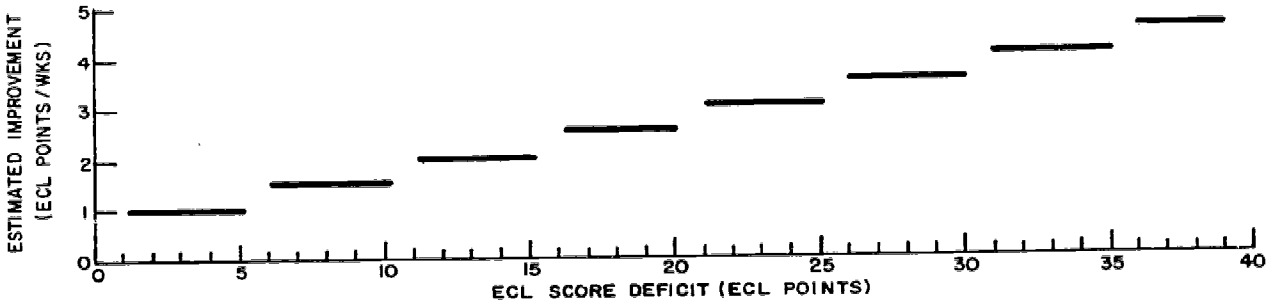
FIGURE 1. HYPOTHESIS TESTING OF COMPROMISE CRITERIA

Of these, the first three could be calculated, but the last two factors could not be developed adequately for use in the study. Unquestionably, there are "administrative" and "remedial" burdens for poor performers, but no effective technique (short of a management audit) could be developed within the limited scope of this task which would reveal the magnitude of these burdens. Therefore, these last two penalties were not included in the analysis.

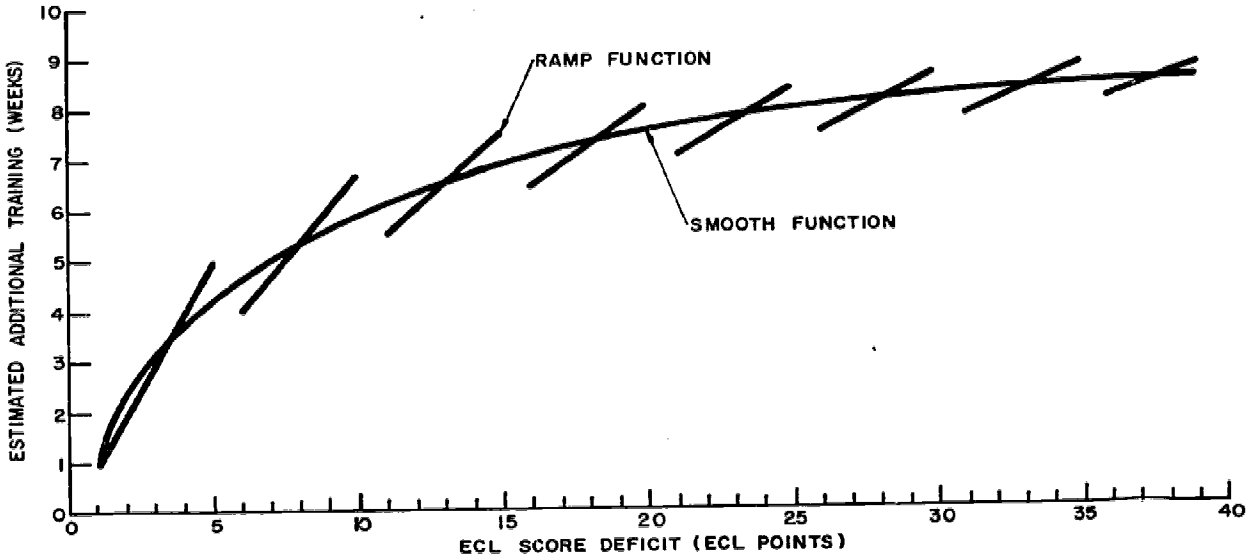
"Failure at English Language Branch" and "Additional Reprogramming" were factors available from the Student Performance Record and readily converted to economic terms. For those students attending English Language Branch on a grant basis, the direct costs borne by the United States are $60 per man-week tuition; $9 per man-day per diem ($63 per week), and $3 to $4 per man-day subsistence ($21 to $28 per week). On the basis of these direct costs, which were provided by the Comptroller of English Language Branch, it was concluded that a conservative estimate of the cost of reprogrammed training was $150 per man-week. This figure admittedly ignores many indirect overhead costs borne by the United States and the pay of students borne by the host country, but it does meet the criterion of conservativeness. In the case of a student who has cheated to gain entry to English Language Branch and fails and is sent home, the entire cost of training plus roundtrip transportation is the direct penalty and is capable of calculation. The training period penalty is $150 per man-week, and the travel costs are documented in a current Air Force directive (D9). Again, many costs are ignored by this approach, but it is conservative.

The penalty for poor performance or failure at subsequent technical training proved impossible to measure directly (P6). However, because of its potential magnitude, it was believed highly desirable to develop relatively indirect techniques for the measurement of this penalty. A brief analysis of 10 cheaters and 10 non-cheaters selected at random indicated that there was a substantial difference between cheaters and non-cheaters; cheaters tended to graduate from the training at Lackland with final ECL's less than the required ECL to a greater degree than did non-cheaters. Based upon this finding, it was concluded that one indirect measurement of the penalty during technical training could be made by calculating the amount of additional training which would have been required to elevate his ECL from the final to the required, but which was not done because external factors dictated the graduation of the student. This technique admittedly is most indirect, but it does give some measure of the penalty during technical training and, if it errs, it is on the conservative side. It was approved for use in the subsequent analysis (P7). Development of the amount of training required was based on estimates of student training time versus achievement rate as a function of present ECL (P2). Figure 2 shows a smoothed curve representing the averaged composite ramp function which relates ECL test score deficit to additional estimated training time required to make up the deficit. Thus, for a stipulated increase in ECL test score, the amount of required incremental training can be obtained and converted into an economic penalty by applying the average $150 per man-week training cost.

It was necessary to include one additional factor because of the flexible environment of foreign national training. Some students fail who have not cheated; others require greater periods of training than originally scheduled, while some require less. In other words, there is a deviation from the normal path even if there has been no cheating. This was termed system variation and provisions were made for its inclusion in the analytical logic. An estimate of this variation was obtained from calculations of the above-mentioned three penalties (failure at English Language Branch, reprogramming, and graduation at less than required ECL) for non-cheater groups, since, by assumption, the penalties for these groups were caused by system variation only.



ESTIMATED IMPROVEMENT IN ECL SCORE



ADDITIONAL TRAINING REQUIRED

FIGURE 2. INCREMENTAL TRAINING DATA

5. Analytic Logic and Approach

For the purposes of analysis, it was found desirable to develop the penalty algorithms into a computer program which solved the penalty expression:

$$P_{cg} = P_f + P_{rt} + P_\Delta - P_{sv}$$

where

P_{cg} = \$ Penalty for the compromise group

P_f = \$ Penalty for failure at English Language Branch

P_{rt} = \$ Penalty for reprogrammed training

P_Δ = \$ Penalty for graduation at less than required ECL

P_{sv} = \$ Penalty for system variation

The input data source, input data format, computer program flow diagram, computer program listing, and computer program variable array keys were made available to the Technical Monitor.

It was not considered necessary to analyze the entire population of enrollees at English Language Branch; yet, in the interests of developing a defensible analysis, a large sample size was considered necessary. The total population enrolling at English Language Branch has been: FY67-2102; FY68-2534; FY69-2533; FY70-3823. It was decided (P7) that 50 percent of the students from the twelve countries sending the largest number of students during calendar year 1969 would be an adequate sample for extrapolation. This resulted in a sample size of 1032, split as evenly as source data permitted into equal samples for each of four quarters. The sampling had to be modified in the case of those countries for which there were not available historical records for the last quarters of the calendar year. In these instances, the third and fourth quarters of 1968 were substituted. In all cases, the data were chosen to run consecutively for four quarters for each country. The sample used in the computer analysis is shown in Table I.

6. Results of the Analysis of the Sample Population

The analysis of the sample population data was conducted on an individual basis for each of the twelve countries comprising the sample.

The output provides data on two groups of students for each country—those students, according to the compromise criteria, who did and who did not cheat in their in-country ECL test. Each of these groups was further subdivided into students who failed their English Language Branch training, those who graduated with less than the required ECL, and those who graduated at or

TABLE I. POPULATION SAMPLE USED FOR ENGLISH LANGUAGE BRANCH COMPROMISE ANALYSIS

	Third Quarter 1968	Fourth Quarter 1968	First Quarter 1969	Second Quarter 1969	Third Quarter 1969	Fourth Quarter 1969	Total
Germany		10	12	12	2		36
Iran		60	55	60	60		235
Israel			30	32	28		90
Korea		35	35	35	16		121
Laos	12	12	12	12			48
Libya	13	12	11*	12			48
Morocco			15	15	15		45
Saudi Arabia		8	12	7	6		33
Spain			12	15	12		39
Thailand		16	15	16	13		60
Turkey		9	11	7	12		39
Viet Nam		20	58	60	60	40	238
							1032

*One student had training interrupted.

TABLE II. DATA FOR PENALTY CALCULATIONS

	Compromise Subgroup	Non-Compromise Subgroup
Subgroup Size	14.15%	85.76%
Failed	1.37%	0.68%
Graduated less than Required ECL	62.33%	26.55%
Graduated at/above Required ECL	36.30%	72.77%
Penalty/Student	$766.83	$232.22
Cost of Compromise/Student Cheating	$ 534.61	
Cost of Compromise for Sample	$77,792.00	

greater than the required ECL, for each of the four quarters considered in the analysis. Finally, a composite tabulation was prepared showing the total results for all countries. The complete computer printout was transmitted to the Technical Monitor.

Highlights from the computer summary tabulation are shown in Table II.

The relative performance of the cheaters versus non-cheaters indicates the detrimental effect which compromise of the in-country ECL has on the efficiency of the instruction program at English Language Branch.

In addition to this analysis, the time variability of compromise was investigated. It had been hypothesized that there would be a consistent trend, but this was not evident from the analysis, the results of which are shown in Table III.

An additional analysis was made on a monthly basis for the students from Viet Nam, this being the only high-proportion cheating group for which there existed a sufficient number of records. The computer result of this analysis was transmitted to the Technical Monitor; this additional analysis also failed to show a consistent time trend.

TABLE III. TIME TREND COMPROMISE DATA

	Number Compromise	Number Non-Compromise	Total Number	Percent Compromise
Third Quarter–1968	4	21	25	16.0
Fourth Quarter–1968	31	151	182	17.0
First Quarter–1969	31	246	277	11.2
Second Quarter–1969	30	253	283	10.6
Third Quarter–1969	28	196	224	12.5
Fourth Quarter–1969	22	18	40	55.0
Totals	146	885	1031	
(Student training interrupted)			+1	
Total Sample Size			1032	

C. Analysis of the Cost of In-Country ECL Test Compromise for the Direct-Entry Program

1. General

It was planned at the outset of this research that an analysis similar to that performed for English Language Branch students would be accomplished for students entering technical training directly from their native country. Results of the direct-entry analysis could be compared with those from English Language Branch analysis. Given data similar to that employed for English Language Branch analysis, the same criterion could be applied as a test for compromise and the same penalties could be applied for student ECL's which did not meet required ECL's. Students who failed technical training and/or were rescheduled for additional English language schooling could be assigned monetary penalties analogous to those applied in similar English Language Branch cases.

However, it became apparent early in the research effort that not all of the input information necessary to accomplish such an analysis was available in a centralized location, if it were available at all. Through discussions with the Air Force Air Training Command (P8), it was learned that the Air Training Command maintains no centralized records on direct-entry students and, in fact, has no capability at any training command for English Comprehension Level testing. Further, these discussions revealed that the Army, which does have a capability for testing English Comprehension Levels, does not maintain a centralized file of records. In view of these findings, it was clear that the analysis of the direct-entry program could be severely limited by the availability of data.

2. Data Availability

Examination of the files of English Language Branch disclosed the existence of a summarized report (D1) on direct-entry students from fifteen Army Training Commands. Review of the report indicated that only a limited analysis could be accomplished from the summary report, although the source data had contained almost all of the data required for an analysis similar to that performed for English Language Branch training program. Unfortunately, this original source data had not been retained at English Language Branch, nor was it available at the parent commands of Defense Language Institute (P8, P9). Because of this lack of source data, two alternatives existed for the analysis of the penalty from in-country ECL test compromise for the direct-entry program:

(1) To collect the original records from the fifteen army commands as was done for the previous study, or

(2) To perform only a limited analysis of the available summary data resulting from the previous study.

Because the first alternative was beyond the scope of the research effort, the second alternative was chosen for the analysis of the direct-entry program (P10).

The summarized report provided the following data on 989 direct-entry students;

(1) The total number of students by country;

(2) The average in-country ECL scores;

(3) The average entry ECL scores;

(4) The average difference between in-country and required ECL scores;

(5) The average difference between entry and required ECL scores;

(6) The average difference between in-country and entry ECL scores;

(7) A tabulation of the number of students whose in-country ECL did not meet the required ECL and the number of students whose entry ECL did not meet the required ECL; and

(8) A tabulation of the number of students whose in-country ECL was at least 18 points greater than entry ECL and the number of students whose entry ECL was at least 18 points greater than their in-country ECL.

3. Compromise and Penality Criteria

Because the entry plus 2-week ECL test scores were not listed in the available data, the criterion for cheating used previously in English Language Branch analysis could not be applied to the direct-entry data. Based primarily upon data availability, it was recommended and approved (P10) that the new criterion for cheating would be a single-decision measure: a student would be classified as a cheater if he obtained an in-country ECL test score of 18 or more points than the entry ECL score. We note that, among English Language Branch candidates in the group using new test forms (see Fig. 1), and thus presumably representing non-cheaters, not one student had as high an in-country/entry score differential as 18 points. In this respect, the single point criterion used for direct entry students would appear to be more conservative than the original two-decision point criterion: The new criterion tends to underestimate the cheating costs because it uses an 18 ECL point score rather than the original 15 ECL point score in-country/entry differential. On the other hand, with respect to the first criterion, some overestimation of costs may occur with the new criterion. If a student had "honestly" qualified in-country, forgotten (or otherwise lost) enough English to score 18 points below his in-country at entry, and regained sufficient comprehension level to reduce his in-country/entry plus 2-weeks score differential to less than ten, he would have been classified as a non-cheater by the criterion applied to English Language Branch students, but would now be classified as a cheater under the direct-entry, single-decision point criterion.

Because the distribution of entry ECL scores of 18 or more points below the in-country ECL score was not available from the report summary, a fixed 18-point (minimum) penalty was assessed to all students in the "18-point or higher difference" category, in line with our conservative approach. All students classified as cheaters by the single point criterion were assigned penalties by the same calculation process used for English Language Branch analysis. Point deficits were converted to required weeks of additional English language training according to Figure 2, and these weeks of additional training were assigned a monetary value by multiplying by $150 per week per student. For those students considered to have compromised their in-country ECL, the term ΔECL was always taken equal to 18 points; therefore, for each student in this category, the resulting compromise penalty is equal to $1080. This approach does not compute the penalties associated with ECL point differentials in excess of 18. However, we understand that these excess points are those most quickly gained during English language instruction and this ameliorates the understatement of the penalty (P2, P3, P4).

4. Analytical Logic

Based upon the delta training time formula developed in Figure 2, the cost penalties are computed by the following expression:

$$P_{\Delta} = \frac{(S)(\$150)(\Delta ECL)}{\frac{(\Delta ECL + 2)}{10} + 0.5}$$

where

S = No. of cheaters

For students who compromised their in-country ECL, the ΔECL is by definition, as previously discussed, 18 points, and the penalty per student is $1080.

5. Results of the Analysis of the Sample Population

The results of the analysis are shown in Table IV.

TABLE IV. DIRECT-ENTRY PENALTY ANALYSIS

Country	Number of Students	Number Compromised	Percent Compromised	Total Compromise Cost Penalty	Average Compromise Cost Penalty
Argentina	8	1	12.5	$ 1,080	$135
Brazil	15	1	6.6	1,080	72
Colombia	11	3	27.3	3,240	294
Ethiopia	49	10	20.4	10,800	220
Greece	77	1	1.4	1,080	14
Guatemala	7	3	42.8	3,240	463
Iran	76	5	6.6	5,400	71
Italy	8	2	25.0	2,160	270
Jordan	31	3	9.7	3,240	104
Korea	114	6	5.3	6,480	57
Lebanon	17	1	5.9	1,080	64
Liberia	14	1	7.1	1,080	77
Thailand	157	26	16.6	28,080	178
Turkey	28	1	3.6	1,080	38
Viet Nam	240	41	18.6	44,280	201
Total	989	105	10.6	$113,400	$117

These findings, which were based upon the assumptions described in Section C3, are reasonably consistent with the finding of the English Language Branch analysis, as shown in Table V.

TABLE V. COMPARISON OF ESTIMATED COMPROMISE FOR DIRECT-ENTRY AND ENGLISH LANGUAGE BRANCH STUDENTS

	Percent Compromised	
	English Language Branch Analysis	Direct-Entry Analysis
Total	14.15	10.6
Iran	0.85	6.6
Korea	23.97	5.3
Laos	27.08	Zero
Libya	10.42	Zero
Thailand	23.33	16.6
Turkey	12.82	3.6
Viet Nam	27.31	18.6
$ Penalty Per Student Enrolled	$75.63	$117.00

D. Summary of Penalty Estimates

Extrapolation of the English Language Branch penalties to past enrollment over the last 4 years results in the listing of Table VI.

Extrapolation of the direct-entry analysis to the total program results in a cost penalty of $117,000 for every 1000 students. Based upon information available to the research team, there were approximately 8000 students in the direct-entry program, and this program level converts to a penalty of almost $1,000,000 per year.

TABLE VI. ENGLISH LANGUAGE BRANCH PENALTIES

	Fiscal Year 1967	Fiscal Year 1968	Fiscal Year 1969	Fiscal Year 1970	Four-Year Total
Number of Students Compromised*	294	355	355	535	1539
Compromise Cost†	$157,000	$189,000	$189,000	$286,000	$821,000

*Based on the average (14%) compromise in the in-country ECL test.
†Based on the average penalty cost ($534.61) per student classified as "compromiser."

E. Personal Interviews and Telephone Contacts

P1 Project Team Meeting with Senior Personnel of the English Language Branch, Defense Language Institute, Lackland Air Force Base, San Antonio, Texas, May 11, 1970.

P2 Project Team Meeting with Chief, General English Section, English Language Branch, Defense Language Institute, Lackland Air Force Base, San Antonio, Texas, May 13, 1970 and July 24, 1970.

P3 Project Team Meeting with Chief, Specialized English Section, English Language Branch, Defense Language Institute, Lackland Air Force Base, San Antonio, Texas, May 13, 1970.

P4 Project Team Meeting with Chief, I & M Section, English Language Branch, Defense Language Institute, Lackland Air Force Base, San Antonio, Texas, May 13, 1970.

P5 Project Team Meeting with Adjutant, English Language Branch, Defense Language Institute, Lackland Air Force Base, San Antonio, Texas, May 13, 1970.

P6 Project Team Telephone Conversation with Deputy Commander, Air Force Air Training Command, Randolph Air Force Base, San Antonio, Texas, May 15, 1970.

P7 Project Team Meeting with Chief, Tests and Measurements Branch, English Language Branch, Defense Language Institute, Lackland Air Force Base, San Antonio, Texas, July 24, 1970.

P8 Project Team Telephone Conversation with DCS/OPS, Defense Language Institute, Washington, D.C., July 24, 1970.

P9 Project Team Telephone Conversation with Headquarters, USCONARC, Fort Monroe, Virginia, July 29, 1970.

P10 Project Team Meeting with Chief, Development Division, Chief, Tests and Measurements Branch, English Language Branch, Defense Language Institute, Lackland Air Force Base, San Antonio, Texas, August 31, 1970.

F. Bibliography

Documents

D1 Defense Language Institute, English Language Branch, Letter to Director, Defense Language Institute, "Administration of English Language Proficiency Examinations," January 14, 1969.

D2 Defense Language Institute, English Language Branch, Memorandum to DLIEC-C, "Vietnam I & M Students (Groups 12-17)," April 27, 1970.

D3 Defense Language Institute, English Language Branch, Quarterly Report, "Quarterly Training Statistics," for Fiscal Years 1967, 1968, 1969, and 1970.

D4 Defense Language Institute, English Language Branch, "Student Performance Records," for Calendar Year 1969.

D5 Defense Language Institute, English Language Branch, "Course Chart for the Specialized Phase of American Language Course," March 17, 1970.

D6 Defense Language Institute, English Language Branch, Memorandum to Record, "Some Solutions to Problems of Test Compromise," no date.

D7 Defense Language Institute, Pamphlet No. 350-3, "Manual for Administration of English Comprehension Level Screening Test," January 1970.

D8 Defense Language Institute, English Language Branch, Memorandum to ELS-C, "ECL Test Statistical Procedures," August 15, 1968.

D9 Department of the Air Force, "MAP Instructions and Training Cost Factors Grant Aid," November 1, 1969.

III. TASK I EFFORT—COMPUTER-GENERATED TESTS (CGTs)

A. Overview

The overall objective of Task I was to develop the necessary conceptual tools and computer programs to enable a digital computer to generate valid ECL test item lists in quantity. To accomplish this objective, we acquired from English Language Branch an initial set of test items derived from 46 operational ECL test forms, organized these test and item data for computer acquisition, and stored the data. We next developed three auxiliary programs to analyze various aspects of the acquired data and, using these programs, investigated the characteristics of the stored items and DLIEL-ECL forms.

An ECL test generation computer methodology was defined. A program was written, and prototype ECL test item lists were generated by the computer and transmitted to the Sponsor. At English Language Branch, ECL test forms were typed and produced from the CGT lists, and evaluated by the Sponsor's usual validation methods. In addition, an "update" program was developed at SwRI which would, on demand, add, delete, update, or correct data pool test items and prepare a report documenting and analyzing the update operations. The major computer programs, along with other appropriate documentation and supplementary descriptive textual materials, were delivered to the Project Technical Monitor.

B. Data Acquisition

1. Categorization

English Language Branch uses a quite complex scheme for categorizing the test items and defining test specifications. After a thorough study of item categorization and test content specifications, we concluded that the specifications could be conveniently defined and dealt with in terms of four separate, independent partitions of the total item data pool (our universe set). We recall that a partition of a universe set is a division of that universe into mutally exclusive and collectively exhaustive subsets; that is, a partition divides the universe set in such a way that every item (member) of the universe belongs to one and only one subset of a partition.

Table VII shows the four basic partitions, the subsets in each partition, and the code symbols assigned to the

TABLE VII. ECL ITEM POOL PARTITIONS, CATEGORIES, AND CODES

Set Partition	Set Property Considered	Subcategories	Code	
#1	Modality of Presentation	Aural Comprehension	AC	
		Reading Comprehension	RC	
#2	Form of Presentation	Question	QU	
		Statement	ST	
		Dialogue	DG	
		Completion	CN	
		Underlined	UN	
#3	Lexical or Structural Subsets	Vocabulary	VO	
		Idiomatic Expression	ID	
		Comparatives, etc.	CO	
		Modals	MO	
		Prepositions	PR	
		Infinitives	IN	
		Gerunds	GE	
		Participles	PA	
		Verb Form	VF	
		Verb Tense	VT	
		Verb Passive	VP	
		Word Order	WO	
		Complex Sentence	CS	
#4	Source Reference (Book)	Elementary	11	
			12	
			13	1
			14	
		Unspecified	00	1
		Intermediate	21	
			22	
			23	
			24	2
		Other	25	2

subsets. The categorical properties of any given test item are completely specified by one combination of four descriptors, one and only one descriptor being chosen from each partition. There is only one unique set of descriptors which is correct for any one given item, because the categories describing an item are uniquely assigned to that item at English Language Branch.

We note that not all logically possible combinations of category codes are acceptable, valid test item descriptors. Some combinations are never used and no current test item is correctly described by such combinations. The subject of category combinations will be treated in greater detail in later sections of this report (III.C,D,F).

2. Transmittal Procedure from English Language Branch

English Language Branch staff prepared a typewritten listing of test items from available DLIEL-ECL forms. The listing contained the following information on each test item:

- Serial identification number assigned to that item
- A four-element (four-level) categorical code descriptor
- Objective
- Answer key
- Average value of the ease index and count (a digit showing the number of previous test administration sessions from which the index was derived)
- Average value of the discrimination index and count
- Transmittal date and
- Code for the DLIEL-ECL test form from which the item was obtained.

The reasons for including a "count" datum with the item index information and the use made of that count will be explained in the section which discusses item updating (Section III.F).

Before proceeding with computer processing of the acquired items, we checked the transmitted data visually to ensure that the serial identification number sequence was consistent, that the item category combinations (mentioned in the previous section) were valid, and that there were no missing elements in any of the test items. Any apparent item discrepancies were resolved through discussions with the English Language Branch staff. The verified item data were entered on punched cards at SwRI.

C. Data Pool Analysis

The initial data pool was assembled from 46 DLIEL-ECL forms containing 5111 test item questions. The kinds of data that were furnished to SwRI on each item, and on the ECL forms, have been listed in the preceding section. We proceeded to analyze these data in order to identify any unique or idiosyncratic characteristics related to ECL test reliability in particular. If we could learn more about the characteristics that accounted for the good performance of the DLIEL-ECL forms, then we could, by selectively duplicating these important characteristics hope to assemble computer-generated ECL test (CGT) forms of comparable quality of performance.

Because of the large size of the data pool, complexity of the analysis task, and availability of the item pool data in punched card format, it was economical and efficient to perform the analyses by computer. We used three analysis programs: a "scan" or "check item data" program, a statistical program, and a histogram plotting program.

The "scan" program, for each test,

(1) Identified and counted the out-of-range items (ease index greater than or equal to 0.94, or less than or equal to 0.37),

(2) Calculated the test mean EI and DI (ease and discrimination index),

(3) Checked for coding misprints in the category descriptors,

(4) Tabulated the distribution of answers, and

(5) Counted the category distributions.

Figure 3 shows a sample printout of this program.

The statistical program summarized test statistics for EI and DI. For each index, statistics were calculated for the total test, aural category, and reading category; the program also assembled tables of distributions for each statistical subanalysis and printed out an ordered array of the item index values. A sample printout is shown in Figure 4. The printout is shown for EI and the test as a whole; printouts for other breakdowns, for example, for DI/aural comprehension, are identical in format to the sample figure.

The histogram program printed histograms for the EI and DI distributions by test. Figure 5 shows sample printouts of this program.

The above three programs were initially applied to 46 DLIEL-ECL tests, as well as to the total item pool, treating all 5111 items as one large test. These programs were subsequently also useful in analyzing the characteristics of our CGT forms.

The analyses performed by these programs generated a sizable body of quantitative information. We will introduce these data, sometimes in summary form and at other times in detail, as needed in the technical discussion that follows. Particularly, the discussion of the next two sections will rely frequently on the information furnished by these data analyses.

D. CGT Methodology

1. Approaches

The major conceptual problems of the Task I effort concerned definition of an effective computer test item assembly methodology. While English Language Branch requirements imposed certain specific and, for the duration of the current project, firm constraints (to be discussed in the next section), there remained a significant amount of leeway in the conceptual development of the assembly program. Consequently, starting with the premise that the general selection methodology would be based on randomization, the computer program development dealt with defining and choosing among admissible conceptual alternatives.

```
SWRI 13 2825 04  D L I  CHECK ITEM DATA   06/01/70    15,56,14,
NO, OF ITEMS = 100  1ST SEQ NO, =  101     E0866R

          E X C E P T I O N S

 115     ,33   1 ,LE, ,37
 121     ,32   1 ,LE, ,37
 122     ,24   1 ,LE, ,37
 123     ,35   1 ,LE, ,37
 125     ,29   1 ,LE, ,37
 126     ,19   1 ,LE, ,37
 127     ,31   1 ,LE, ,37
 128     ,28   1 ,LE, ,37
 129     ,37   1 ,LE, ,37
 139     ,29   1 ,LE, ,37
 157     ,23   1 ,LE, ,37
 160     ,37   1 ,LE, ,37
 161     ,25   1 ,LE, ,37
 165     ,22   1 ,LE, ,37
 197     ,32   1 ,LE, ,37

E,I, MEAN =  ,56   D,I, MEAN =   ,23  NO, OF EXCEPTED ITEMS = 15 (   15)   GOOD =  85

 DISTRIBUTION OF ANSWERS  15 (A)   29 (B)   42 (C)   14 (D)
```

FIGURE 3a. SAMPLE PRINTOUT OF "CHECK DATA" PROGRAM–ITEM CHECK

SWRI 13 2825 04 D L I CHECK ITEM DATA 06/01/70 15.56.14.

DISTRIBUTION OF CATEGORIES E0866R

AC	QU	VO	1	1										
AC	ST	VO	1	4	AC	VO	1	5						
RC	CN	VO	1	7	RC	VO	1	7	VO	1	12			
AC	ST	VO	2	7										
AC	CN	VO	2	7	AC	VO	2	14						
RC	CN	VO	2	23										
RC	UN	VO	2	5	RC	VO	2	28	VO	2	42	VO	54	
RC	CN	ID	1	1	RC	ID	1	1	ID	1	1			
AC	QU	ID	2	2										
AC	ST	ID	2	9										
AC	CN	ID	2	1	AC	ID	2	12						
RC	CN	ID	2	6										
RC	UN	ID	2	2	RC	ID	2	8	ID	2	20	ID	21	
RC	CN	CO	1	7	RC	CO	1	7	CO	1	7	CO	7	
AC	ST	MO	1	1	AC	MO	1	1						
RC	CN	MO	1	2	RC	MO	1	2	MO	1	3	MO	3	
RC	CN	PR	1	1	RC	PR	1	1	PR	1	1	PR	1	
RC	CN	IN	1	1	RC	IN	1	1	IN	1	1	IN	1	
RC	CN	GE	1	1	RC	GE	1	1	GE	1	1	GE	1	
AC	CN	PA	1	1	AC	PA	1	1						
RC	CN	PA	1	1	RC	PA	1	1	PA	1	2	PA	2	
RC	CN	VT	1	3	RC	VT	1	3	VT	1	3	VT	3	
RC	CN	VP	1	3	RC	VP	1	3	VP	1	3	VP	3	
AC	QU	WO	1	1	AC	WO	1	1						
RC	CN	WO	1	2	RC	WO	1	2	WO	1	3	WO	3	
RC	ST	CS	1	1	RC	CS	1	1	CS	1	1	CS	1	

FIGURE 3b. SAMPLE PRINTOUT OF "CHECK DATA" PROGRAM–CATEGORY DISTRIBUTION

FOR TEST ECL 49 ALL 120 E.I.

SUMMARY STATISTICS

NUMBER OF VARIATES =	120
ARITHMETIC MEAN =	7.12666667E-01
STANDARD DEVIATION =	1.98602339E-01
VARIANCE =	3.94428889E-02
COEFF OF VAR (PCT) =	2.78670000E+01
STANDARD SKEWNESS =	-7.77000000E-01
STANDARD EXCESS =	-2.56000000E-01

ORDER STATISTICS

SMALLEST VARIATE =	2.10000000E-01
LOWER DECILE =	3.83000000E-01
FIRST QUARTILE =	5.95000000E-01
MEDIAN =	7.65000000E-01
THIRD QUARTILE =	8.57500000E-01
UPPER DECILE =	9.49000000E-01
LARGEST VARIATE =	9.90000000E-01
TOTAL RANGE =	7.80000000E-01
DECILE RANGE =	5.66000000E-01
SEMI-QUARTILE RANGE =	1.31250000E-01
BOWLEYS SKEWNESS =	-2.95000000E-01
PEARSON SKEWNESS =	-7.91000000E-01

FREQUENCY DISTRIBUTION

FROM	UP TO BUT NOT INCLUDING	FREQUENCY	PERCENT FREQUENCY
1.500E-01	2.000E-01	0	0.0
2.000E-01	2.500E-01	2	1.7
2.500E-01	3.000E-01	4	3.3
3.000E-01	3.500E-01	3	2.5
3.500E-01	4.000E-01	3	2.5
4.000E-01	4.500E-01	4	3.3
4.500E-01	5.000E-01	4	3.3
5.000E-01	5.500E-01	5	4.2
5.500E-01	6.000E-01	5	4.2
6.000E-01	6.500E-01	8	6.7
6.500E-01	7.000E-01	10	8.3
7.000E-01	7.500E-01	6	5.0
7.500E-01	8.000E-01	16	13.3
8.000E-01	8.500E-01	15	12.5
8.500E-01	9.000E-01	13	10.8
9.000E-01	9.500E-01	10	8.3
9.500E-01	1.000E+00	12	10.0
1.000E+00	1.100E+00	0	0.0

FIGURE 4a. SAMPLE PRINTOUT OF STATISTICAL PROGRAM

C U M U L A T I V E D I S T R I B U T I O N

VALUE	NUMBER LESS THAN VALUE	PERCENT LESS THAN VALUE	VARIATE SUM * PCT LESS THAN VALUE
2,000E-01	0	0,0	0,0
2,500E-01	2	1,7	,5
3,000E-01	6	5,0	1,8
3,500E-01	9	7,5	2,9
4,000E-01	12	10,0	4,2
4,500E-01	16	13,3	6,2
5,000E-01	20	16,7	8,4
5,500E-01	25	20,8	11,5
6,000E-01	30	25,0	14,9
6,500E-01	38	31,7	20,7
7,000E-01	48	40,0	28,6
7,500E-01	54	45,0	33,7
8,000E-01	70	58,3	48,1
8,500E-01	85	70,8	62,6
9,000E-01	98	81,7	75,7
9,500E-01	108	90,0	86,3
1,000E+00	120	100,0	100,0

FIGURE 4b. SAMPLE PRINTOUT OF STATISTICAL PROGRAM

ORDERED ARRAY

2.1000000E-01	6.1000000E-01	7.7000000E-01	8.6000000E-01
2.2000000E-01	6.1000000E-01	7.7000000E-01	8.6000000E-01
2.6000000E-01	6.2000000E-01	7.7000000E-01	8.7000000E-01
2.7000000E-01	6.3000000E-01	7.7000000E-01	8.7000000E-01
2.8000000E-01	6.3000000E-01	7.7000000E-01	8.8000000E-01
2.9000000E-01	6.4000000E-01	7.7000000E-01	8.8000000E-01
3.1000000E-01	6.4000000E-01	7.8000000E-01	8.8000000E-01
3.1000000E-01	6.4000000E-01	7.9000000E-01	8.8000000E-01
3.3000000E-01	6.6000000E-01	7.9000000E-01	9.0000000E-01
3.6000000E-01	6.7000000E-01	7.9000000E-01	9.0000000E-01
3.8000000E-01	6.7000000E-01	8.1000000E-01	9.1000000E-01
3.8000000E-01	6.7000000E-01	8.1000000E-01	9.2000000E-01
4.1000000E-01	6.7000000E-01	8.2000000E-01	9.2000000E-01
4.3000000E-01	6.7000000E-01	8.2000000E-01	9.2000000E-01
4.3000000E-01	6.8000000E-01	8.2000000E-01	9.3000000E-01
4.3000000E-01	6.8000000E-01	8.2000000E-01	9.3000000E-01
4.6000000E-01	6.8000000E-01	8.2000000E-01	9.4000000E-01
4.7000000E-01	6.9000000E-01	8.2000000E-01	9.4000000E-01
4.8000000E-01	7.0000000E-01	8.3000000E-01	9.5000000E-01
4.9000000E-01	7.0000000E-01	8.3000000E-01	9.5000000E-01
5.0000000E-01	7.3000000E-01	8.3000000E-01	9.5000000E-01
5.3000000E-01	7.3000000E-01	8.4000000E-01	9.5000000E-01
5.3000000E-01	7.3000000E-01	8.4000000E-01	9.5000000E-01
5.3000000E-01	7.4000000E-01	8.4000000E-01	9.6000000E-01
5.3000000E-01	7.5000000E-01	8.4000000E-01	9.6000000E-01
5.6000000E-01	7.5000000E-01	8.5000000E-01	9.7000000E-01
5.8000000E-01	7.6000000E-01	8.5000000E-01	9.8000000E-01
5.8000000E-01	7.6000000E-01	8.5000000E-01	9.8000000E-01
5.8000000E-01	7.6000000E-01	8.5000000E-01	9.9000000E-01
5.8000000E-01			
7.6000000E-01			
8.5000000E-01			
9.9000000E-01			

FIGURE 4c. SAMPLE PRINTOUT OF STATISTICAL PROGRAM

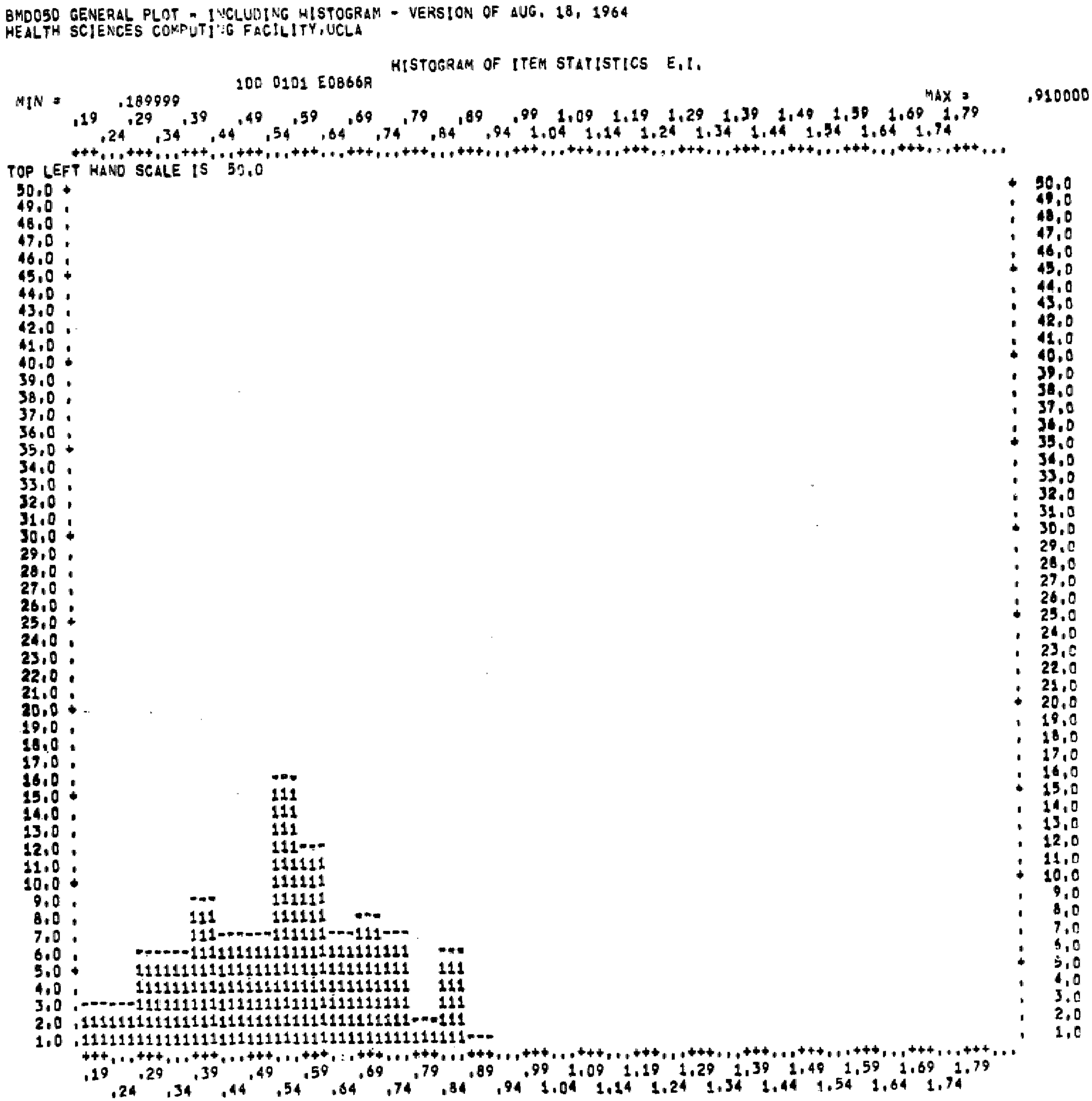


FIGURE 5a. SAMPLE PRINTOUT OF HISTOGRAM PROGRAM

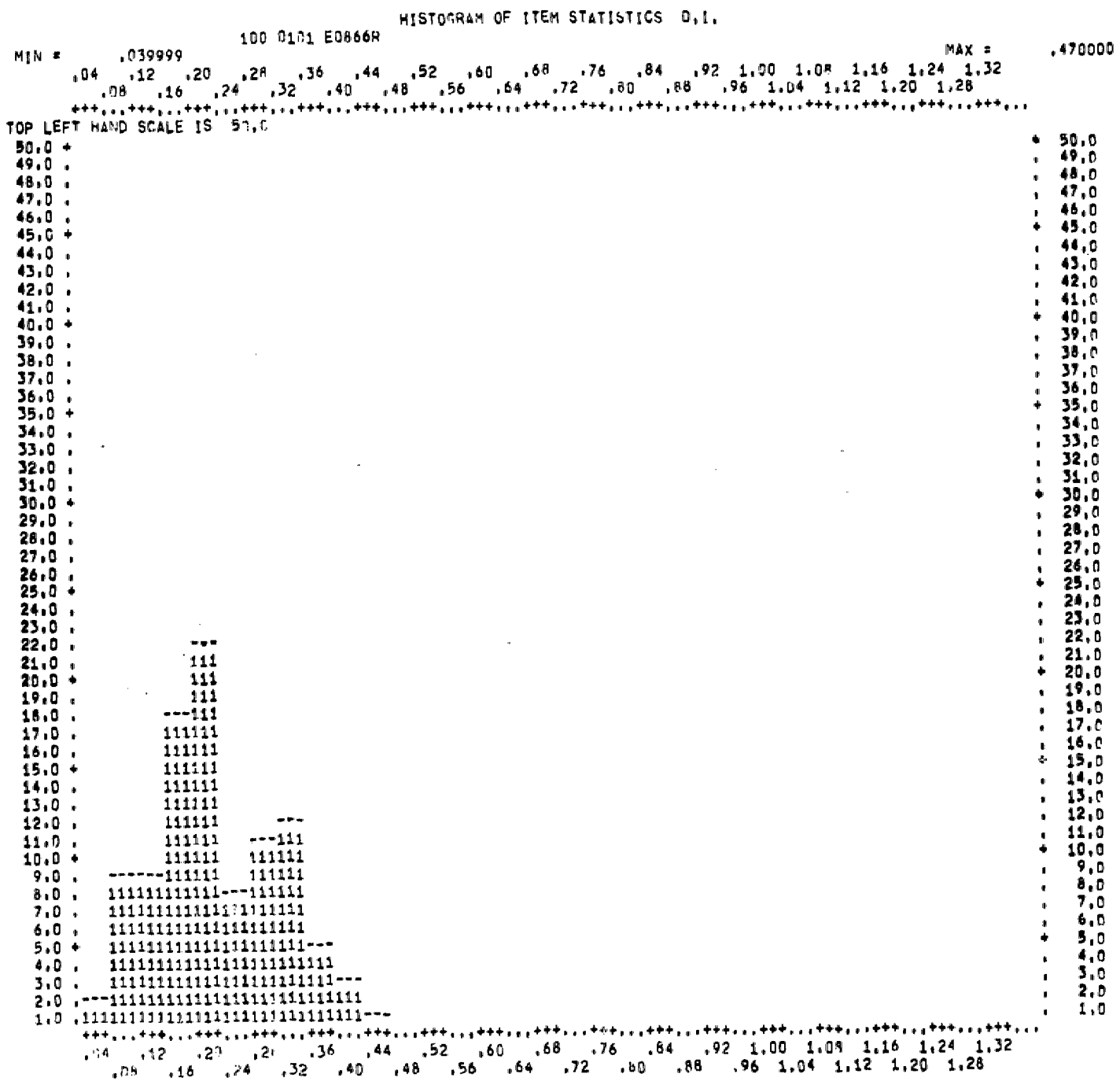


FIGURE 5b. SAMPLE PRINTOUT OF HISTOGRAM PROGRAM

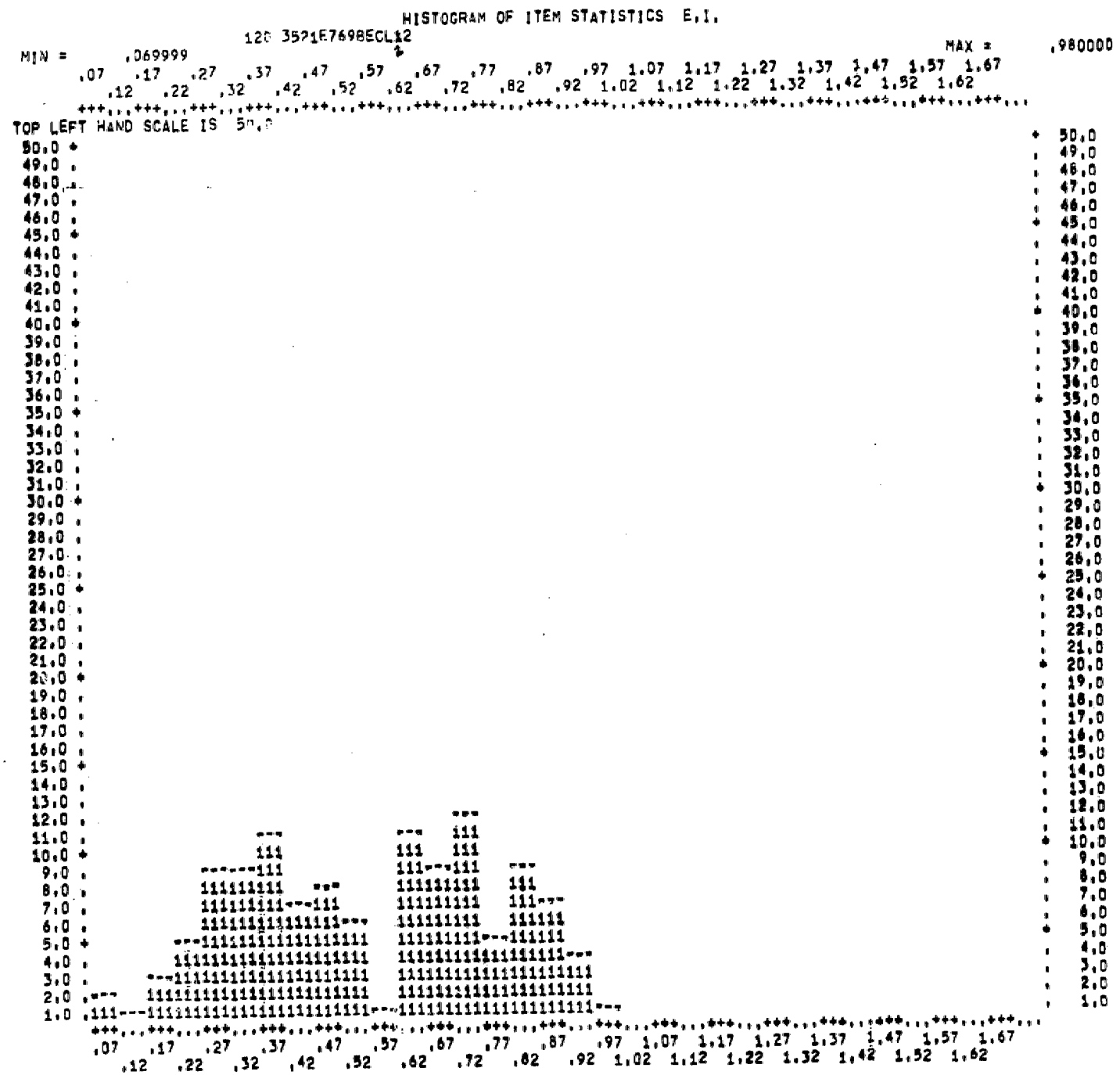


FIGURE 5c. SAMPLE PRINTOUT OF HISTOGRAM PROGRAM

```
BMD05D GENERAL PLOT - INCLUDING HISTOGRAM - VERSION OF AUG. 16, 1964
HEALTH SCIENCES COMPUTING FACILITY,UCLA

                              HISTOGRAM OF ITEM STATISTICS  E.I.
                  120 3401E-0769A
MIN =      .179999                                                             MAX =      .970000
```

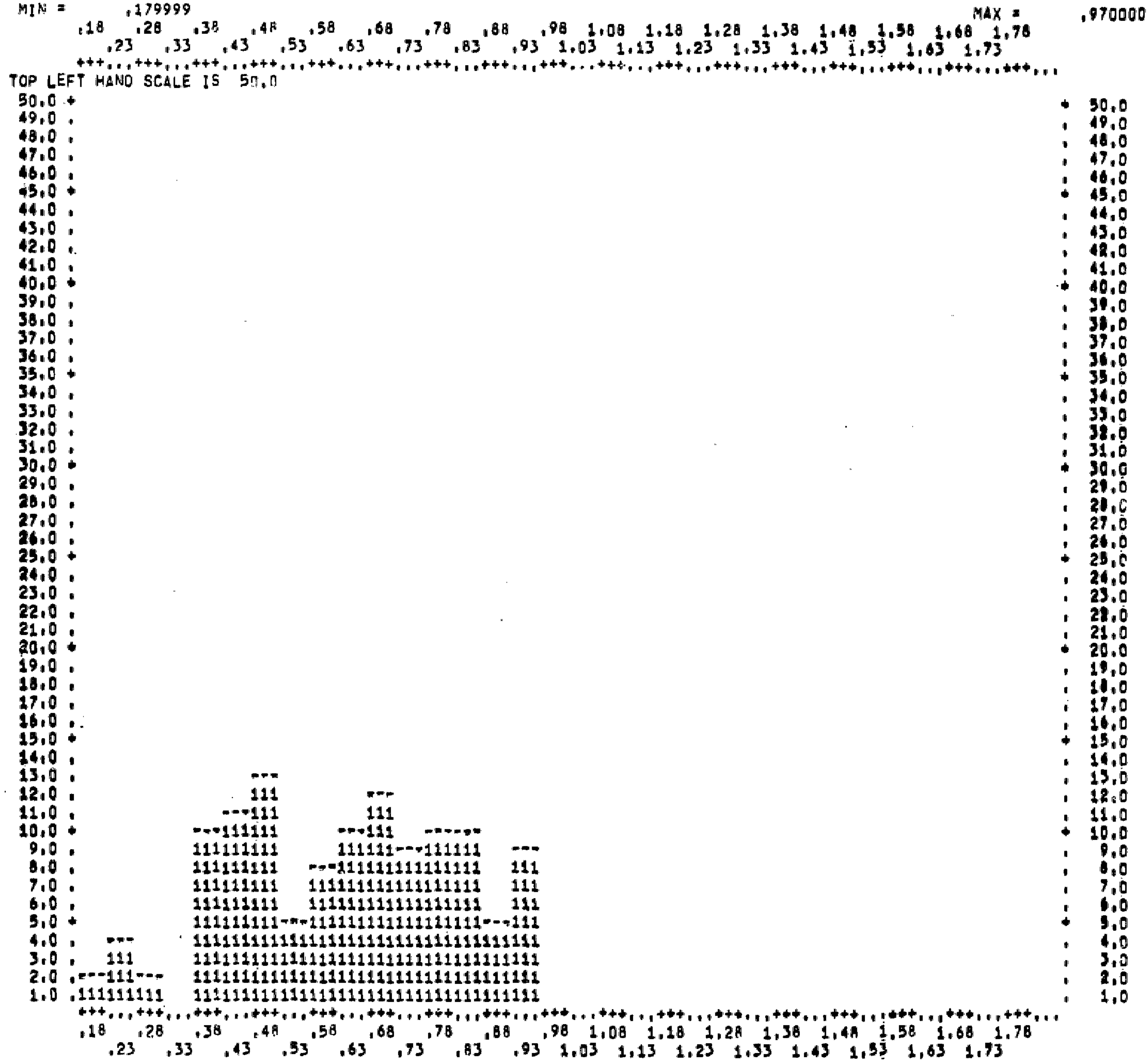


FIGURE 5d. SAMPLE PRINTOUT OF HISTOGRAM PROGRAM

33

Ideally, when evaluated according to present English Language Branch criteria, a CGT form should perform at least as well as the operational DLIEL-ECL forms now in use. Therefore, a vital question that had to be considered before choosing a CGT methodology was whether or not the ECL tests at English Language Branch might possess properties which, although unspecified and as yet unidentified, were contributing in important ways to the excellent performance characteristics of the DLIEL-ECL forms. In other words, were there any other identifiable ECL test form characteristics that would be vital for us to copy in order to obtain CGTs whose performance would be at least equivalent to the ECL forms' performance?

It was not possible to obtain conclusive answers to these questions from experimental investigations for several reasons: controlled studies of the effects on test performance of varying the test parameters were beyond the program scope; such studies would also have been difficult to implement at English Language Branch, since such investigations would place significant additional burdens on the staff, affect student schedules, and, by increasing exposure of operational test forms, make those forms more vulnerable to compromise. We therefore turned to a study of the characteristics of the 46 ECL forms which were the source of our computer data item pool, hoping that analysis and study of the test characteristics would provide helpful guidelines for achieving good CGTs. Using the computer analysis programs discussed in Section III.C, we obtained summary data on the performance history of 46 tests; tabulations were obtained for each of the 46 tests, showing the number of items in the test, the number of items with an ease index equal to or less than 0.37, the number of items with an ease index equal to or greater than 0.94 (at the time of this initial analysis, these constituted "acceptable" limits, but these limits were modified later, as will be discussed in Sections F and G), the total number of items falling outside the acceptable EI range, the mean ease index and discrimination index, and the type of ease index distribution (obtained from inspection of the computer-generated histogram). The test analysis data are summarized in Table VIII.

We had expected that the results of our DLIEL-ECL test analysis would guide us toward suitable computer methodologies for assembling ECL forms. The formal content specifications defined at English Language Branch would have to be satisfied at any rate, but we expected that the analyses would indicate the desirability of certain additional sampling or statistical constraints. Contrary to these expectations, our study of the DLIEL-ECL forms failed to identify any new critical test form characteristics; for all the investigated parameters (as summarized in Table VIII), there were significant variations between forms.

However, we knew that these forms perform very well in spite of their apparent dissimilarities. We also knew that, in any case, the moderate size of our item pool would not allow excessive constraints to be imposed on the assembly procedure; too severe constraints would make the generation of complete (120-item) CGT form lists difficult. We therefore chose, after several extensive discussions with English Language Branch staff, to generate tests with a constraint methodology based only on the current English Language Branch formal content specifications. Thus, our first set of prototype CGTs would be assembled without any additional conceptual constraints on the test generation methodology. We would rely on the statistical properties of the item pool and on quasi-random* sampling to achieve satisfactory test form parameter values and distributions. We decided to generate a set of CGT prototypes, submit them for a standard evaluation at English Language Branch, and, should the evaluation so dictate, to subsequently modify our initial approach.

*We use "quasi-random" to signify that at any one stage of test item assembly, all items in the data pool have an equal chance of being selected by the program; however, the item categories from which admissible item candidates may be acquired decrease as the assembly of a given form progresses. This point will be discussed in Section D3.

TABLE VIII. SUMMARY OF DLIEL ECL TEST FORMS

Test ID	No. of Items	Exceptions, EI		Total Exceptions		No. of "Good" Items	Mean EI	Mean DI	Content										Type EI Distribution
									Lexical		Struc-	Books		Lexical		AC			
		⩽0.37	⩾0.94						VO	ID	tural	VO1	VO2	RC	AC	QU	ST	DG	
ECL49	120	10	14	24	20%	96	0.713	0.14	78	15	27	30	48	22	71	24	39	8	1-Mode
ECL53	120	23	4	27	23%	93	0.587	0.21	75	19	26	24	51	19	75	30	30	15	Flat
ECL55	120	7	10	17	14%	103	0.698	0.20	78	16	26	33	45	19	75	24	33	18	Bimodal
E1069A	120	21	3	24	20%	96	0.602	0.21	75	15	30	41	34	15	75	30	30	15	Bimodal
ECL47	120	19	5	24	20%	96	0.604	0.21	76	18	26	38	38	19	75	31	29	15	Bimodal
E0866R	100	15	0	15	15%	85	0.556	0.23	54	21	25	12	42	44	31	3	20	0	1-Mode
E0267R	100	14	2	16	16%	84	0.647	0.15	48	21	31	24	24	26	43	11	30	0	Flat
E764FR	100	19	1	20	20%	80	0.543	0.19	59	10	31	–	59	15	54	15	33	1	1-Mode
E7067RR	100	7	0	7	7%	93	0.620	0.23	62	8	30	22	40	17	53	21	28	1	1-Mode
E0569A	120	14	13	27	23%	93	0.719	0.14	59	24	37	45	14	8	75	22	42	11	1-Mode
E0869B	120	11	8	19	16%	101	0.683	0.16	84	10	26	45	39	19	75	29	33	13	1-Mode
E268BR	100	2	2	4	4%	96	0.708	0.19	61	11	28	39	22	18	54	19	33	2	Bimodal
FF0466	100	15	3	20	20%	80	0.616	0.19	55	15	30	16	39	16	54	18	36	0	Flat
E667RECL14	100	12	2	15	15%	85	0.656	0.18	53	21	26	17	36	22	52	10	42	0	Flat
E769BECL22	120	30	5	35	29%	85	0.563	0.14	78	13	29	46	32	13	78	30	32	15	Bimodal
E669BECL10	120	13	9	22	18%	98	0.678	0.17	80	15	25	49	31	21	74	27	36	11	1-Mode
E0469A	120	10	10	21	18%	99	0.720	0.16	61	25	34	25	36	19	67	19	40	8	1-Mode
E0969A	120	17	7	25	21%	95	0.623	0.21	79	14	27	52	27	15	78	28	33	17	Flat
E0567R	100	7	0	7	7%	93	0.656	0.18	51	25	24	28	23	31	45	9	35	1	1-Mode
E0769A	120	8	6	15	13%	105	0.636	0.17	75	14	30	40	36	15	75	30	30	15	Flat
E768BECL6	120	10	5	15	13%	105	0.688	0.20	82	3	35	63	19	22	63	24	38	1	1-Mode
E868DECL7	100	15	0	15	15%	85	0.551	0.18	59	16	25	30	29	28	47	24	23	0	Bimodal
E468BECL27	120	5	3	8	7%	112	0.680	0.27	73	17	30	46	27	22	68	26	42	0	1-Mode
FF366ECL33	100	3	4	7	7%	93	0.681	0.18	54	12	34	16	38	29	37	16	20	1	Bimodal
E0169ECL17	120	12	16	28	23%	92	0.722	0.14	51	27	42	45	6	10	68	15	50	2	1-Mode
E0668CR	100	7	0	7	7%	93	0.610	0.23	58	18	24	38	20	17	59	22	36	1	1-Mode
E669ARECL18	100	4	0	4	4%	96	0.730	0.17	53	12	35	33	20	5	60	10	37	3	1-Mode
ECL57	120	14	3	17	14%	103	0.636	0.18	72	18	30	30	42	14	76	31	30	15	1-Mode
ECL61	120	22	10	34	28%	86	0.654	0.16	82	10	28	35	47	16	76	36	30	9	Bimodal
ECL60	120	27	5	32	27%	88	0.575	0.16	76	16	28	39	37	18	74	26	34	14	1-Mode
ECL40	120	3	1	4	3%	116	0.699	0.24	80	16	24	49	31	21	75	26	47	1	Bimodal
ECL42	100	3	0	3	3%	97	0.653	0.32	81	7	12	39	42	20	68	24	44	0	1-Mode
E0369	120	4	4	8	7%	112	0.699	0.24	85	4	31	55	30	20	69	25	38	6	Bimodal
ECL32	100	1	0	1	1%	99	0.643	0.23	66	11	23	30	36	18	59	24	35	0	Flat
ECL41	100	2	0	10	8%	90	0.716	0.10	67	12	21	53	14	17	62	20	42	0	1-Mode
ECL54	120	12	10	22	18%	98	0.673	0.13	84	9	27	33	51	19	74	30	29	15	Flat
ECL59	120	22	6	28	23%	92	0.595	0.15	78	14	28	39	39	13	79	31	35	13	Flat
ECL29	120	23	5	28	23%	92	0.608	0.15	82	15	23	54	28	21	76	31	30	15	Flat
ECL56	120	12	9	27	23%	93	0.666	0.18	86	8	26	38	48	18	76	17	45	14	Bimodal
E0968D	100	1	4	5	4%	95	0.693	0.21	60	18	22	37	23	19	59	25	33	1	Flat
ECL44	100	11	6	17	14%	83	0.655	0.19	67	12	21	34	33	23	56	18	38	0	Bimodal
E0569B	105	9	4	13	11%	92	0.654	0.23	62	14	29	36	26	18	58	8	47	3	Bimodal
E1068D	120	10	1	11	9%	109	0.657	0.23	73	20	27	46	27	30	63	23	40	0	Bimodal
ECL34E0565FR	84	13	0	13	11%	71	0.579	0.21	41	18	25	23	18	20	39	14	25	0	1-Mode
ECL30	100	0	0	0	0%	100	0.689	0.25	69	9	22	46	23	18	60	20	40	0	1-Mode
ECL36	99	16	7	23	19%	76	0.631	0.17	60	15	24	38	22	19	56	13	36	2	Flat

2. Constraints and Content Specifications

It was decided early in the program that a CGT must, at minimum, strictly meet English Language Branch content specifications for a 120-item test. The current content specifications are shown in Table IX.

In order to discover the effect that these specifications have on category sums, we compare Tables VII and IX. This comparison shows that the content specifications in some instances impose specific numerical sums requirements on subsets of a particular partition, while in other cases they impose numerical sums requirements on set intersections between subsets from different partitions. First, to explain the four partitions (previously referred to in Section III.B), we note, with reference to Table VII, that partition #1 identifies each item as belonging to either the AC or RC subcategories; partition #2 identifies each item as belonging to one of five subcategories

(QU, ST, DG, CN, UN); partition #3 identifies each item as belonging to one of 13 categories (VO, ID, CO, MO, PR, IN, GE, PA, VF, VT, VP, WO, CS); and partition #4, as used at English Language Branch at this time, identifies each item as being either from an elementary or intermediate book. The category membership of an item is completely specified by assigning the item four descriptors, one and only one descriptor from each partition, each descriptor showing to which subset of a partition the item belongs. The set properties of each test item are therefore fully specified by one and one only combination of four code symbols.

We can now interpret the specifications shown in Table IX in set terminology. The aural/reading requirement is a simple sums requirement on partition #1. The listening/(questions, statements, dialogues) requirement is a sums requirement on the intersection of one subset from partition #1 (AC) with three of the five subsets of partition #2 (QU, ST, DG). The vocabulary requirement is a simple sums requirement on a subset of partition #3 (VO). The vocabulary/(elementary, intermediate) requirement is a sums requirement on the intersection of a subset of partition #3 (VO) with the subsets of partition #4 (elementary, intermediate). The idioms requirements, as well as the requirements on the structural items, are simple sums requirements on the remaining subsets of partition #3.

TABLE IX. CONTENT SPECIFICATIONS FOR CGT (120-ITEM TEST)

Item Type	Model ECL Test
Listening	**75 Items**
Questions	30 Items
Statements	30 Items
Dialogues	15 Items
Reading	**45 Items**
Vocabulary	**72 Items**
Elementary	29 Items
Intermediate	43 Items
Idioms	**18 Items**
Structural Items	**30 Items**
Comparative	1 Item
Modal	4 Items
Preposition	3 Items
Infinitive	2 Items
Gerund	2 Items
Participial	2 Items
Verb Form	3 Items
Verb Tense	3 Items
Verb Passive	3 Items
Word Order	3 Items
Complicated Sentence	4 Items

We note in passing that there are redundancies in the content specifications of Table IX, since some of the subset requirements are sufficient to specify certain super-set sums. For example, since the form of each listening item (AC) is either a question, statement, or dialogue, the three subset sum specifications on AC/(QU, ST, or DG) serve to specify the sum total of the AC items. Other redundancies arise for analogous reasons. The redundancies can serve as a numerical check on the content specifications.

The number of distinct subset combinations logically obtainable from the four partitions is 260 (2 X 5 X 13 X 2); however, the English Language Branch content specifications define requirements for only 23 (and these can be restated in terms of only 18 sums, as will be later seen from Figure 6) of the 260 possible combinations. Furthermore, the logically possible number of 260 subcategory combinations cannot be fully used because certain subset combinations are not used at English Language Branch at this time; for example, since all currently used listening (aural) items are lexical (either vocabulary or idiom) items, any combination of item descriptors, including listening and structural (non-lexical) subcategories, would, at present, be considered an invalid category combination. There remain, after removal of the at present "forbidden" categories, 142 valid categorical subset descriptors. Since only 23 subset sums are firmly specified, this leaves a large number of subcategory sums unspecified for each ECL test form. We could have assigned specific sums to these subcategories, but, in line with the discussion of the preceding section, we chose to satisfy only the sums requirements defined at English Language Branch, allowing fluctuations in the other subcategory sums to occur as a consequence of the sampling procedure and item pool composition.

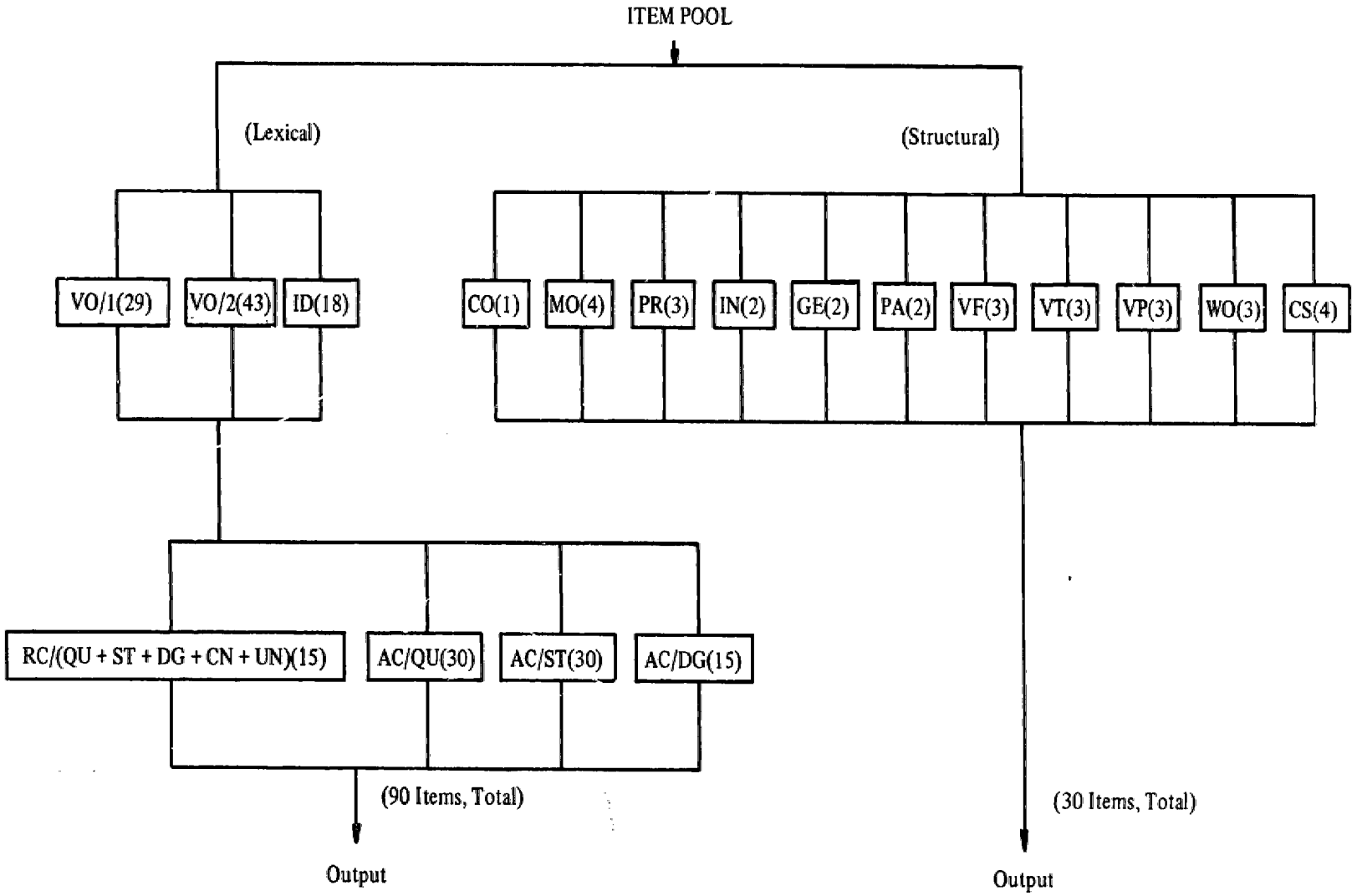


FIGURE 6. CGT COMPUTER PROGRAM NETWORK AND BRANCH TOTAL COUNTS

3. Generation of the First Set

The main function of the computer test assembly program is to assemble test form lists of 120 items each, each form being constructed to satisfy the content specifications described in the preceding section. Each set of 120 items is generated independently, in the sense that the computer program begins generation of each new form anew, "without memory" of its previous generation history. The content specifications are met by using a "pipe flow" structure consisting of a network of diverging and converging branches in which each branch has a predetermined total count setting which reflects the number of items of a given subset required to satisfy the content specifications. As the test form is assembled, a counter in each branch counts the items that have already passed through it. When the required total has been reached for that branch, the program rejects all items that would have normally been routed through the now-closed branch.

The program samples the item pool randomly, selects an item, and attempts to pass it through the pipe network. Each item's set of four category descriptors (Reference Table VII) uniquely specifies the item's path through the network. If all the pipes for that item are open, the item is accepted for the CGT form, and the program selects another item. On the other hand, if any branch of the network is closed to that item, or if the item had already been acquired for this form, the computer program rejects the item and samples the data base to acquire the next item. We see that the closing of a branch, by rejecting a certain set of candidate items, in effect restricts and reduced the item pool available for the remainder of the CGT form assembly. The program networks together with the prescribed branch total counts are shown in Figure 6.

By way of example, suppose we were sampling an "AC QU VO1" item and that 28 "VO1" items had been previously accepted in this test generation run. Then, assuming that the current "AC QU" subtotal was less than 30 and that the candidate item had not already been acquired for the partially assembled test, the item would be accepted. The "VO1" total would be increased to 29, closing that branch. (The AC QU total would also be raised by one.) Thereafter, any sampled item which had "VO1" as part of its descriptor would be rejected.

The computer program, in addition to assembling the desired category totals, keeps a count of the number of items rejected and terminates a CGT form assembly program when a predetermined number of rejections have occurred. This feature may be used to improve the test generation efficiency; it acts as a safeguard against anomalous conditions which could result in costly, unproductive, uncontrolled use of computer time.

Two formats of hard copy output are available. The first is a detailed listing, showing various steps and events in the test generation sequence for SwRI's study and evaluation of the program. The second arranges the 120 test items in the format in which it is transmitted to English Language Branch. In the latter format, the items are arranged into a group of 75 aural comprehension items and a group of 45 reading comprehension items and sorted within each group in order of their serial (item identification) number. The format shows the item serial number, category, ease index and count, discrimination index and count, answer key, item objective, new test (CGT) number, original test (DLIEL-ECL) number*, and the date of item acquisition into the data item pool. A sample transmitted output page is shown in Figure 7. The format printed for study and evaluation at SwRI is discussed in the next section.

*This is English Language Branch's code number for the DLIEL ECL test form in which the item happened to be used when the item was entered into the item pool at SwRI.

CGT B

08/27/70 PAGE 2

ITEM	CATEGORY	EI	CT	DI	CT	ANS	OBJECTIVE	NEW ID	OLD ID	ENTRY DATE
1856	AC DG VO 23	.56	1	.09	1	B	PRECISION	CGT B	ECL 56	05/28/70
2036	AC DG VO 11	.17	1	.06	1	D	CALL	CGT B	ECL 59	05/26/70
2088	AC DG ID 11	.21	1	.19	1	C	CHECK ON IT	CGT B	ECL 59	05/26/70
2148	AC QU VO 21	.52	3	.41	3	C	COMPREHENSION OF TWO ADJECTIVE	CGT B	E09660	04/22/70
2161	AC QU VO 14	.82	1	.11	1	D	RAPIDLY	CGT B	ECL 60	05/28/70
2203	AC ST VO 11	.80	1	.05	1	B	CHILDREN	CGT B	ECL 60	05/28/70
2228	AC DG VO 11	.59	1	.30	1	B	HOT	CGT B	ECL 60	05/28/70
2292	AC QU VO 24	.86	1	.16	1	D	DESIRE	CGT B	ECL 61	05/27/70
2307	AC QU VO 12	.92	1	.14	1	A	ENJOYABLE	CGT B	ECL 61	05/27/70
2326	AC ST VO 11	.99	1	.02	1	B	HARD	CGT B	ECL 61	05/27/70
2334	AC ST VO 23	.94	1	.05	1	C	EXCELLENT	CGT B	ECL 61	05/27/70
2347	AC DG ID 12	.36	1	.11	1	B	MORE THAN ENOUGH	CGT B	ECL 61	05/27/70
2351	AC DG ID 24	.60	1	.24	1	B	THINK IT OVER	CGT B	ECL 61	05/27/70
2475	AC DG ID 22	.46	2	.13	2	D	BAD SHAPE	CGT B	E0169	03/25/70
2525	AC QU VO 12	.81	3	.20	3	C	PORTABLE	CGT B	E0968D	04/22/70
2556	AC ST VO 12	.73	3	.25	3	C	CONSIDER	CGT B	E0968D	04/22/70
2659	AC ST VO 23	.41	2	.33	2	A	DISMISSED	CGT B	ECL 39	02/13/70
2718	AC QU VO 21	.64	1	.37	1	B	RANGE/DISTANCE	CGT B	E0369	10/15/69
2827	AC QU VO 25	.29	2	.24	2	B	ARTILLERY	CGT B	E0469A	12/17/69
2999	AC ST VO 13	.86	2	.13	2	B	CLOSE	CGT B	E0569A	04/01/70

FIGURE 7. SAMPLE CGT PRINTOUT OF TRANSMITTAL FORMAT

The CGT preparation comprises the following steps:

(1) The program assembles a test form.

(2) The form is printed out in the study format.

(3) The "statistics" analysis program (see Section C) is applied to the form.

(4) Representative forms are selected for transmission to the Sponsor.

(5) A punched card output of the selected forms is prepared by computer.

(6) The punched cards for each CGT form are sequenced and categorized on a sorter to order the items into a sequence most convenient to the Sponsor (selected by him).

(7) The CGT card sequence is printed, and the listing transmitted to the Sponsor.

4. Computer Program Evaluation and Analysis at SwRI

Seventeen completed CGTs were assembled in mid-August. A summary of their major characteristics is presented in Table X, which follows the same general format as Table VIII. A detailed discussion of the comparisons between the CGTs and DLIEL-ECLs, based on English Language Branch test results, will be presented in Section III.E.

On comparing Tables VIII and X, it appears that in several respects the CGT forms are more uniform than the ECL test forms. Between CGT forms, there is less variability in the number of out-of-index range items and in the distribution of the test mean ease and discrimination indexes. Also, all the CGT forms have the same total number of items and identical content distributions (per Table VII), since these characteristics are predetermined by the computer program, whereas the corresponding aspects of the ECL tests are variable. The greater uniformity of the CGT forms reflects the acquisition of items by near-random sampling of the large item pool and the firm content constraints imposed by the assembly methodology.

To aid SwRI's study of the test generation process, each CGT was printed out in a format shown in Figure 8. This format supplies a detailed account of the events occurring during a test generation. The format provides the following information (reference Figure 8):

- Line (1) shows the heading for the test items (identification number, category code, ease and discrimination index, answer key, objective, DLIEL-ECL source, and item acquisition date);

- Line (2) shows a typical item acquired for this specific CGT form;

- Line (3) is an intermediate summary printout which appears when one of the category sums requirements has been completed with the incorporation of the last acquired item; the line shows that, to this point, 82 items had been rejected because a sum requirement related to their category membership had already been satisfied, a total of 192 items had been sampled, 110 items had been accepted, and the last item accepted was item Serial No. 4589, categorical descriptor RC ST CS Book 1;

TABLE X. SUMMARY OF THE FIRST SET OF CGT FORMS

Test No.	No. of Items	Exceptions, EI		Total Exceptions		No. of "Good" Items	Mean EI	Mean DI	Content										Type EI Distribution
									Lexical		Struc-	Books		Lexical		AC			
		⩽0.37	⩾0.94						VO	ID	tural	VO1	VO2	RC	AC	QU	ST	DG	
1	120	14	4	19	16%	101	0.63	0.18	72	18	30	29	43	15	75	30	30	15	Bimodal
2	120	10	0	10	8%	110	0.63	0.20	DITTO			DITTO		DITTO		DITTO			Bimodal
3	120	18	1	20	17%	100	0.60	0.20	"			"		"		"			Bimodal
6	120	14	6	20	17%	100	0.61	0.18	"			"		"		"			Flat
7	120	19	4	23	19%	97	0.61	0.19	"			"		"		"			1-Mode
8	120	18	4	22	18%	98	0.63	0.20	"			"		"		"			1-Mode
9	120	15	4	20	17%	100	0.63	0.18	"			"		"		"			1-Mode
10	120	19	1	20	17%	100	0.62	0.19	"			"		"		"			1-Mode
12	120	15	4	19	16%	101	0.63	0.20	"			"		"		"			1-Mode
13	120	12	1	13	11%	107	0.61	0.20	"			"		"		"			Flat
14	120	15	1	16	13%	104	0.63	0.20	"			"		"		"			Bimodal
15	120	12	1	13	11%	107	0.63	0.20	"			"		"		"			Bimodal
19	120	17	4	22	18%	98	0.64	0.18	"			"		"		"			Bimodal
21	120	13	2	15	13%	105	0.63	0.21	"			"		"		"			Flat
22	120	9	3	12	10%	108	0.66	0.18	"			"		"		"			Bimodal
23	120	22	3	25	21%	95	0.62	0.19	"			"		"		"			1-Mode
25	120	15	6	21	18%	99	0.64	0.18	"			"		"		"			1-Mode

```
                    ECL COMPUTER GENERATED TEST     2
08/18/70                                                    PAGE  4                    5,878

ITEM  CATEGORY    EI    DI  ANS  OBJECTIVE                          FILE DATE                    <-- 1

2139 RC ST CS C   ,45   ,18  C   THERE MAY BE SEVERAL, ETC          ECL 59  05/26/70             <-- 2
4589 RC ST CS D   ,66   ,19  C   HOLD THE PHONE                     ECL 55  03/18/70

 INOFIT =   82  IRANCT =  192  ITMCT = 110  ITEMS =4589  RCSTCS00                               <-- 3

 ITALLY =   8   0   0   0   0   1   0   1   0   0   0   0   0   0   8   0   8   3   0   5    <-- 4
3712 AC DG VO 2   ,57   ,14  B   GUARANTEE                         ECL 29  01/16/70
1722 AC DG VO 2   ,58   ,25  C   APPROVED                          ECL 54  05/26/70
4554 AC DG VO 2   ,64   ,42  C   HOBBY                             ECL 55  03/18/70
3768 AC GU VO 2   ,91   ,13  B   GOOD/QUITE WELL                   E08698  04/30/70
2296 AC GU VO 2   ,43   ,38  A   REPLY                             ECL 61  05/27/70
3657 AC GU VO 2   ,52   ,13  A   ABOARD                            ECL 29  01/16/70

 INOFIT =  127  IRANCT =  243  ITMCT = 116  ITEMS =3657  ACGUVO25

 ITALLY =   2   0   0   0   0   1   0   1   0   0   0   0   0   0   2   0   2   0   0   2    <-- 4a
4551 AC DG VO 2   ,29  0,00  C   IMPOSING                          ECL 55  03/18/70
4312 AC DG VO 2   ,49   ,12  A   TO GRAB,,,PURSE                   ECL 47  03/06/70

 INOFIT =  146  IRANCT =  265  ITMCT = 118  ITEMS =4312  ACDGVO22

 ITALLY =   0   0   0   0   0   1   0   1   0   0   0   0   0   0   0   0   0   0   0   0
2374 RC CN IN D   ,76   ,10  C   TO HEAR                          ECL 61  05/27/70

 INOFIT =  188  IRANCT =  308  ITMCT = 119  ITEMS =2374  RCCNIN00

 ITALLY =   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0
3486 RC CN PA D   ,42   ,22  D   FIXED                             E0769A  04/21/70

 INOFIT =  359  IRANCT =  487  ITMCT = 120  ITEMS =3486  RCCNPA00

 ITALLY =   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0

  487 = NO, OF RANDOM SELECTIONS                                                             <-- 5
  359 = NO, OF ELEMENTS NOT FITTING REQUIREMENTS                                             <-- 6
    6 = NO, OF DUPLICATES                                                                    <-- 7
  120 = NO, OF ELEMENTS                                                                      <-- 8
```

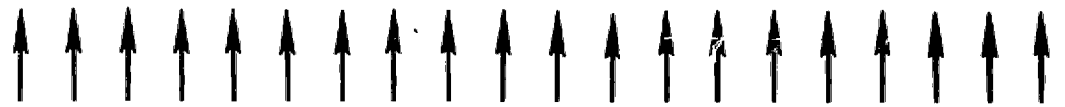


FIGURE 8. SAMPLE CGT PRINTOUT OF STUDY FORMAT

- Line (4) shows the category sum totals still required at this point (the key explaining the 20 positions is found superposed at the bottom of Figure 8). Comparing line 4 with its later counterpart line 4a, we see that the "QU" count has gone from 3 to 0. This shows that the last item accepted (Serial No. 3657) completed the "QU" sums requirement.

- Line (5) shows the total number of random selections made before the 120-item test was completed;

- Line (6) shows the total number of elements rejected because their categorical requirements had been satisfied at time of their sampling;

- Line (7) shows the total number of sample duplicates (item already acquired for this test form) rejected during the sampling procedure; and

- Line (8) shows that there were 120 elements in the completed test.

We chose to terminate a CGT run when the number of elements rejected for not fitting category requirements (Line 6, Figure 8) reached 1000. We chose this number because an estimate indicated that most computer assembly runs should be completed before 1000 elements were rejected, and, thus, the limit should provide a reasonable margin of safety against anomalous long runs without terminating too many runs with incomplete tests. In the first group of 25 generated test lists, 17 completed tests were generated with rejection totals in the range of 100 to 500 items per test. Complete and incomplete tests alike required approximately 4 sec of computer time per test form to generate.

The actual computer costs directly associated with CGT item selections are nominal. If we assume that about 25 or more form lists are being assembled at one time, the cost to generate each test form list is approximately $2.80, including all processing and hard copy outputs. This is the cost, after the CGT computer technology has become operational, of assembling, in the sequence preferred by English Language Branch, a listing of 120 items for transmittal to the Sponsor.

The study format furnished information concerning the reasons for not completing some of the test assemblies before the 1000th reject item count was reached. It appears that, on occasion, a test generation run would encounter difficulties because of certain imbalances in the data pool subsets. With an ideal, numerically balanced item pool, the probability of sampling a given category would be independent of the subset being sampled. To assure this statistical independence, the proportion of items in an intersection subset should be the product of the proportions of the relevant supersets. This is the empirical equivalent of multiplying unconditional probabilities to achieve statistical independence for conditional probabilities. For example, with reference to Figure 6, since the AC/QU set is required to have 30 items or 30/90 = 1/3 of the lexical items per test and the ID set must have 18 items or 18/90 = 1/5 of the lexical items per test, then, for a balanced, ideal pool, the AC/QU/ID subset should contain $1/3 \times 1/5 = 1/15$ of all the lexical items in the data pool. When a given subset has a disproportionately large membership compared to the "ideal" (statistically independent) proportion, then with random sampling that subset will be sampled disproportionately frequently, and this in turn will cause the branch total counts in certain branches to be reached prematurely, closing off those branches and requiring that the remaining totals be satisfied from other, numerically deficient, subsets. (The argument can, of course, be restated in terms of problems caused by disporportionately small subsets.) Most incomplete tests

had accumulated approximately 115 items at their termination, using 1000-item rejection cutoff limit. A study of these CGTs showed that the failure to generate a complete 120-item test was most commonly due to certain item pool numerical imbalances; because of these imbalances, the ID and AC/DG sums requirements both were among the last to be satisfied. Thus, sometimes only AC/DG/ID items could be accepted toward the end of a test generation run. However, since there were in this category only 19 members in the whole item pool at the time of this CGT assembly, the program was weighted heavily against finding sufficient items to complete the test run when the previously mentioned conditions existed. Table XI shows the actual versus ideal data subset sizes as of August 1970.

One could devise methods for modifying the computer generation program which would circumvent incomplete test generation. For example, one could impose constraints on the sampling procedure, in addition to English Language Branch's content specifications, so that the computer program would randomly sample within selected independent subsets but acquire predetermined totals from each of the 23 subsets identified in the content requirements. We recall from Figure 6 that English Language Branch specifications can be met by constraining only 18 subset sums; we therefore are free, at least in theory, to introduce further constraints. However, additional constraints, while perhaps facilitating test generation, could introduce unknown characteristics into the CGT forms. These effects would be difficult to assess, given the limited amount of pretest experimentation possible at English Language Branch. The possibility of introducing unknown effects by these or similar computer program modifications seemed undesirable; in addition, the occasional generation of an incomplete test, infrequent at present, should become even less frequent as new items from English Language Branch expand and balance the item pool. In any case, the generation of an incomplete test in even as many as one out of three runs is a trivial matter in terms of time, convenience, and economy. Therefore, no specific plans exist at present for dealing with it.

TABLE XI. ITEM POOL COMPOSITION, 31 AUG, 1970

a. Lexical Items			
No.	Set Code	No. of Items	
		Actual	Ideal
1	AC QU VO 1	485	408
2	AC QU VO 2	425	612
3	AC ST VO 1	718	408
4	AC ST VO 2	579	612
5	AC DG VO 1	156	204
6	AC DG VO 2	107	306
7	AC QU ID	107	255
8	AC ST ID	310	255
9	AC DG ID	19	128
10	RC VO 1	307	204
11	RC VO 2	243	306
12	RC ID	240	128

B. Structural Items			
No.	Set Code	No. of Items	
		Actual	Ideal
13	CO	201	43
14	MO	101	170
15	PR	131	128
16	IN	80	85
17	GE	80	85
18	PA	39	85
19	VF	170	128
20	VT	163	128
21	VP	62	128
22	WO	127	128
23	CS	131	170

5. System Compatibility of Computer Programs

All of the computer programs developed during this research activity are written in the FORTRAN IV language. They have been checked and run extensively on a CDC 6400 system using the SCOPE 3.2 executive system and the RUN compiler.

The programs were written with the idea that, at a future date, they may be transferred to another computer system. Therefore, the input data forms are conventional 80-column fixed length records (key to tape data transcribers with variable length records were not used); the output formats use standard FORTRAN specifications (especially nH instead of *. . . .* or '. . . .'); the Hollerith constants are short in length (1 to 4 characters) so that

they may fit the short word size of IBM systems. The numerical calculations are of a statistical nature and should not be affected by round-off error on IBM systems; integer arithmetic does not exceed 4 digits (also compatible with IBM's 1/2 word integer). All subroutine returns are standard and do not use any of the special features that are unique to the CDC compilers (e.g., ENCODE and DECODE). EQUIVALENCE and COMMON statements have been set up to avoid conflicts between different systems. All output has less than 60 lines per page and less than 132 characters per line.

All of the programs are short or built in modular pieces so that they might be used on systems with small partitions of core storage.

The above compatibility considerations limit the speed, ease, and flexibility in writing programs; however, it is hoped that the chosen approach will facilitate and encourage the widespread use of these programs, thus offsetting the above-mentioned limitation.

E. Sponsor's First CGT Evaluation

1. Test Transmittal

Six sample test lists (CGT 8, 9, 13, 14, 21, and 22) were chosen from the first group of 17 completed CGTs and delivered on 31 August 1970 to the Sponsor for validation. The six tests were selected to give two examples each of the three types of ease index distributions identified in Table X (1-Mode, Flat, Bimodal). At the same time, the 46 DLIEL-ECL test forms were screened, and three tests with characteristics similar to those of the selected CGT samples were identified. The three forms were proposed as criterion tests. Each of the three suggested criterion tests was similar to a corresponding pair of CGT sample forms in index means, content, and type of ease index distribution.

2. Validation Results

Validation procedures were performed at English Language Branch, and an evaluation report was transmitted to SwRI on 2 Dec 70. It had not been feasible to use the three ECL forms suggested as validation criterion tests; however, for each pair of CGTs, three sets of DLIEL-ECL test forms were administered for validation. Each of the three validating ECL test administrations used several different ECL forms.

The results of English Language Branch's evaluations, in brief, were:

(1) The CGTs met content specifications very well. (This was to be expected, as the computer program insured correct categorical subset totals.)

(2) There were no significant differences between the observed and computer-calculated CGT index means.

(3) There were no significant differences between the index means of the CGT and DLIEL-ECL tests.

(4) The CGT/CGT correlations were higher for two out of three sample pairs than for the corresponding ECL/CGT or ECL/ECL pairs.

(5) The sampling of CGT test items seemed to be well distributed among the 46 ECL test forms.

(6) Item duplication between CGT test forms was acceptable.

(7) The number of statistically unacceptable items (out of ease or discrimination index tolerance range) for each CGT remained substantially constant before and after pretesting.

(8) The CGT reliability indexes were comparable to those of the DLIEL-ECL forms.

Interpreting the results of the validation analysis is made difficult by the presence of several subtle factors whose influence cannot be readily assessed. These include: the unavoidable use of several DLIEL-ECL forms for each "single" criterion test administration; the use of index range criteria to screen items at English Language Branch in contrast to SwRI's use of the full unrestricted item pool for the first CGT set; effects on item index stability of a recent changeover at English Language Branch to a different formula for calculating the item indexes; and possible effects due to the historical grouping (influenced by item age) of items in the forms of the DLIEL-ECL tests, contrasted with the chronologically random items assembled in a sample CGT form. These factors suggest that validation results must be interpreted with care. With that reservation, we present a summary of the results of the Sponsor's correlations analysis in Table XII. It can be seen from Table XII that, for each pretesting group (three DLIEL-ECL and two corresponding CGT scores), the three kinds of correlation coefficients (ECL/ECL; ECL/CGT; CGT/CGT) are comparable. As has been mentioned, two of the CGT/CGT correlation coefficients are high compared to the other coefficients in the group, and the Sponsor's analysis showed these to be significant (CGT #13 and #14, at the 0.01 level, CGT #21 and #22, at the 0.05 level).

TABLE XII. CORRELATION SUMMARY

CGT Pair #	ECL/ECL r	ECL/CGT r	CGT/CGT r
8 and 9	0.58	0.64	0.72
	0.72	0.65	
	0.75	0.75	
		0.62	
		0.66	
		0.78	
13 and 14	0.74	0.75	0.86
	0.77	0.79	
	0.72	0.75	
		0.73	
		0.77	
		0.74	
21 and 22	0.87	0.85	0.93
	0.83	0.84	
	0.85	0.89	
		0.86	
		0.84	
		0.86	

It was pointed out by the Project Technical Monitor that the number of out-of-index range items in these CGTs is higher than the number found in the current DLIEL-ECL tests (which are hand screened after administration for validation purposes). The Technical Monitor suggested that remedial measures be developed with the hope that consistent achievement of higher interest correlations would result. Since at the same time a large quantity of new ease and discrimination index data was furnished by English Language Branch for updating the original item pool, the development of data base modifications was timely and accordingly implemented as described in the next sections.

F. Item Pool Updating

An "update" computer program was developed so that item pool changes, additions, or deletions can be made. The program is used to keep the CGT item pool status current.

When we receive updating information and instructions from English Language Branch, the computer program will perform the requested operations and prepare a report (to be discussed below).

The staff at English Language Branch will inspect the report and at an appropriate time prepare a new set of updating instructions, repeating the cycle. The CGT program will use the most recent item file generated by the update program, unless special instructions to the contrary are received. Copies of each new update report will be filed at English Language Branch and SwRI. In general, the program performs the following functions on instruction:

(1) It changes or corrects existing item data by incorporating new information concerning item categories, objectives, answers, and/or index values.

(2) It deletes items.

(3) It accepts new items.

(4) It identifies items requiring attention. (Input errors; EI, or DI out-of-range.)

(5) It prepares a summary status report on the revised item pool.

We mentioned in Section B.2 that an "index count" datum is included in the item information. The reason for its inclusion is that it is used in revising the item's EI and DI as new history accumulates on its performance. Each time a new DI and EI is calculated on an item by English Language Branch, the calculation is based on administrations to approximately the same size student population. To update the cumulative index, it is therefore desirable to weight the new index figure in order to ascribe to each test administration the same weight. For example, if an item previously had an (accumulated) EI of 0.55 and has an EI of 0.63 from the current administration of that item, we would compute the new EI as $(0.55N + 0.63)/(N + 1)$, where N is the old count.

We currently classify items as out-of-range if the EI is < 0.30 or > 0.93, and/or if the DI < 0. These limits, which are incorporated in the update program, could be adjusted by a trivial change in the program should the current index range criteria be revised by the Sponsor.

To illustrate the "Update" program, a sample pool of 33 items has been created to illustrate most of the features of the program. Figure 9 is a dump of these items. For purposes of this illustration, the "Objective" field is used to comment about the various items (normally, the item objective appears in this field). Figure 10 is a listing of the cards used to make an update run on the sample pool. The first 14 cards are changes to the old pool. The 15th card is a new item to be added to the pool. Figures 11 and 12 are examples of the report format generated by the program.

Each update card is documented in "Exceptions and Update Report," (Fig. 11). The column of item numbers in the center of the report lists all of the items updated. The column of item numbers on the left-hand side indicates all items in the pool that have fields of information that are out-of-range or invalid. At the end of the report, we list 5 transaction statistics: the number of records read from the input master, the number of records written on the output master, the number of new items created, the number of items deleted and, finally, the number of items changed.

Figure 12 is a one-page report on the content of the entire pool. Counts and percentages are given for individual categories, combined categories, and the distribution of the answer keys. The

THIS IS THE IN PUT MASTER FILE SAMPLE

```
 0 01/14/71
9001RCCNC000024,733024,0330ASHOWS SIZE OF EI DI FIELDS              SMPLE1011471
9002RCCNM001  .99 1  .99 1BSHOWS EI OUT OF RANGE HIGH               SMPLE1011471
9003RCCNPR02  .28 1  .00 1CSHOWS EI OUT OF RANGE LOW                SMPLE1011471
9004RCCNIN1300,2901-0.0101DSHOWS DI OUT OF RANGE                    SMPLE1011471
9005RCCNGE1400,2701-0.0101ASHOWS BOTH OUT OF RANGE EI LOW           SMPLE1011471
9006RCCNPA2100,9401-0.0101BSHOWS BOYH OUT OF RANGE EI HGH           SMPLE1011471
9007RCCNVF2200,300100.1901E SHOES BAD ANSWER KEY                    SMPLE1011471
9008RCCNVT2300,310100.2001  SHOWS BLANK ANS KEY                    SMPLE1011471
9009RCCNVP24          .2101ASHOWS BLANK E1 FIELD                    SMPLE1011471
9010RCCNW03500,3201       B8HOWB BLANK DI FIELD                     SMPLE1011471
9011RCQUN0  00,330100,2201CSHOWS BLANK BOOK NO, OK                  SMPLE1011471
9012RCBTW01100,340100,2301DCORRECT ITEM                             SMPLE1011471
9013RCBTCS0000,350100,2401ACORRECT ITEM                             SMPLE1011471
9014RCQUC80000,360100,2501BCORRECT ITEM                             SMPLE1011471
9015RCCNCS0000,370100,2601CCORRECT ITEM                             SMPLE1011471
9016ACQUV01100,380100,2701DCORRECT ITEM                             SMPLE1011471
9017ACBTV01200,390100,2801ACORRECT ITEM                             SMPLE1011471
9018ACBTV02300,400100,2901BCORRECT ITEM                             SMPLE1011471
9019ACDGV01400,410100,3001CCORRECT ITEM                             SMPLE1011471
9020ACDGV02100,420100,3101DCORRECT ITEM                             SMPLE101 71
9021ACQUV02200,430100,3201ACORRECT ITEM                             SMPLE1011471
9022ACDGV01300,440100,3301BCORRECT ITEM                             SMPLE1011471
9023ACDGV02400,450100,3401CCORRECT ITEM                             SMPLE1011471
9024ACQUID1100,460100,3501DCORRECT ITEM                             SMPLE1011471
9025ACSTID2100,470100,3601ACORRECT ITEM                             SMPLE1011471
9026ACDGID1200,480100,3701BCORRECT ITEM                             SMPLE1011471
9027RCCNV01300,490100,3801CCORRECT ITEM                             SMPLE1011471
9028RCCNV02500,500100,3901DCORRECT ITEM                             SMPLE1011471
9029RCCNID1100,510100,4001ACORRECT ITEM                             SMPLE1011471
9030ARQUV01100,520100,4101BBADLISTENING/READING CATGRY              SMPLE1011471
9031ACVNID1200,530100,4201CBAD FORM OF PRESENTATION CTGRY           SMPLE1011471
9032ACQUN01300,540100,4301DBAD LEXICAL/STRUCTURAL CATGRY            SMPLE1011471
9033ACQUV01500,550100,4401ABAD BOOK NO.                             SMPLE1011471
```

FIGURE 9. SAMPLE INPUT MASTER FILE

```
 THIS IS THE SAMPLE UPDATE FILE ALL NEW ITEMS
9002      11                                                          U
9003      12                                                          U
9004          .35 1                                                   U
9007                      D                                           U
9008                      A                                           U
9009              0         ZERO COUNT FLAGS NO HISTORY               U
9010                      0 ZERO COUNT FLAGS NO HISTORY               U
9012                  .04 1                                           U
9017          .20 1                                                   U
9029                                                                  U
9030AC                                                                U
9031  UN                                                              U
9032    MO                                                            U
9033      25                                                          U
9034RCSTID12  .96 1  .48 INCORRECT ITEM                  SMPLE1020171
```

FIGURE 10. SAMPLE NEW ITEM CARD LISTING

OLD MASTER FILE VOL 0 DATE 01/14/71

NEW MASTER FILE VOL 1 DATE 02/02/71

DEFENSE LANGUAGE INSTITUTE ENGLISH LANGUAGE BRANCH
DATA BASE EXCEPTIONS AND UPDATE REPORT

PAGE 1

EXCEPTIONS					CHANGES							COMMENTS
	E.I.		D.I.	INVALID		E.I.		D.I.		ANS.	TEST ID	MISCELLANEOUS
ITEM	LT .30	GT .93	LTD.00	CATEGORY ANS	ITEM	OLD	NEW	OLD	NEW	OLD NEW	OLD NEW	EXCEPTIONS OR CHANGES
					9002							01 TO 11 CATG
9002		.99										
					9003							02 TO 12 CATG
9003	.28											
					9004	.29 1	.64 2					
9004			-.01									
9005	.27											
9005			-.01									
9006		.94										
9006			-.01									
					9007					E D		
					9008					A		
					9009							CHANGED-OBJECTIVE E.I. COUNT = 0
9009												
9009=0.00												
					9010							CHANGED-OBJECTIVE D.I. COUNT = 0
9010												
					9012			.23 1	.27 2			
					9017	.39 1	.59 2					
9017	.29					.39 *						
					9029							DELETED
					9030							AR TO AC CATG
					9031							VN TO UN CATG
9031				ACUN								
					9032							NO TO NO CATG
9032				AC NO								
					9033							13 TO 25 CATG
					9034							NEW

ITEMS IN = 33
OUT = 33
NEW = 1
DELETED = 1
CHANGED = 13

FIGURE 11. SAMPLE UPDATE REPORT–ITEM DISPLAY

CATEGORY	NO. . .	DATA BASE	PERCENT	IDEAL DATA BASE PERCENT	ACTUAL VS IDEAL PERCENT
AC		15	45.45	62.50	72.73
RC		18	54.55	37.50	145.45
QU		8	24.24		
ST		6	18.18		
DG		5	15.15		
CN		13	39.39		
UN		1	3.03		
VO		12	36.36	60.00	60.61
ID		5	15.15	15.00	101.01
CO		1	3.03	.83	363.65
MO		2	6.06	3.33	181.82
PR		1	3.03	2.50	121.21
IN		1	3.03	1.67	181.81
GE		1	3.03	1.67	181.81
PA		1	3.03	1.67	181.81
VF		1	3.03	2.50	121.21
VT		1	3.03	2.50	121.21
VP		1	3.03	2.50	121.21
MO		3	9.09	2.50	363.64
CS		3	9.09	3.33	272.73
ACQUVO 1		2	6.06	8.00	75.76
ACQUVO 2		2	6.06	12.00	50.51
ACSTVO 1		1	3.03	8.00	37.88
ACSTVO 2		1	3.03	12.00	25.25
ACDGVO 1		2	6.06	4.00	151.52
ACDGVO 2		2	6.06	6.00	101.01
ACQUID--		1	3.03	5.00	60.61
ACSTID--		1	3.03	5.00	60.61
ACDGID--		1	3.03	2.50	121.21
RC--VO 1		1	3.03	4.00	75.76
RC--VO 2		1	3.03	6.00	50.51
RC--ID--		1	3.03	2.50	121.21
ANS CODE					
A		9	27.27		
B		9	27.27		
C		7	21.21		
D		8	24.24		

OUT OF RANGE BREAK DOWN

E.I. LOW = 4
E.I. HIGH = 2
D.I. LOW = 3

TOTAL ITEMS = 7

MEAN E.I. = .44
MEAN D.I. = .30

FIGURE 12. SAMPLE UPDATE REPORT–ITEM POOL SUMMARY

number of items with out-of-range values is summarized and, finally, the pool mean EI and DI are given.

Figure 13 shows a dump of the new "Master Item Pool File."

G. Second Prototype CGT Set

Prior to generating the second set of CGTs, we incorporated updating information, furnished by English Language Branch, on the ease and discrimination index counts of 3746 data pool items. After this revision, it was found that there were 549 out-of-range items in the pool. On instructions from the Technical Monitor, we removed all those items from the pool and constructed a new updated file which contained 4848 "good" in-range items. The program described in the preceding section (F) was used for the updating; the category summary of this revised item pool is shown in Figure14.

Using the new item pool, the CGT program assembled a second prototype set of 25 ECL test form lists. The characteristics shown by the CGT program during this second run substantially duplicated the features encountered during the first CGT assembly. Once again, the set of 25 tests contained 17 completed CGTs, and the history of test generation, the reasons for not completing all tests, and other general features of the computer program performance were substantially identical to the characteristics described for the first CGT set in Section D.4. The slight reduction in size of the item pool used for the second set (4848 items versus 5111 items in the first 25 CGT data pool) did not appear to affect the effectiveness of the program. Since not only the item pool size but also its categorical composition remained approximately the same after the final update, the similarities of outcome of the two test generation assembly runs seem reasonable.

The characteristics of the 17 completed CGTs are shown in Table XIII. Comparing Tables XIII and X, we note that the major difference is the absence of out-of-index range items in the second set of tests. This is, of course, the direct consequence of removing out-of-range items from the item pool samples by the Update program. The mean EI of the second set of 17 tests was slightly higher than that of the first set of 17 tests (0.646 versus 0.626), while the mean DI was essentially the same for both sets of 17 tests (0.194 versus 0.192).

From the second set of test forms lists, six representative CGTs were selected and transmitted to English Language Branch on January 5, 1971. These research end items were presented to the Sponsor for approval.

H. Sponsor's Second CGT Evaluation

The evaluation study performed at English Language Branch on the second set of CGT tests supported the conclusions reported in Section E (sponsor's first CGT evaluation). It appears that the computer program is selecting 120-item sets which generate valid ECL test forms.

In summary, English Language Branch's conclusions from the second evaluation were:

(1) The CGTs precisely met the content specifications of the model ECL tests.

(2) There were no significant differences between the observed and computer-calculated CGT index means. This implies that the cost of validating CGT forms may be lessened; the test index means now may not require recalculation after administration, since the computer-calculated index means appear to be acceptable estimators.

THIS IS THE OUTPUT MASTER FILE

```
  1 02/02/71
9001RCCNCC0024,733024.033CASHOWS SIZE OF EI DI FIELDS                SMPLE1011471
9002RCCNMX011  ,99 1  .99 1BSHOWS E' OUT OF RANGE HIGH               SMPLE1011471
9003RCCNPR12  ,28 1 0.00 1CSHOWS E. OUT OF RANGE LOW                 SMPLE1011471
9004RCCNIN13  ,64 2 -.01 1DSHOWS DI OUT OF RANGE                     SMPLE1011471
9005RCCNGE14  ,27 1 -.01 1ASHOWS BOTH OUT OF RANGE EI LOW            SMPLE1011471
9006RCCNPA21  ,94 1 -.01 1BSHOWS BOTH OUT OF RANGE EI HGH            SMPLE1011471
9007RCCNVF22  ,30 1  .19 1D SHOFS BAD ANSWER KEY                     SMPLE1011471
9008RCCNVT23  ,31 1  .20 1A SHOWS BLANK ANS KEY                      SMPLE1011471
9009RCCNVP24-0.00-0  .21 1AZERO COUNT FLAGS NO HISTORY               SMPLE1011471
9010RCCNWO25  ,32 1-0.00-0BZERO COUNT FLAGS NO HISTORY               SMPLE1011471
9011RCQUWO    ,33 1  .22 1CSHOWS BLANK BOOK NO. OK                   SMPLE1011471
9012RCSTWO11  ,34 1  .27 2DCORRECT ITEM                              SMPLE1011471
9013RCSTCS00  ,35 1  .24 1ACORRECT ITEM                              SMPLE1011471
9014RCQUCS00  ,36 1  .25 1BCORRECT ITEM                              SMPLE1011471
9015RCCNCS00  ,37 1  .26 1CCORRECT ITEM                              SMPLE1011471
9016ACQUVO11  ,38 1  .27 1DCORRECT ITEM                              SMPLE1011471
9017ACSTVO12  ,59 2  .28 1ACORRECT ITEM                              SMPLE1011471
9018ACSTVO23  ,40 1  .29 1BCORRECT ITEM                              SMPLE1011471
9019ACQGVO14  ,41 1  .30 1CCORRECT ITEM                              SMPLE1011471
9020ACQGVO21  ,42 1  .31 1DCORRECT ITEM                              SMPLE1011471
9021ACQUVO22  ,43 1  .32 1ACORRECT ITEM                              SMPLE1011471
9022ACQGVO13  ,44 1  .33 1BCORRECT ITEM                              SMPLE1011471
9023ACQGVO24  ,45 1  .34 1CCORRECT ITEM                              SMPLE1011471
9024ACQUID11  ,46 1  .35 1DCORRECT ITEM                              SMPLE1011471
9025ACSTID21  ,47 1  .36 1ACORRECT ITEM                              SMPLE1011471
9026ACQGID12  ,48 1  .37 1BCORRECT ITEM                              SMPLE1011471
9027RCCNVO13  ,49 1  .38 1CCORRECT ITEM                              SMPLE1011471
9028RCCNVO25  ,50 1  .39 1DCORRECT ITEM                              SMPLE1011471
9030ACQUVO11  ,52 1  .41 1BRADLISTENING/READING CATGRY               SMPLE1011471
9031ACUNIO12  ,53 1  .42 1CRAD FORM OF PRESENTATION CTGRY            SMPLE1011471
9032ACQUMO13  ,54 1  .43 1DRAD LEXICAL/STRUCTURAL CATGRY             SMPLE1011471
9033ACQUVO25  ,55 1  .44 1ARAD BOOK NO.                              SMPLE1011471
9034RCSTID12  ,56 1  .45 1BCORRECT ITEM                              SMPLE1020171
```

FIGURE 13. SAMPLE UPDATED MASTER FILE

CATEGORY	NO. IN DATA BASE	PERCENT	IDEAL DATA BASE PERCENT	ACTUAL VS IDEAL PERCENT
AC	2781	57.36	62.50	91.78
RC	2067	42.64	37.50	113.70
GU	1043	21.51		
ST	1703	35.13		
DG	305	6.29		
CN	1539	31.75		
UN	258	5.32		
VO	2979	61.45	60.00	102.41
ID	656	13.53	15.00	90.21
CO	188	3.88	.83	465.37
MO	98	2.02	3.33	60.64
PR	124	2.56	2.50	102.31
IN	76	1.57	1.67	94.06
GE	77	1.59	1.67	95.30
PA	40	.83	1.67	49.50
VF	158	3.26	2.50	130.36
VT	153	3.20	2.50	127.89
VP	61	1.26	2.50	50.33
WO	107	2.21	2.50	88.28
CS	129	2.66	3.33	79.83
ACGUVO 1	455	9.39	8.00	117.32
ACGUVO 2	416	8.58	12.00	71.51
ACSTVO 1	672	13.86	8.00	173.27
ACSTVO 2	539	11.12	12.00	92.65
ACDGVO 1	165	3.40	4.00	85.09
ACDGVO 2	110	2.27	6.00	37.82
ACGUID--	110	2.27	5.00	45.38
ACSTID--	288	5.94	5.00	118.81
ACDGID--	20	.41	2.50	16.50
RC--VO 1	277	5.71	4.00	142.84
RC--VO 2	345	7.12	6.00	118.61
RC--ID--	238	4.91	2.50	196.37
ANS CODE				
A	997	20.57		
B	1347	27.78		
C	1383	28.53		
D	1121	23.12		

OUT OF RANGE BREAK DOWN

E.I. LOW = 0
E.I. HIGH = 0
D.I. LOW = 0

TOTAL ITEMS = 0

MEAN E.I. = .66
MEAN D.I. = .19

FIGURE 14. STATISTICS OF LATEST ITEM POOL (JANUARY 1971)

TABLE XIII. SUMMARY OF THE SECOND SET OF CGT FORMS

Test No.	No. of Items	Exceptions, EL		DI	Total Exceptions	No. of "Good" Items	Mean EI	Mean DI	Content									Type EI Distribution
									Lexical		Struc-	Books		Lexical		AC		
		< 0.30	> 0.93	> 0					VO	ID	tural	1	2	RC	AC	QU	ST	DG	
26	120	0	0	0	0	120	0.64	0.20	72	18	30	29	43	15	75	30	30	15	Flat
29	120	"	"	"	"	120	0.68	0.19	DITTO			DITTO		DITTO		DITTO			1-Mode
31	120	"	"	"	"	120	0.68	0.19	"			"		"		"			1-Mode
32	120	"	"	"	"	120	0.63	0.19	"			"		"		"			Flat
33	120	"	"	"	"	120	0.64	0.20	"			"		"		"			Bimodal
34	120	"	"	"	"	120	0.64	0.19	"			"		"		"			1-Mode
35	120	"	"	"	"	120	0.63	0.19	"			"		"		"			Flat
36	120	"	"	"	"	120	0.64	0.20	"			"		"		"			Bimodal
37	120	"	"	"	"	120	0.66	0.18	"			"		"		"			1-Mode
38	120	"	"	"	"	120	0.67	0.18	"			"		"		"			Flat
39	120	"	"	"	"	120	0.65	0.20	"			"		"		"			Flat
40	120	"	"	"	"	120	0.65	0.20	"			"		"		"			1-Mode
41	120	"	"	"	"	120	0.65	0.20	"			"		"		"			Flat
43	120	"	"	"	"	120	0.63	0.21	"			"		"		"			Bimodal
44	120	"	"	"	"	120	0.65	0.20	"			"		"		"			1-Mode
46	120	"	"	"	"	120	0.65	0.21	"			"		"		"			1-Mode
50	120	"	"	"	"	120	0.67	0.20	"			"		"		"			1-Mode

(3) There were no significant differences between the index means of the CGT and DLIEL-ECL tests.

(4) The CGT/CGT correlations were higher than the corresponding ECL/CGT correlations (typically, by 0.05 to 0.09).

(5) The sampling of CGT test items seemed to be well distributed among the 50 different ECL tests.

(6) There were no unacceptable items before pretesting, since these were removed from the item pool. After pretesting, 46 items emerged out-of-range (according to calculations based only on these pretesting data) from the 6 CGT prototype forms.

(7) The CGT reliability indexes were comparable to those of the DLIEL-ECL forms; the reliability index (K-R No. 21) for the 6 tests ranged between 0.91 and 0.93.

On the basis of these results of English Language Branch's evaluation, it appears that overall performance of the CGT forms was improved as a result of the item pool update, which removed all out-of-range items. Further, the evaluation recommended that the 6 forms "be put into operational use at DLIEL."

I. Other Related Efforts

A small amount of project time was devoted to acquainting ourselves with relevant techniques and developments in the fields of computer-generated tests and analysis of test data. This section will summarize these supplementary efforts.

Our Technical Monitor called attention to a computer test generation project conducted by the U.S. Army.(1) In following up his suggestion, we held several conversations with personnel of the U.S. Army Enlisted Evaluation Center, Fort Benjamin Harrison, Indiana (Dr. R. O. Waldkoetter and Mr. J. L. Finucane). We also had the opportunity to examine relevant report drafts made available, courtesy of these researchers. It was concluded that, while the Army's project was of intrinsic interest, the problems it was addressing were different from our problems in major respects so that the MOS Item Bank techniques were not directly applicable to our program development at this time. However, it is quite possible that further developments at Fort Benjamin Harrison and at SwRI may make the cited effort applicable to our program, and we plan to remain in communication with the staff of the U.S. Army Enlisted Evaluation Center.

A second development called to our attention by the Project Monitor was the potential application of a "Rasch" model to ECL tests.(2) The referenced paper was reviewed, but it was concluded that a detailed investigation of this methodology would be required before the usefulness of the model could be properly established and that such an investigation was outside of the scope of the present program.

A third area of preliminary investigation concerned the automatic generation of transformation graphs and charts by computer. At the present time, pretesting of a new ECL form at English

(1) Finucane, J. L., "Development of Specification for MOS Test Item Bank," Proceedings of the 11th Annual Conference, Military Testing Association, 1969, pages 28-34.
(2) Moonan, W. J., "Evaluating Trainee Test Performance By a G. Rasch Measurement Model; A Dialogue," Paper at the 11th MTA Conference, 1969.

Language Branch is followed by time-consuming hand preparation of a transformation graph which is used to convert the new test raw scores to ECL scores. The utility of a computer program using polynomial least mean squares approximations was briefly explored and looks promising at this time. Figure 15 shows the result of this investigation. The figure shows a scatter diagram for a new test and ECL criterion test and also a comparison of the hand-prepared transformation graph and the machine-computed second- and fourth-order polynomial approximations. We note the good agreement between the graphs in the higher scoring ranges; in this particular pretesting population, there were few low scorers, so that the transformation graphs in the lower half of the score range tend to be less meaningful.

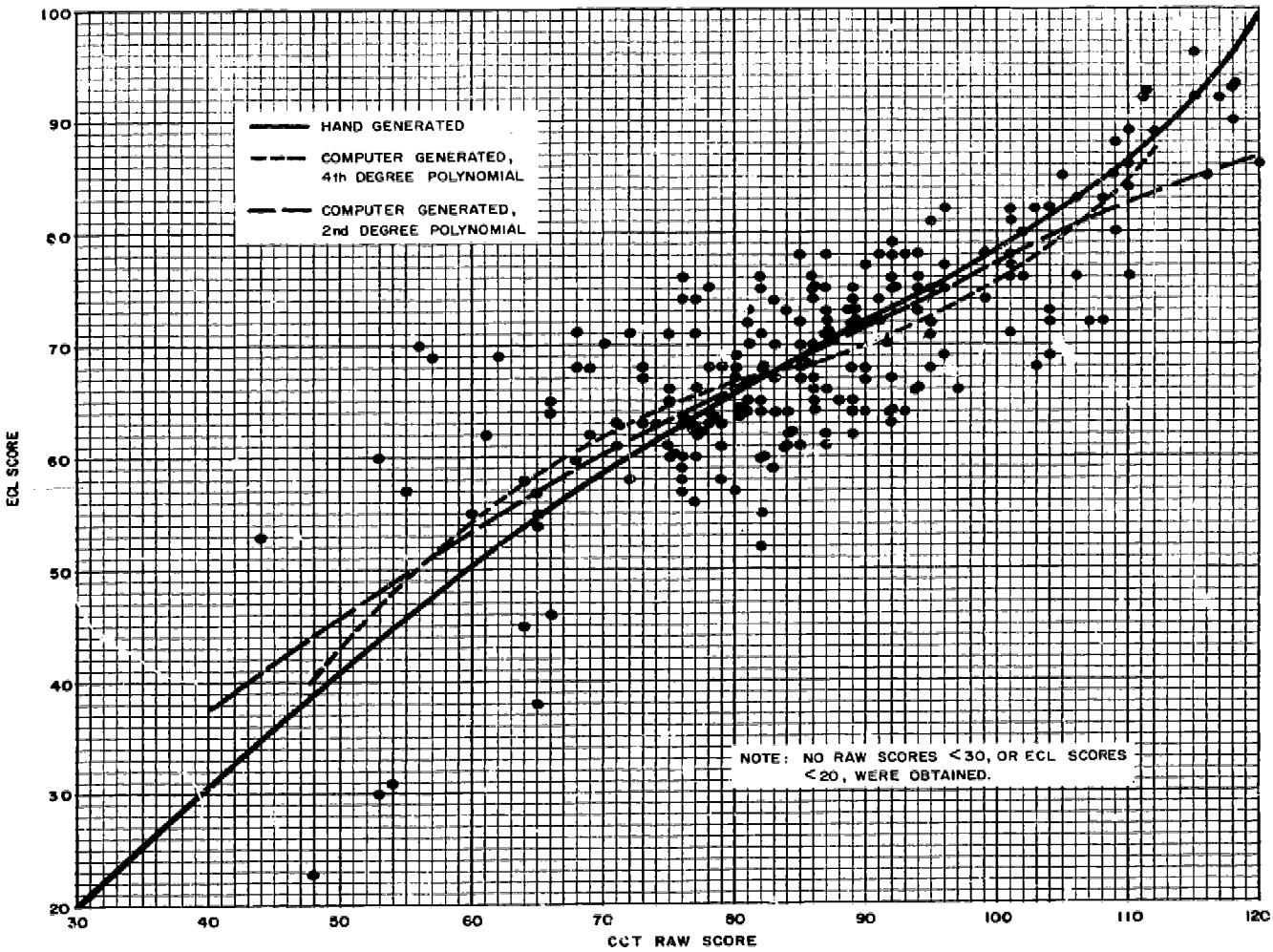


FIGURE 15. DLIEL ECL/CGT SCATTERGRAM AND COMPARISON OF HAND VERSUS COMPUTER-GENERATED TRANSFORMATION GRAPHS

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Southwest Research Institute
8500 Culebra Road
San Antonio, Texas 78228

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
NA

3. REPORT TITLE
A Methodology to Achieve Secure Administration of English Comprehension Level Tests–Phase I

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Final Report, 1 May 1970–11 May 1971

5. AUTHOR(S) *(First name, middle initial, last name)*
Louis S. Berger
Thomas R. Jackson
Thomas E. Hawkins

6. REPORT DATE
10 May 1971

7a. TOTAL NO. OF PAGES
52 plus 7 prelims.

7b. NO. OF REFS
11

8a. CONTRACT OR GRANT NO.
F41609-70-C-0030

b. PROJECT NO.
Work Unit No. 0024

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
None

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES
SwRI Project: 13-2825

12. SPONSORING MILITARY ACTIVITY
Defense Language Institute
English Language Branch
Lackland AFB, Texas

13. ABSTRACT

The English Comprehension Level (ECL) test is Defense Language Institute's (DLI) basic test for measuring English language proficiency of foreign military personnel. It is used for overseas screening and evaluation and during language training. The motivation to perform well on an ECL test is high. Consequently, test compromise is prevalent, particularly in overseas screening tests, causing economic disbenefits to the governments and countries involved.

The reported project comprised two tasks pertaining to secure ECL test administration: Task I, development of a computer methodology which would construct many ECL forms, thus reducing the opportunity for compromise; and Task II, a cost/benefits analysis to provide DLI management with guidelines and information on typical penalty costs. The Task II analysis, based on staff interviews, student records, and related DLI studies, indicated that compromise costs were substantial. Task I developed a computer methodology which will generate well-functioning ECL test forms in virtually unlimited number. The test forms conform to the Sponsor's specifications. Test items are computer-selected in sets of 120 from a 5000-item data pool, listed, and transmitted to the Sponsor for test preparation. Both tasks are reported in detail.

FORM 1 NOV 65 1473

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Achievement Tests						
Psychometrics						
Computer Science						
Automatic Test Assembly						
Language Tests						
English Comprehension Level Tests						
Item Analysis						
Screening Tests						
Test Compromise						
Operations Research Cost Benefits Study						
Defense Language Institute						
Economic Analysis						