THE PREDICTIVE VALIDITY OF SECRETARIAL TESTS GIVEN BY THE INSTITUTE OF EDUCATIONAL RESEARCH IN ADDIS ABABA UNIVERSITY

BY DESALEGN CHALCHISA

A thesis Submitted in (part) fulfilment for the degree of Masters of Arts in Educational Psychology at the Addis Ababa University.

School of Graduate Studies

Addis Ababa University

June, 1991

ADDIS ABABA UNIVERSITY

THE PREDICTIVE VALIDITY OF SECRETARIAL TESTS GIVEN BY THE INSTITUTE OF EDUCATIONAL RESEARCH IN ADDIS ABABA UNIVERSITY

BY

DESALEGN CHALCHISA

FACULTY OF EDUCATION

Approved by Board of Examiners:

LAKEW WOLDETEKLE

Chairman, Department Graduate Committee

MAKONNEN YIMER

Advisor

SHASHI K BARI

Examiner

Darge Wole

Examiner

ACKNOWLEDGEMENTS

I wish to express my deep gratitude to Dr. Mekonnen Yimer, my advisor, for his valuable guidance and suggestion for improvement during the writing of the thesis.

I am also greatly indebted to the Tests and Measurement Unit of the Institute of Educational Research and the Personnel Department of the Addis Ababa University for their help in the collection of the data for writing the thesis.

Finally I would like to acknowledge the staff of the Institute of Educational Research, Ato Fekade G- Egziabher (Director, Minilik Secondary School) and W/ro Serkalem Abate (Guidance officer, Minilik Secondary School) for their never relinquishing stimulation to keep my thesis going and bringing it to a good end.

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF APPENDICES

THE PREDICTIVE VALIDITY OF SECRETARIAL TESTS GIVEN BY THE INSTITUTE OF EDUCATIONAL RESEARCH IN ADDIS ABABA UNIVERSITY

Abstract

The purpose of the present study was to determine the predictive validity of G-III, G-II and G-I Secretarial tests given by Test and Measurement Unit (TMU) of the Institute of Educational Research (IER). The subjects were 102, 97, and 58 G-III, G-II and G-I secretaries of AAU, respectively. The predictors were secretarial tests for G-III, G-II and G-I, work experience and Educational Status. The criterion measures were supervisory ratings(last ratings before testing and two ratings subsequent to promotion). The predictors and criterion measures from the complete data of 76, 71, and 42 G-III, G-II and G-I secretarial candidates, respectively were analysed by the use of simple and stepwise regression method. Items of the objective tests, typing and shorthand tests were analysed for 100 subjects.

The study showed significant ($P<0.05$) correlations between G-III Amharic typing-revised manuscript and the last supervisory ratings before test ($r=0.28$), G-III Amharic typing-straight copy and the first supervisory ratings after promotion ($r=0.21$), G-II English typing-straight copy and

the last supervisory ratings before test (r=0.26), G-II English typing-straight copy and the first supervisory ratings after promotion (r=0.19) and G-I Shorthand-letter II and the first supervisory ratings after promotion (r=-0.32). No other secretarial tests of G-III, G-II and G-I was found to be significant predictor at 0.05 level.

1. INTRODUCTION

1.1 Background of tests

Tests are generally used for the purpose of decision making. There are various types of decisions to be made with the use of tests. Some of the decisions are made at organizational levels and some are made at personal levels. The types of decisions to be made with the use of tests can be classified into administrative decisions which include: selection, screening, classification and placement; guidance decisions; instructional decisions and research decisions (Mehrens and Lehmann, 1984).

In selection decisions one decides whether to accept or reject an individual for a particular job, treatment or program. In the process of employee selection, personnel managers are aided by testing and assessment center - a program where applicants are engaged in specific individual and group exercises and observed by managers or trained observers who evaluate their ability and potential for a particular job. Testing is a tool for examining human resources for qualities relevant to performing available jobs and thereby choosing a competent individual. Although many different kinds of tests are available for organizational use, they generally can be classified into

aptitude tests, achievement tests, vocational interest tests and personality tests (Certo and Appelbaum, 1985).

If carefully designed and professionally interpreted psychological tests can provide very reliable information about a prospective job candidate. According to Zippo (1980) an effective testing program can help to:

(1) reduce turnover, improve organizational functioning and cut-down the need for training programs;

(2) identify unsuitable candidates; and

(3) meet equal employment opportunity requirements.

In Ethiopia the use of specific tests for personnel selection or promotion in the various job categories is very limited. Organizations mostly rely on high school grades, Ethiopian School Leaving Certificate Examination (ESLCE) and College and University grades for hiring employees. There are only few organizations which use their own tests for initial selection. Specially, the lack of testing service units in the nation inhibited the wide use of specific tests

in the various job categories for the purpose of promotion and selection. The Tests and measurement Unit (TMU) of the Institute of Educational Research (IER) is one of the testing centers of its kind in the nation which provides testing service to the various organizations and individuals.

Before its merger with the IER in 1987 the TMU was called the University Testing Center. As stated by Yusuf (1989) the University Testing Centre was established in Addis Ababa University (AAU) in 1966 with the main aim to undertake research on test development, to offer testing services for personnel recruitment, placement and promotions and to develop appropriate standardised aptitude tests for AAU's use .

The principal services rendered by the University Testing Centre were the development and administration of the University Aptitude Tests, the consultation of the Facility of Education and staff member of the AAU on problems of measurement and evaluation and the conduct of research and studies of an institutional nature (Jones, 1973).

At present the TMU renders testing services to various

institutions and individuals in the country. The main services include administrating local and foreign tests, developing the AAU Aptitude Test and conducting research to update the tests now in use by the unit.

The local tests are basically administered to give professional assistance to the Personnel Department of the AAU on employment and promotion decisions. Besides, currently the Unit offers testing services to 106 local institutions. Yearly, the unit gives local tests for about 700-800 individuals. These local tests include: accounting, auditing, clerical, secretarial and other jobs category tests. The other job category tests are tests for retail inspectors, cash collectors, purchasers, record officers, house masters, cleaners, gardeners and guards. The secretarial tests now in use by the unit are executive secretary, senior secretary (G-I), Secretary (G-II), secretary (G-III) and Copy Typist.

The unit also serves as a representative of foreign testing agencies. In this connection the unit administers foreign tests to students seeking admission to graduate and undergraduate programs in various foreign universities and to other individuals for certification purposes. Yearly the unit administers foreign tests for about 600-800

individuals. The foreign tests administered by the unit include among others: the Graduate Record Examination (GRE), Test of English as a Foreign Language (TOEFL), Michigan English Language Assessment Battery (MELAB), the Chartered Association of Certified Accountants Examination and the Association of International Accountants Examination.

1.2 Statement of the problem

Tests are generally used to provide information in making better decision. Although tests appear to be good in their use for decision making a problem of empirically justifying their usability is indispensable. That is, before the tests are used in making decisions it is necessary to know if they provide useful information. To maximize the accuracy of the decision the test which leads to the best decision should be selected. Then, how should the best test be selected? The decision mainly depends on the validity, reliability and other test characteristics.

Validity is one of the most important consideration in judging the adequacy of the test for making better decision. Before a test is ready for decision evidences of its validity must be determined.

Validation research are mainly dominated by criterion related validity (Anastasi, 1976; Mehrens and Lehmann, 1984; Shoenfeldt, 1984; Thorndike and Hagen, 1977). The criterion related validity, particularly predictive validity is the appropriate validity to be determined in personnel selection. Shefritz (1985) stated that predictive validity is the type of validity most strongly advocated. A test needs to have high predictive validity if it is used to predict persons who will be successful in a particular job because valid tests are excellent indicators of future performance.

A test may be constructed by experts consulting and strictly following test construction guidelines. Although the items of the test are perfectly constructed, in the sense of psychometric characteristics one cannot pass judgments on the adequacy of the test for predicting a certain performance unless its predictive validity is empirically determined. In this regard, Lindeman and Merenda (1979) argued that one does not proceed to construct a test with no guidelines whatsoever, but one should reserve judgment concerning the value of an instrument for a particular purpose until it has been determined as much as possible quite objectively and accurately.

At present predictive validity evidences for the secretarial tests administered by TMU are not available. For over two decades the psychological, educational and occupational (vocational) tests that were offered by the unit to its prospective consumers were hardly evaluated scientifically (Yusuf, 1989). Consequently the role and relevance of these tests, their validity and reliability were questioned. Similarly Jones (1973-2) indicated the absence of follow-up validation studies to determine the effectiveness of the test batteries that were in use then in the Testing Centre of the AAU.

Therefore, this study is aimed to answer the following research question: Are the secretarial tests for the senior secretary (G-I), secretary (G-II) and Junior Secretary (G-III) administered by TMU for AAU's use valid in predicting successful secretaries when validated against semi-annual supervisor's ratings.

In connection to the above research question the study also deals with the item analysis of the secretarial tests administered by the Unit. This item analysis is aimed at identifying items that are inconsistent with the total test or parts of the test.

1.3 The Need for the Study

The use of validated tests is one of the requirements of effective testing. Thorndike (1949) argued that whenever a test is being used for personnel selection for some job speciality, it is most desirable that it be validated empirically. A number of authors in the field of measurement today still strongly advocate the need for test validation in personnel selection. (Brown, 1983; Cronbach, 1984; Hopkins and Stanely, 1981; Shefritz, 1985; Sussman, 1986). Unless there is sufficient evidence on the predictive validity of a given selection test, it may be misused in predicting future performance.

Hence, the use of test results for the purpose of prediction without considering their predictive validity may lead to serious errors. As indicated by Shoenfeldt (1984) in the absence of formal validation, one would never know the extent to which the testing program was successful or superior to another assessment procedure. A testing program that does not involve validation research is at best unknown and at worst may be an outright fraud (Shoenfeldt, 1984).

If a test whose validity has not been demonstrated,

that is, test whose outcome does not empirically relate to its purpose, is being used it may be considered unsuitable for prediction. Tuckman (1975) maintained that such a test does a disservice to the test taker since conclusions based on it may reflect on some other characteristics of the test taker than the one that the test was intended to measure.

The battery of secretary test given by TMU for the various secretarial positions were as follows:

Secretarial Position	Test Battery
Copy Typist	1. English: Usage, Comprehension and Grammar 2. English Typing: Straight Copy and Revised Manuscripts 3. Amharic Typing: Straight Copy
G-III Secretary	1. Arithmetic 2. English: Usage, Comprehension and Grammar 3. English Typing: Straight Copy, Revised Manuscripts & Tables 4. Amharic Typing: Straight Copy and Revised Manuscripts
G-II Secretary	1. English Comprehension 2. English Typing: Straight Copy, Letters, Revised Manuscripts and Tables 3. Amharic Typing: Straight Copy and Revised Manuscripts
G-I Secretary	1. English Comprehension 2. English Typing: Straight Copy, Letters, Revised Manuscripts and Tables 3. Amharic Typing: Straight Copy and Revised Manuscripts 4. Shorthand 70 W/M: Letter I and letter II
Executive Secretary	1. English Comprehension 2. English Typing: Straight Copy, Letters Revised Manuscripts and Tables 3. Amharic Typing: Straight Copy and Revised Manuscripts 4. Shorthand 80 W/M: Letter I and letter II

The results of these tests are used for the selection and promotion of secretaries by AAU and other 106 government institutions such as: Commercial Bank of Ethiopia, Public Transport Corporation, Ethiopian Retail Trade Corporation, Foot wear and Leather Articles Enterprises and Ethiopian AirLines. However, at present empirical justifications of the degree of inferences to be made from these test results were hardly present (Yusuf, 1989). Without validity evidence no one knows whether the result of these tests accurately predict or not for the positions of the various secretarial jobs. Hence, the need to determine the predictive validity becomes the purpose of the present study.

In addition, predictive validity evidences could be practically useful not only for the test users but also for the test developers and takers. For the test developers validity data provides evidences about the quality of a test and its potential uses. The test takers could benefit from the validity evidences because the test results are the basis for the decisions that affect their lives.

In the item analysis, item difficulty, item discrimination, item correlation and item reliability

indices are to be computed. The data from item analysis helps to provide sug- gestions to the test writer in case poor items are identified.

1.4 The Scope of the Study

The subjects of the study were office secretaries of AAU. The secretarial position at AAU include: Copy Typist, G-III Secretary, G-II secretary, G-I secretary and Executive secretary. Copy Typist and Executive secretaries were not included in the study because the number of candidates sitting each year for these positions were few in number. In addition, G-III, G-II and G-I candidates from the colleges under AAU but outside Addis Ababa were not included for the same reason.

With respect to the criterion measures, there are a number of them that could be used in predictive validation research such as supervisor's ratings, peer ratings, self-assessment, production , proportions of spoilage in pieces produced and the like. In this study, the writer chose supervisory ratings as a criterion measure. Although each type of criterion measure has its own shortcomings, supervisory ratings usually suffer from common problems such

as halo-effect, and restriction in range (Meyer, 1987). The supervisory ratings used in the present study as a criterion measure inevitably suffer from these common problems.

1.5 Methods of the Study

Subject

The subjects of the study were all office secretaries in the AAU who took secretarial tests from September, 1982 to July, 1990 to be promoted or selected to the positions of G-III, G-II and G-I secretaries. All in all, from September, 1982 to July, 1990 there were 960, 818 and 464 G-III, G-II and G-I AAU secretarial Candidates, respectively who took secretarial tests at the TMU of the IER. Out of these candidates 102, 97 and 58 were either selected or promoted to the positions of G-III, G-II and G-I secretarial positions, respectively. These were the subjects of the study. All the subjects were females except 3 males in G-II and a male in G-I. The item analysis of the objective tests were based on 100 randomly selected candidates who took the secretarial tests from January, 1990 to December, 1990.

Materials

The materials were secretarial test scores obtained

from the records of the TMU of the IER and educational status, work experience, number of supervisory ratings from the test date to promotion, and supervisory ratings obtained from the personnel Department of the AAU (see appendices A, B, C). The test scores, educational status, work experience and number of supervisory ratings from test date to promotion formed the independent variables. The test scores for G-III secretary included: four objective type tests (Arithmetic, English Usage, English Comprehension and English Grammar), three English typing tests (straight copy, revised manuscript and tables) and two Amharic typing tests (straight copy and revised manuscript). The G-II and G-I test score materials contained: English Comprehension test scores, English typing test scores (Straight copy, letters, revised manuscript and tables) and Amharic typing test scores (straight copy and revised manuscript) In addition the materials for G-I contained Shorthand test scores (Letter I and Letter II).

The supervisory ratings were the dependent variables. These were last supervisory ratings before the tests were administered, the first and the second supervisory ratings subsequent to promotion for each position. These ratings are semi-annual performance evaluations done by various departments and collected form the files of each candidate

available in the Personnele Department of the AAU.

The performance evaluation form was prepared by the Central Personnel Agency (CPA) of the Ethiopian Government. The AAU government and the other government institutions use the form to evaluate, develop and promote government employees. So far, the researcher was not able to find a study conducted on the validity and reliability of the performance evaluation form. However, in the present study, the average performance ratings of supervisors by using the CPA performance evaluation form reported before the tests were administered and after promotion (to subsequent ratings) were used as the dependent variables.

It was decided to use three semi-annual performance ratings because secretaries are rated by different supervisors before and after promotion. This enables also to see the stability of the raters and arrive at proper conclusion.

Procedures

The general paradigm for investigating criterion related validity involves establishing the relationship between scores on the test (predictors) and the criterion.

The scores on the secretarial tests administered to secretarial candidates who were promoted or selected to the positions of G-III, G-II and G-I secretaries from September, 1982 to July, 1990 were collected from the records of the TMU of the IER. Other potential predictors including educational status - grade level of subjects, work experience - number of years served by the subjects being a secretary and number of supervisory ratings from the date of the test to the date of promotion or selection were collected from the records of the personnel Department of the AAU. The criterion data was also collected from the records of the Personnel Department of the AAU.

The data were analyzed by means of correlating the predictors with the criterion, and simple and multiple linear regression. The best predictors of the criterion were found using step-wise procedures. The statistical tests were carried out at 5% level of significance.

The items of the objective tests were analysed using item difficulty, item discrimination, the point biserial correlation and reliability indices. For the subjective tests the difficulty and discrimination indices were estimated using the formula proposed by Whitney and Sabers (cited in Mehrens and Lehmann, 1984).

APPENDIX - A

Predictors and Criterion Data for G - III

No.	Arithmetic (x_1)	English Usage (x_2)	English Comprehension (x_3)	English Grammar (x_4)	English Typing - Straight Copy (x_5)	English Typing - Revised Manuscript (x_6)	English Typing - Tables (x_7)	Amharic Typing - Straight Copy (x_8)	Amharic Typing - Revised Manuscript (x_9)	Work Experience in years (x_{10})	Educational Status (x_{11})	No of ratings from test date to promotion (x_{12})	Last ratings before test (y_1)	First ratings after promotion (y_2)	Second ratings after promotion (y_3)
1.	68	45	48	60	63	76	63	68	71	04.4	1	1	4.00	2.55	4.33
2.	44	45	50	60	63	68	56	59	70	10.5	1	1	-	4.13	4.13
3.	58	50	51	60	65	66	68	69	72	07.4	1	1	-	3.14	3.14
4.	55	50	61	51	53	65	64	57	51	06.1	2	3	3.63	4.22	4.10
5.	52	48	53	32	67	66	63	50	61	05.5	1	1	3.77	3.77	3.55
6.	44	34	48	35	66	68	61	66	72	09.3	1	1	4.50	4.20	4.67
7.	56	49	55	57	75	74	70	64	60	07.6	3	1	-	3.13	2.75
8.	47	35	28	39	57	63	53	66	71	05.4	1	1	-	4.00	3.88
9.	52	62	50	74	52	54	54	45	45	09.4	3	1	-	2.12	4.75
10.	56	65	68	70	69	72	50	51	68	08.3	1	1	-	3.35	2.63
11.	59	43	43	35	49	56	64	66	69	09.8	1	1	3.89	4.44	3.66
12.	52	43	38	44	66	67	63	59	70	09.9	1	1	3.54	3.78	3.67
13.	63	69	55	67	51	61	54	47	50	07.5	1	1	-	4.50	4.25
14.	63	43	53	50	66	64	56	66	74	10.9	3	1	3.67	3.67	4.56
15.	49	35	56	40	66	55	53	66	71	04.4	1	1	-	4.22	4.00
16.	58	52	61	47	49	64	48	59	54	05.6	1	1	-	4.00	3.75
17.	49	50	56	45	59	56	35	65	58	10.5	1	1	4.00	3.67	3.70
18.	73	62	58	69	62	54	68	59	60	11.9	3	1	3.88	3.63	3.63

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Y_1	Y_2	Y_3
19.	52	53	53	54	47	58	53	58	62	03.9	1	1	4.44	4.33	3.92
20.	76	65	55	54	57	67	64	68	72	03.9	1	1	4.44	4.44	4.44
21.	59	43	56	60	63	68	64	57	68	05.4	1	1	4.22	4.44	4.44
22.	65	65	35	75	44	35	45	58	60	06.1	1	2	3.44	3.43	3.95
23.	47	47	61	60	44	46	60	50	55	08.4	1	1	4.22	4.66	4.66
24.	54	45	48	56	63	65	58	63	70	10.2	1	1	4.50	4.92	4.89
25.	42	52	53	56	69	68	63	63	62	08.5	1	1	-	3.00	3.88
26.	57	63	53	64	53	63	63	50	58	09.0	1	1	4.25	4.25	4.37
27.	59	51	62	64	50	44	52	47	60	03.0	1	1	-	3.50	3.50
28.	59	45	48	54	49	67	63	57	72	08.1	1	1	4.77	4.55	3.33
29.	59	45	43	52	44	53	53	48	63	07.1	1	2	3.50	3.44	3.78
30.	59	53	53	56	52	55	56	50	62	05.0	3	3	4.44	4.22	4.11
31.	54	54	58	64	57	68	64	57	71	04.3	1	3	4.22	4.88	4.44
32.	59	53	48	60	52	58	53	44	50	06.1	2	1	3.66	3.55	3.66
33.	51	51	53	52	51	56	55	50	50	04.1	1	1	5.00	5.00	5.00
34.	49	54	57	57	51	53	56	56	50	07.4	1	2	4.00	3.90	2.67
35.	40	41	53	40	66	56	60	59	64	09.1	1	1	4.77	3.67	4.22
36.	52	58	50	55	53	63	63	57	68	09.5	1	1	3.89	2.78	4.00
37.	59	55	58	54	49	67	63	57	72	08.7	1	2	-	4.55	3.33
38.	50	51	56	59	55	59	60	54	62	07.5	1	1	-	2.50	3.75
39.	56	43	58	57	52	59	56	58	61	04.1	1	2	3.70	4.33	3.92
40.	52	50	53	56	59	56	63	54	64	09.3	1	1	3.75	3.78	4.00

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Y_1	Y_2	Y_3
41.	57	52	56	56	63	65	53	57	68	08.0	1	1	4.37	4.11	4.22
42.	50	55	59	59	49	53	61	50	64	05.6	2	1	4.22	4.22	3.77
43.	74	54	67	60	70	66	69	59	71	08.4	3	1	4.33	4.22	4.22
44.	57	48	58	50	59	53	53	50	61	11.4	3	1	3.78	3.78	3.77
45.	63	51	38	50	66	64	61	57	71	05.8	3	1	4.56	4.44	4.43
46.	57	60	60	52	53	53	52	54	63	04.4	1	1	4.00	3.70	3.11
47.	57	50	48	47	63	65	61	71	73	04.4	1	2	3.40	3.50	3.50
48.	50	54	53	49	49	56	56	57	62	12.3	1	1	4.25	4.44	4.44
49.	44	50	53	54	63	64	60	50	58	10.4	1	1	4.25	3.75	4.14
50.	61	46	61	72	66	67	58	57	62	07.6	2	2	3.56	3.70	3.70
51.	57	59	46	50	66	64	60	63	63	07.5	1	1	-	4.00	4.38
52.	54	53	53	72	63	71	64	63	70	07.0	1	1	-	3.50	3.60
53.	54	50	53	56	63	64	53	50	63	08.7	2	1	4.44	4.44	4.33
54.	57	58	58	55	53	41	58	54	51	09.1	3	1	4.00	4.44	4.67
55.	52	45	38	56	57	55	58	50	54	15.1	1	2	3.87	4.00	4.00
56.	57	63	56	36	53	67	64	48	54	10.1	1	1	3.60	4.00	4.00
57.	59	52	49	67	64	67	67	54	62	05.5	2	1	4.00	4.00	4.22
58.	52	56	51	72	65	58	60	59	63	02.4	1	1	-	4.56	4.56
59.	52	55	50	73	63	71	69	70	74	04.9	3	1	4.00	4.44	4.22
60.	61	50	48	64	44	65	56	68	58	04.4	3	1	4.33	4.00	3.77
61.	52	48	50	50	66	70	64	74	62	08.1	3	1	4.33	4.33	4.44
62.	47	45	53	45	57	64	56	68	68	12.4	1	1	4.33	4.00	4.00
63.	49	43	50	50	66	72	64	68	68	06.2	3	1	4.44	4.67	4.44
64.	50	54	44	50	59	55	56	57	61	09.0	1	3	4.00	3.89	2.89
65.	57	55	50	52	63	65	46	54	62	10.6	1	1	4.67	5.00	5.00

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Y_1	Y_2	Y_3
66.	50	45	56	50	63	68	63	54	64	09.9	1	2	4.00	3.11	3.22
67.	45	40	45	40	57	63	64	48	54	13.0	1	1	-	4.00	3.67
68.	50	61	56	49	45	50	60	54	62	01.5	1	1	-	4.00	3.67
69.	59	41	50	56	45	60	46	63	64	08.6	1	1	4.02	3.66	3.44
70.	59	53	57	65	50	64	71	68	50	06.5	1	1	-	2.89	3.44
71.	49	38	35	56	61	46	63	45	43	07.4	2	1	-	3.30	3.50
72.	26	30	46	47	57	56	60	54	56	07.4	2	1	-	3.70	3.70
73.	54	57	64	50	44	50	51	36	46	07.4	2	1	-	3.37	2.75
74.	59	52	62	35	63	67	63	59	70	10.1	1	1	4.80	4.44	4.66
75.	50	45	50	73	66	66	64	66	71	02.6	1	1	4.00	4.56	4.56
76.	63	53	56	64	63	48	63	57	64	03.3	1	1	4.00	4.56	-
77.	52	38	53	49	57	56	63	54	62	07.7	1	1	4.50	4.50	4.50
78.	50	66	60	64	53	45	63	50	61	03.2	1	2	3.88	4.44	4.56
79.	44	47	50	50	53	58	58	54	58	08.2	1	2	3.90	2.67	3.67
80.	50	50	48	45	66	49	60	57	64	07.7	1	1	4.11	4.11	4.00
81.	64	54	53	60	50	53	60	50	54	01.4	2	1	-	3.73	4.33
82.	50	48	34	52	53	56	58	57	63	08.9	1	1	3.89	4.00	4.56
83.	44	41	28	40	57	56	53	63	64	10.1	1	2	3.77	4.88	3.89
84.	56	37	46	35	53	68	61	71	74	04.4	2	1	-	5.00	5.00
85.	57	41	38	60	53	58	60	54	50	10.6	1	1	3.56	4.00	3.78
86.	54	51	28	40	57	56	53	63	64	06.1	1	1	4.75	4.75	4.00
87.	45	35	65	56	56	54	86	36	63	04.4	1	1	4.87	4.63	3.32
88.	54	50	50	57	52	46	58	54	54	07.1	2	1	4.22	3.55	3.56
89.	57	50	46	56	57	71	64	50	62	04.1	2	1	4.44	4.44	4.67
90.	52	51	50	53	55	59	58	54	51	06.2	1	1	3.67	3.77	3.82

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Y_1	Y_2	Y_3
91.	52	50	50	62	50	52	60	58	60	09.8	1	1	4.44	4.44	4.49
92.	59	47	51	63	57	56	56	57	63	05.1	3	1	4.63	4.50	4.53
93.	57	57	54	68	52	44	46	63	63	07.6	1	2	4.20	4.25	4.50
94.	57	50	50	51	50	58	56	54	50	11.1	1	2	3.13	3.40	3.85
95.	74	52	64	63	42	48	58	68	51	08.8	1	1	3.54	3.68	3.78
96.	57	51	53	64	67	42	50	70	49	10.5	1	1	3.89	4.00	4.12
97.	68	53	46	66	53	56	42	50	51	09.9	1	1	4.00	4.33	4.56
98.	54	56	46	63	53	63	53	59	62	06.1	1	1	3.88	4.00	4.06
99.	64	53	43	74	62	50	48	50	54	05.2	3	1	4.25	4.26	4.30
100.	61	57	56	73	74	53	53	58	58	06.6	1	1	3.90	3.80	4.00
101.	57	57	61	60	54	56	60	50	55	05.4	1	1	-	2.40	3.00
102.	57	58	58	54	64	69	50	68	58	04.4	1	2	3.78	4.56	4.78

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y_1	Y_2	Y_3
63.	57	55	53	54	52	55	59	08.2	1	1	5.00	5.00	5.00
64.	48	58	57	42	52	55	56	13.2	1	1	4.22	4.22	4.22
65.	57	54	67	58	74	50	51	07.5	2	1	3.67	3.11	3.22
66.	44	54	53	56	66	79	53	07.4	2	1	4.11	4.67	4.67
67.	51	52	53	47	52	55	56	10.0	1	1	4.89	4.44	4.33
68.	50	50	55	57	68	57	56	05.6	3	1	4.66	4.66	4.20
69.	54	52	54	62	67	55	56	06.1	1	1	4.00	3.88	4.22
70.	45	52	60	59	53	49	55	04.5	1	1	4.00	4.33	4.45
71.	34	50	58	48	68	50	56	13.0	1	1	4.00	4.33	4.22
72.	55	52	45	45	53	48	54	07.8	1	1	3.67	4.00	4.00
73.	41	46	51	54	66	57	56	07.7	2	3	4.44	4.44	4.33
74.	31	55	55	57	68	58	69	07.6	2	1	3.89	4.11	4.11
75.	57	46	50	50	66	49	46	13.7	1	3	3.33	3.33	3.46
76.	55	54	45	58	57	57	61	10.2	2	1	4.67	4.08	3.90
77.	57	58	45	55	50	54	49	13.0	2	2	4.00	4.30	4.20
78.	41	66	56	57	50	51	55	23.7	3	1	3.60	4.40	4.25
79.	57	52	55	55	47	55	53	11.1	1	1	4.56	4.55	4.44
80.	57	48	43	57	66	60	69	13.2	2	3	4.77	4.78	4.33
81.	47	58	47	64	63	57	66	15.1	1	1	4.67	4.67	3.00

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y_1	Y_2	Y_3
82.	57	56	47	54	57	57	50	21.7	2	2	3.00	4.67	4.67
83.	51	69	50	69	79	57	66	06.4	3	1	-	4.50	4.75
84.	54	57	59	61	70	57	56	09.4	1	1	4.75	4.50	3.80
85.	57	50	51	51	54	55	49	21.1	1	3	4.87	4.87	4.87
86.	62	50	48	60	47	49	53	12.8	1	1	3.57	4.12	4.62
87.	45	44	46	53	47	51	56	12.3	2	1	4.40	2.75	3.66
88.	55	46	43	58	54	60	62	12.3	2	2	4.00	4.90	3.00
89.	60	49	45	61	60	53	57	12.0	3	1	4.80	4.90	4.22
90.	54	55	56	61	50	49	53	13.4	2	1	4.30	4.56	4.78
91.	58	58	53	57	57	49	56	13.4	2	1	4.00	4.00	4.00
92.	41	50	50	53	58	55	49	11.1	1	1	3.90	4.11	4.56
93.	54	51	50	51	46	55	52	11.5	1	1	4.38	4.12	4.78
94.	57	44	36	56	76	59	57	12.0	3	3	4.11	4.44	4.78
95.	45	50	52	54	66	59	65	13.9	1	1	3.89	3.67	3.44
96.	50	50	43	54	67	54	56	12.5	2	1	4.71	4.67	4.33
97.	57	49	62	54	68	55	69	12.3	2	3	4.00	4.66	4.66

APPENDIX - B

Predictors and Criterion Data for - G-II

No.	English Comprehension (X_1)	English Typing - Straight Copy (X_2)	English Typing - Letters (X_3)	English Typing - Revised Manuscript (X_4)	English Typing - Tables (X_5)	Amharic Typing - Straight Copy (X_6)	Amharic Typing - Revised Manuscript (X_7)	Work Experience in years (X_8)	Educational Status (X_9)	No of ratings from test date to promotion (X_{10})	Last ratings before test (Y_1)	First ratings after Promotion (Y_2)	Second ratings after promotion (Y_3)
1.	62	48	56	49	53	55	56	08.4	1	1	4.67	4.67	-
2.	54	51	53	64	67	59	59	12.4	1	1	4.11	4.22	-
3.	47	56	46	51	52	51	56	10.0	1	1	4.11	4.00	-
4.	45	39	24	53	52	57	59	08.4	1	1	3.56	3.78	-
5.	37	41	33	57	71	41	50	09.3	1	1	4.00	4.00	-
6.	51	48	40	51	71	49	50	07.4	1	1	4.22	3.78	-
7.	45	52	51	44	50	66	85	08.4	1	1	4.11	4.77	-
8.	47	54	43	61	74	54	56	10.2	1	1	4.44	4.78	-
9.	41	54	41	64	50	58	52	07.5	1	1	3.77	4.33	-
10.	47	46	50	59	69	47	56	05.5	1	1	4.11	4.00	-
11.	47	51	54	58	67	53	54	06.2	2	1	4.40	4.44	-
12.	50	50	33	57	50	49	58	09.9	1	1	4.44	4.33	-
13.	54	66	51	51	47	44	58	10.9	3	1	4.89	4.89	4.89
14.	55	49	21	55	45	54	44	09.5	2	1	4.11	4.11	4.00
15.	45	58	45	58	51	50	59	03.1	1	1	3.88	4.00	-

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y_1	Y_2	Y_3
16.	47	56	45	61	62	53	54	07.6	1	1	4.78	4.78	4.78
17.	55	51	36	59	78	66	66	07.8	3	1	4.78	4.78	4.78
18.	47	50	46	60	69	51	58	09.3	1	1	3.88	4.22	–
19.	54	54	44	45	68	47	38	10.5	1	1	3.56	3.67	–
20.	45	48	53	56	67	43	50	12.3	1	2	4.78	4.67	–
21.	47	41	52	46	46	40	49	09.1	3	1	3.78	4.67	–
22.	45	50	55	57	52	37	44	10.1	1	1	5.00	4.58	–
23.	57	66	56	66	50	60	57	06.1	3	1	4.67	4.78	–
24.	28	54	55	39	47	51	56	05.8	3	1	4.56	4.89	4.89
25.	57	52	40	49	54	53	56	04.4	1	1	4.17	3.56	3.74
26.	54	46	55	59	47	51	56	07.1	2	3	4.44	4.22	4.22
27.	54	44	40	63	76	57	56	04.4	1	1	4.77	4.88	4.89
28.	57	59	54	44	53	55	54	08.8	2	3	4.67	4.67	4.67
29.	28	50	43	54	52	56	69	10.9	3	3	4.85	4.56	4.44
30.	68	68	56	53	54	53	56	16.3	1	3	5.00	5.00	5.00
31.	37	59	50	49	67	58	69	07.5	2	2	4.56	4.56	4.55
32.	54	61	55	62	74	51	56	07.7	2	3	4.50	4.17	5.00
33.	47	50	35	59	52	55	56	11.1	2	1	4.11	4.00	4.89
34.	51	51	43	59	51	64	67	09.0	1	2	3.87	4.67	4.78
35.	44	59	56	54	50	50	56	07.5	1	1	4.56	4.67	4.76
36.	54	58	50	58	67	59	55	07.5	2	2	4.89	4.90	4.90
37.	44	35	44	75	76	57	59	06.6	1	3	3.55	4.11	4.14
38.	52	58	53	66	64	54	60	08.0	2	2	4.88	4.88	4.90

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y_1	Y_2	Y_3
39.	47	44	36	49	76	57	54	07.5	1	4	2.67	4.00	4.00
40.	25	56	61	66	68	61	69	06.6	1	1	4.67	4.22	4.55
41.	57	50	43	54	67	51	56	09.5	2	2	4.78	4.67	4.55
42.	50	59	66	37	52	41	56	07.5	2	4	4.00	3.33	3.33
43.	54	64	53	56	53	49	56	06.8	2	1	4.10	4.44	4.56
44.	57	54	40	46	53	51	56	09.4	1	2	4.55	3.66	3.78
45.	54	58	50	58	67	49	55	07.6	2	2	4.39	4.50	4.11
46.	45	55	37	54	76	56	52	10.5	2	1	3.89	3.78	4.78
47.	57	51	50	48	67	50	50	09.0	1	1	5.00	4.78	4.78
48.	54	54	45	59	71	51	55	10.6	1	1	4.89	4.89	4.90
49.	45	46	50	50	68	46	50	06.1	1	1	3.89	4.22	4.56
50.	48	46	52	57	52	51	56	13.2	1	2	4.80	4.67	4.89
51.	57	48	37	54	56	60	65	21.5	1	4	3.56	4.66	3.78
52.	51	50	56	58	51	59	49	21.6	1	1	4.70	4.80	4.80
53.	50	68	50	59	67	60	56	09.2	1	1	4.50	4.80	4.56
54.	55	54	58	45	66	53	56	15.8	1	2	4.25	4.63	4.75
55.	45	52	54	57	47	61	76	12.7	1	1	4.75	3.89	4.44
56.	57	50	45	59	52	55	56	13.2	1	1	3.90	4.44	4.55
57.	47	54	56	60	68	53	56	15.3	1	1	4.39	3.44	3.39
58.	57	48	58	53	68	49	52	25.0	1	2	4.60	4.22	4.60
59.	47	50	53	56	68	56	58	16.4	1	2	4.44	4.20	4.66
60.	55	54	58	45	66	53	56	07.0	1	2	4.77	4.33	4.10
61.	57	55	52	48	68	47	51	10.2	2	1	5.00	4.67	4.67
62.	37	54	55	51	70	55	61	08.0	1	4	4.22	4.33	4.10

APPENDIX - C

Predictors and Criterion data for G - I

No.	English Comprehension (x_1)	English Typing - Staight Copy (x_2)	English Typing - Letters (x_3)	English Typing - Revised Manuscript (x_4)	English Typing - Table (x_5)	Amharic Typing - Straight Copy (x_6)	Amharic Typing - Revised Manuscript (x_7)	Shorthand - Letter I (x_8)	Shorthand - Letter II (x_9)	Work Experience in years (x_{10})	Educational Status (x_{11})	No of ratings from test date to promotion (x_{12})	Last supervisory ratings before test (y_1)	First supervisory ratings after promotion (y_2)	Second supervisory ratings after promotion (y_3)
1.	57	52	56	51	66	48	45	45	48	11.2	1	1	4.44	4.78	-
2.	47	41	61	38	50	54	44	40	50	21.5	1	1	3.89	4.22	-
3.	45	54	62	51	69	52	50	40	50	13.3	1	1	4.78	4.78	-
4.	54	52	56	48	68	51	50	47	48	16.4	1	1	4.78	4.78	-
5.	37	52	57	40	52	43	36	37	48	12.3	2	1	4.22	4.33	-
6.	54	50	59	59	68	37	44	44	42	11.1	1	1	4.78	4.56	-
7.	47	54	52	53	70	41	47	46	48	12.3	2	1	4.66	4.89	-
8.	57	48	48	61	80	42	51	40	50	23.7	3	1	4.22	4.83	-
9.	50	55	54	51	69	41	45	47	48	13.0	1	1	4.00	4.83	-
10.	60	48	43	46	67	42	48	38	20	11.5	1	4	3.63	4.33	3.67
11.	47	59	71	80	71	59	77	55	52	19.4	1	1	4.50	4.50	4.50
12.	57	50	48	44	66	42	46	43	44	12.4	3	4	4.44	4.44	4.56
13.	47	54	50	64	74	50	45	52	54	16.3	2	3	4.89	4.78	4.33
14.	57	55	52	58	68	51	58	53	67	16.4	1	3	4.13	4.00	4.33
15.	47	56	71	58	68	54	69	37	46	14.5	3	6	4.31	4.56	4.40
16.	57	56	53	64	74	53	62	53	44	23.2	1	1	5.00	5.00	4.50

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Y_1	Y_2	Y_3
17.	41	48	37	47	52	28	44	34	47	10.2	1	1	4.67	4.67	-
18.	54	50	57	77	68	62	50	46	57	11.5	2	1	4.44	4.33	-
19.	62	54	40	74	66	38	43	34	45	07.4	2	1	4.67	4.67	-
20.	50	44	43	48	41	36	42	39	50	12.0	2	1	4.44	4.38	-
21.	62	48	53	79	50	43	51	44	50	01.3	2	1	4.33	4.44	-
22.	54	63	40	51	50	38	37	40	48	11.1	1	1	4.44	4.44	-
23.	57	56	52	77	68	41	51	28	47	07.6	2	1	4.22	4.22	4.44
24.	57	52	48	53	67	43	48	43	57	19.9	2.	4	4.00	4.29	4.22
25.	37	50	50	56	71	53	60	50	44	29.7	3	5	4.39	3.89	4.22
26.	56	49	55	57	65	64	50	54	60	21.3	1	3	4.66	4.00	4.22
27.	50	48	43	54	60	62	66	51	56	17.0	1	2	5.00	5.00	5.00
28.	50	50	56	56	73	51	62	37	66	14.9	1	2	4.60	4.00	3.56
29.	50	52	66	59	76	45	56	47	44	21.5	2	3	4.50	4.78	4.78
30.	45	59	66	77	67	66	62	56	67	23.6	1	1	4.25	3.90	4.20
31.	62	50	59	59	74	53	56	50	61	20.8	1	4	4.86	4.88	4.89
32.	53	54	50	61	55	55	58	56	57	21.5	2	2	4.50	4.75	5.00
33.	47	54	46	51	53	41	47	38	44	21.3	1	1	4.78	4.78	4.78
34.	41	56	51	54	67	48	56	58	44	13.3	1	1	4.89	4.55	4.44
35.	34	56	51	49	71	53	56	38	61	06.8	1	2	4.22	4.44	4.46
36.	54	56	53	53	71	53	54	50	57	12.1	3	2	4.11	4.56	4.70
37.	54	52	64	58	79	50	59	51	57	12.1	1	2	4.67	5.00	4.89
38.	66	56	53	48	70	55	64	59	70	09.3	3	3	4.11	4.11	4.10
39.	37	56	62	48	54	42	46	38	44	09.0	1	2	3.67	3.78	3.89
40.	64	55	52	58	76	41	50	48	44	06.5	3	1	4.74	4.52	4.87

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Y_1	Y_2	Y_3
41.	60	55	48	58	71	48	56	50	55	17.8	3	5	4.55	4.66	4.78
42.	45	50	51	48	52	41	50	51	59	21.3	2	4	4.56	4.56	4.56
43.	50	50	51	48	68	43	41	48	70	12.8	2	3	4.89	4.22	4.22
44.	47	48	52	58	76	41	50	48	44	06.5	3	4	5.00	4.89	4.89
45.	50	54	51	45	52	43	59	43	50	09.5	1	1	4.61	4.78	4.89
46.	57	55	53	55	50	41	45	53	70	28.9	1	4	5.00	4.77	4.89
47.	68	50	60	54	57	68	63	57	64	16.3	1	2	5.00	4.00	4.13
48.	53	60	56	45	48	59	55	50	51	27.0	3	2	4.55	4.44	4.88
49.	54	50	51	70	42	47	58	38	66	24.5	1	4	3.78	4.11	3.83
50.	63	56	54	60	50	51	59	62	67	12.7	1	4	4.00	4.00	4.13
51.	50	55	44	58	67	50	66	59	60	19.0	3	3	4.75	4.63	4.44
52.	47	52	50	46	68	51	55	37	46	19.0	2	2	4.75	4.50	4.66
53.	45	48	50	48	67	43	50	53	55	17.1	3	4	4.33	4.00	4.20
54.	45	48	53	58	54	44	45	33	41	27.6	1	4	4.90	4.85	4.85
55.	47	56	59	59	53	44	48	53	50	14.3	3	1	4.37	4.75	4.13
56.	55	56	62	64	73	53	62	43	50	16.8	1	3	4.05	4.62	4.47
57.	54	55	54	48	67	48	46	53	50	19.1	2	3	4.00	5.00	4.89
58.	45	52	51	60	50	45	55	53	44	16.6	1	3	4.11	4.89	5.00

2. BASIC THEORETICAL ASPECTS OF VALIDATION RESEARCH

2.1 Types of Validity Analysis

When Tests are used in making decisions it is necessary to know whether or not they provide useful information. The decision partly depends on the validity of a test. Thorndike and Hagen (1977) stated that validity is among many considerations entering into the evaluation of a test. The first and foremost question to be asked with respect to any testing procedure is "How valid is it?"

Validity refers to the degree to which the test provides information that is relevant to decisions which are to be made. The definition of validity in the current standards of Educational and Psychological Tests (Cited in Lawshe, 1985) runs as follows "Validity refers to the appropriateness of inferences from test scores or other forms of assessment." In discussing the concept of validity, Lawshe argued that a number of professional psychologists persistently refer to the validity of tests when, in fact, it is not tests which are valid or invalid but rather inferences from test scores. On the other hand,

Cronbach (1984) used interchangeably the two concepts: the validity of a test and interpretations from the test scores. He explained that whether it is called validity of a test or validity of inferences drawn from test scores researchers refer validity to the extent to which test results serve the particular uses for which they are intended.

In summary, validity is the degree to which inferences from test scores or from the results of other assessments are justified on the basis of empirical evidence. Hence, validity analysis is a procedure whereby researchers collect data to determine the accuracy of inferences that can be made from test scores. In the technical vocabulary of employment testing, to validate is to investigate, to conduct research and to determine the use of a test (Shefritz, 1985).

The majority of the psychological literatures discuss three types of validity (content, criterion - related and construct). Nevertheless, Lawshe (1985) argued that talking about types of validity is misleading and particularly troublesome in the field of employee selection. When a candidate takes a personnel test for employment or

promotion, the result is a sample of that person's behaviour to the extent that the test satisfies psychometric standards and is used in accordance with professionally accepted procedures (Lawshe, 1985).

In this regard, according to Lawshe (1985) most inferences made in the employment context fall into one of the three broad categories each of which requires its own unique analytical strategy to determine the validity of the inference.

First, if the purpose is to infer how well a candidate will perform on the job, i.e., to predict future behaviour, a criterion - related strategy is indicated. Here the mathematical relationship between test scores and some numerical index of job success is to be determined. Second, if the wish is to infer the extent to which a candidate currently possesses a relatively simple proficiency, that is, a component of the job or knowledge required to perform the job, thus to evaluate the present job competence, a content validity analysis is indicated. Third, if the wish

is to infer the degree to which the candidate currently possesses a trait or other characteristic (usually a psychological construct) critical to job performance, thus to assess an attribute, a construct validity analysis is indicated (Lawshe, 1985).

Among the three categories of validity inferences the criterion - related validity strategy has dominated validation research (Shoenfeldt, 1984). The criterion - related validity strategy refers to the empirical techniques of studying the relationship between the test scores or other measures (Predictors) and some independent external measures (Criteria) (Mehrens and Lehmann, 1984). Shoenfeldt (1984) listed nine tenets of criterion - related validity research which are summarized as follows:

1. The purpose of validation research is to predict future performance.
2. Predictors and criterion should be selected on the basis of job analysis.
3. Measuring instruments must be standardized.
4. Tests should be empirically validated.
5. Validation is situation specific.

6. More than one test should be used.
7. Only one criterion should be used.
8. Tests are preferred over non-test assessments.
9. Individual differences should be recognized in evaluating tests.

Most writers make a distinction between two kinds of criterion - related validity: Predictive validity and concurrent validity (Anastasi, 1976; Brown, 1983; Mehrens and Lehmann, 1984; Thorndike and Hagen, 1977). For instance, Mehrens and Lehmann (1984) discussed the logical and procedural distinction between predictive validity and concurrent validity. The logical distinction between the two criterion related validity pertains to the purpose of testing or the inferences we wish to make . In the aspect of predictive validity the concern is with the usefulness of the test score in predicting some future performance, whereas in concurrent validity the concern is with substituting the sore with some less efficient way of gathering criterion data. The procedural distinction is based on the time period when the criterion data are gathered. When the criterion data are collected at the same time with the test data we speak of concurrent validity and when the criterion data are gathered at a later date, we

validity involves the assessment of individuals followed by the collection of criterion information at some subsequent time (Shoenfeldt, 1984).

Decisions involving selection are predictive in nature. According to Gronlund (1976) predictive validity is of special significance in all types of aptitude testing, but it is pertinent whenever test results are used to make specific predictions. In agreement with Gronlund, Anastasi (1976) contended that the information provided by predictive validity is almost relevant to tests used in the selection and classification of personnel. For instance, hiring job applicants, selecting students for admission to colleges or professional schools and assigning military personnel to occupational training programs represent examples of the sort of decisions requiring knowledge of the predictive validity of tests.

In the main, the value of tests has been judged by their power to predict (Shelley and Cohen, 1986). In personnel promotion and selection tests are used as part of a decision making process. Brown (1983) stated that implicit in the concept of predictive validity is the idea that tests are used as a part of a decision making process. When a test is administered to a job applicant the decision

is whether he or she should be hired. Such a decision is influenced by the individuals test score.

In conclusion, validation research is directed to the investigation of inference form test results about the success of the test taker at present or in the future. Predictive validity study is the investigation of inferences from test results about how well the candidate will perform on the future job. In the selection decisions, the use of a particular test will lead to more correct decisions than simply choosing the candidates by other means or at random (Nitko, 1983).

2.2 Methods of Expressing Test Validation

The common methods of expressing predictive validity studies are the Pearson Product Moment Correlation, the simple and multiple regression analysis and the expectancy table (Nitko, 1983). Among these methods the Pearson Product Moment Correlation coefficient (r) is probably the most common method used in reporting validity studies (Mehrens and Lehmann, 1984). Predictive validity is expressed in terms of the correlation coefficient between the predictor variable and the criterion. Such a coe - fficient is called a validity coefficient (Lindeman and

Merenda, 1979). A validity coefficient is represented by the symbol r_{xy}, which stands for the correlation between predictor (x) and criterion (y). The basic procedure in determining Validity coefficient is to give a test to a group who are entering some job or training program, follow them up later, get for each one a specified criterion measure of success on the job or in the training program and then to compute the correlation between test score and criterion measure of success. The higher the correlation the more effective the test as a predictor (Brown, 1983).

The square of the validity coefficient, r^2_{xy} is called coefficient of determination. It is the amount of variance accounted for (predicted) in the criterion from knowledge of predictor test scores (Mehrens and Lehmann, 1984). The standard error used to set confidence limits about an estimated criterion score is

$$S_{y.x} = S_y \sqrt{1-r^2_{xy}}$$

When S_y = criterion standard deviation

The equation commonly called a regression equation, used to estimate the criterion score (y) is

$$Y' = r_{xy} \left(\frac{Sy}{Sx} \right) \left(X-\bar{X} \right) + \bar{Y}$$

Where r_{xy} = Correlation between the predictor and

the criterion

s_y = The criterion standard deviation

s_x = The predictor standard deviation

x = The predictor score

y = The criterion score

$\bar{x}$ = Mean of the predictor score

$\bar{y}$ = Mean of the criterion score

In a traditional validity study the main tool of analysis is the calculation of the multiple regression equation. Multiple regression is a procedure for determining weights for each predictor so that the standard deviation of the errors of prediction will be a minimum for the group studied (Schrader, 1971). Multiple regression weights are commonly expressed by b-weights and B (beta) - weights. The b- weights give the actual numbers that would be used in calcuating predicted values from predictor data. The B-weights give the weights that would be appropriate if each predictor were expressed in terms of standard scores with a mean of 0 and standard deviation of 1 for the group studied (Schrader, 1971).

Regression analysis also could yield correlation coefficients between each pair of variables. The coefficients relating each predictor to the criterion are

the familiar single-predictor validity coefficients. They are useful for evaluating predictors and for comparing results from one study to another (Schrader, 1977).

In summary, the common methods for reporting validity studies are the Pearson Product Moment Correlation Coefficient and both simple and multiple regression analysis.

2.3 Criteria in Predictive Validity Studies

Dunnettee (1963) defined a criterion as a standard which can be used as a yardstick for measuring employees success or failure. Theoretically, for any job there is some ultimate or true criterion or job standard (Cronbach, 1984). Shefritz (1985) defined a criterion as a measure of job performance or other work related behaviour against which performance on a test or other predictor measure is compared.

A number of criteria have been used in validation studies. According to Lawshe and Bolma (Cited in Nitko, 1983) personnel classification and selection research in government and industry have tended to use the four general types: production (quantity and quality of goods, sales),

personal data (accidents on the job, length of service, group memberships, training course grades), sample of job performance and judgements by others (checklists, supervisory ratings).

Among the various types of criteria used in validation studies the supervisory ratings are the most frequently used measure for both research and personnel decision making purposes (Meyer, 1987; Orpern, 1973). The criterion used to predict performance or success may be a representative for actual performance. Glueck (1982) argued that since good job performance is usually a combination of many things (quality of work, quantity of work, and so on), a criterion such as a supervisor's rating is a proxy for the real measure job success or performance. In personnel work, in business and industry where specialized tests are used to select individuals for specific jobs it is possible, indeed essential, to use actual production records or performance ratings as a criteria of a test validity.

Although industrial predictors are validated against supervisory ratings, these ratings are not entirely satisfactory as a criteria (Cronbach, 1984). Often a rating reflects the personal relation between the worker and the supervisor rather than the quality of work. Cronbach (1984)

stated that when a test fails to predict it is hard to locate whether the fault lies with a criterion or with a test. It is a problem always to chose the most satisfactory criterion. Any way, as stated by Cronbach (1984) a measure is a criterion only when an audience accepts it as important and relevant.

There are certain qualities desired in a criterion measure. These qualities are: relevance, freedom from bias, reliability and availability. (Annastasi, 1976; Brown, 1982; Lindeman and Merenda, 1979, Thorndike and Hagen, 1977). Thorndike and Hagen (1977) wrote, we rely upon professional judgment to provide the appraisal of the degree to which some available criterion measure is relevant to the ultimate criterion of success in an educational program or in a job.

By freedom from bias it means that the measure should be one on which each person has the same opportunity to make a good score (Thorndike and Hagen, 1977). Anastasi (1976) wrote, variation in generosity of ratings by the supervisors resulted from the knowledge of test scores may be an example of biasing factors. This source of error in test validation would obviously raise the correlation between test scores and criterion in a manner that is entirely superious or

artificial.

To be reliable a criterion measure should be stable or consistent from day to day or from time to time (Lindeman and Merenda, 1979). If a criterion score is one that jumps around from day to day so that the person who receives a high rating from one supervisor gets a low rating from another, then, there is no possibility of finding a test that will predict that rating (Thorndike and Hagen, 1977).

Finally, in the choice of the criterion measure a researcher encounters practical problems of convenience and availability (Thorndike and Hagen, 1977). In choosing the criterion measure the researcher should consider such factors as the time and the expense required to obtain the measure (Lindeman and Merenda, 1979).

2.4 Reliability of Tests

With respect to any test, one should raise at least two basic questions. The first is how accurately the domain defined by the test represents the domain in which we are

really interested. This is the issue of validity which is absolutely central to any given use of a testing procedure (Thorndike, 1985). The second question is how accurately and precisely the test score assesses the domain from which the test does in fact draw a sample. This estimates of precision determines the tests reliability (Thorndike, 1985).

Precision can be seen in absolute or relative terms. Relative precision asks the question: How accurately does the test reflect the individuals standing in the group? The answer is typically expressed as a correlation coefficient. This statistic indicates the extent to which individuals in a group maintain relatively consistent positions (scores) when two sets of measures are obtained and correlated, using the same test or its equivalent forms (Freeman, 1962). In relative precision, Correlational indices of different precision are used for comparing different tests when scores are reported on scales that are not directly comparable, but data are available for the same or comparable groups (Thorndike, 1985).

Absolute precision asks the question: How much can the individual be expected to vary in terms of some meaningful score scale from one testing to another. The answer to this question is typically expressed as a standard error of

measurement (SEM), indicating the standard deviation of a person's hypothetical series of measures.

This two - fold purpose of reliability coefficients is reflected in the several methods which have been developed for estimating reliability. Methods which provide reliability estimates based on a single testing offer evidence as to the precision of the test itself. These include internal consistency estimates such as those obtained by the use of the Kuder-Richardson when the test is given only once (Harold, 1952).

2.5 Item Analysis

Item analysis is a procedure by which individual item performance by a group of test takers is compared to their performance on the total test (Tuckman, 1975). Analysis of the relationship between item scores and total test scores very often reveals items that are inconsistent with the total test or parts of it. Therefore, the purpose of item analysis is to make each item consistent with the total test (Tuckman, 1975).

Ever since Binet and Simon in 1966 plotted the proportion of correct response to an item as a function of age, item analysis has been an important component in the field of educational measurement (Baker, 1977). Since then

two basic theoretical models, the classical psychometric and the item characteristic curve have been developed and a wide variety of item analysis procedures have been created (Baker, 1977).

However, most practioners of item analysis use the classical psychometric model in which item difficulty and the item - criterion correlation are used to describe the item (Baker, 1977).

In the classical psychometric model the simplest and most commonly used measure of item difficulty is the proportion of correct answers on the item (P). The desired difficulty level of items depends on the purpose of a test. For tests intended for selection, the bulk of the items must be at a difficulty level where selection will be made (Henrysson, 1970). The point - biserial (r_{pbis}) is used as an item discrimination index. It has the formula.

$$r_{pbis} = \frac{M_R - M_x}{S_T} \sqrt{P/q}$$

Where: M_R = Mean criterion score for the subjects choosing the right answer

M_x = Mean criterion score for all subjects

S_t = Standard deviation of criterion scores for all subjects

P = Proportion choosing the right answer

q = 1 - P

Guilford (cited in Henrysson, 1970) argued that the point-biserial correlation tells more about the contribution from the particular item to the predictive validity of the total test than does the biserial correlation. This is true in the sense that the point biserial correlation will tend to favour items of average difficulty (Henrysson, 1970).

Most people think of item analysis only in terms of objective tests (Mehrens and Lehmann, 1984). But it is just as important, if not more important, for essay test to have good questions, considering that they have so few questions (Mehrens and Lehamnn, 1984). Whitney and Sabers (cited in Mehrens and Lehman, 1984) proposed the following approach to compute the difficulty and discrimination indices of essay items.

1. Identify the upper and lower 25% of the group tested
2. Compute the sum of scores for the "highs" and

"lows" for each item

3. Apply the formula

$$\text{disc} = \frac{\Sigma_H - \Sigma_L}{N(\text{Score}_{max} - \text{Score}_{min})}$$

$$\text{diff} = \frac{\Sigma_H + \Sigma_L - (2N\ \text{Score}_{min})}{2N\ (\text{Score}_{max} - \text{Score}_{min})}$$

Where : Σ_H = Sum of scores for upper 25%

Σ_L = Sum of scores for lower 25%

N = 25% of numbers tested

Score_{max} = Highest possible score on the question

Score_{min} = Lowest possible score on the question.

The size of N bias the results of the difficulty and discrimination values. As the size of N increases the difficulty and discrimination values decreases. For instance, if the size of N is doubled the difficulty and discrimination values would be decreased by half. However, Whitney and Sabers (cited in Mehrens and Lehmann, 1984) did not offer any information regarding the size of N.

To conclude, item analysis is an important tool to improve the effectiveness of test items. In order to have good questions item analysis is equally important for

objective and essay tests.

As it has been already discussed validity, reliability and item analysis are closely linked. As stated by Brown (1983) item analysis helps to select items of appropriate difficulty and discrimination. Since discriminating items correlate positively with the total test score and validity refers to the relationship between test scores and some other variables or constructs, consequently reliability of a test a necessary condition for validity.

3. REVIEW OF LITERATURE

Lord, Cowles and Cynamon (1950) designed a battery of seven objective tests (Pre-Engineering Inventory) to assist the selection of students who would succeed in engineering schools. The tests included: passages requiring interpretation, problems for solution, diagrams to comprehend and questions on specific information. The test was administered to random samples ranging from 98 to 181 individuals. Correlation between half test scores gave a reliability coefficient of 0.75 - 0.94. The validity coefficient ranging from 0.14 - 0.68 were obtained by correlating the Pre-Engineering Inventory scores with first term average grades. It was found that the Pre-Engineering Inventory composite score was in general the most predictive of engineering college grades.

Ghiselli & Barthol (1953) summarized the results of personality tests and inventories used as aids in the selection of employees so that the usefulness of personality inventories can be more easily assessed. He reviewed related literatures from 1919-1953 and found that the mean validity coefficients of personality inventories for various occupational groups was between 0.14 - 0.36. He concluded, personality inventories on the average did have reasonably good predictive value in selecting clerical workers and salesmen.

Springer (1953) compared the ratings made by two supervisory personnels representing two levels of supervision over the 100 candidates and by three coworkers who work closely with the 100 leadsman jobs. Each candidate was evaluated on the basis of his work experience, education, work record, scores on mental ability and job knowledge tests. The Pearson Product Moment Correlation showed a reliability coefficient from 0.56 - 0.71 and 0.34 - 0.48 for the relationships between ratings of supervisors and coworkers respectively. It was concluded that there is a much higher degree of agreement among the ratings given by members of supervision than among ratings made by coworkers. The correlations obtained indicated a fairly high statistical reliability for the supervisory ratings.

Bray and Campbell (1968) assessed the efficiency of assessment centre method in personnel selection. Seventy-eight newly hired candidates in 14 telephone companies for sales position were evaluated by means of Assessment Centre consisting of paper and pencil test, an interview and individual and group stimulations. The assessment staff judgments, supervisory ratings and trainers ratings were compared with job performance some months later as evaluated by special observational team. Assessment results were strongly related to job performance criterion. Supervisors'

and trainers' ratings were not significantly related to the job performance criterion nor to assessment results. It was concluded that Assessment Centre method was efficient in personnel selection.

Grant and Bray (1970) were presented test validation data for telephone company installation and repair occupations. The sample was 211 from minority and 219 from non-minority groups of five geographically speared telephone companies. Five aptitude tests were administered at or immediately following employment. The scores from these tests were correlated with proficiency measures obtained from Learning Assessment Program. Correlation coefficients of 0.16 -0.41, 0.17 - 0.50 and 0.19 - 0.44 for minority, non minority and combined groups were obtained, respectively. It was concluded that all the tests were significantly predict success in the Learning Assessment Program.

In the study, the effect of language and anxiety on test performance in Ethiopian situation Jones (1973-2) administered geography and intelligence tests of identical ability constructed in Amharic and English to a random sample of 474 postal clerk candidates. It was found that the mean scores on the geography test for Amharic and English forms were 20.84 and 20.88 respectively. On the

intelligence tests, the mean scores for Amharic and English tests were 21.82 and 29.47, respectively. In the light of these findings it was concluded that a decision on whether to construct a test in Amharic or English depends on the use to which the test results are to be put. In this study, the English and Amharic geography tests were correlated with certain other measures of Amharic ability and English ability. The study revealed correlation coefficients of 0.45 between geography (Amharic) and Amharic ability tests, 0.45 between geography (Amharic) and English ability tests, 0.30 between geography (English) and Amharic ability tests and 0.40 between geography (English) and English ability tests. On the basis of these results the author generalized that tests constructed in the same language. correlate more highly with one another than in the case when tests in one language correlated with those in another.

To determine the effect of anxiety on test performance Jones (1973-2) administered short form (21 items). Test Anxiety scale to 97 postal clerk and 82 secretarial Trainee Candidates. The test scores of these candidates on Test Anxiety Scale was correlated with Geography (Amharic), Geography (English), DAT-V, DAT-N, Ottis Intelligence Test, Cattel's Culture Fair Test and Amharic Reading Tests, From this study the author concluded that test anxiety did have a

deleterious effect upon test performance. It was suggested that study should be designed to reduce anxiety associated with test taking.

Jones (1973-2) expressed his dissatisfaction with the prospective users of the AAU Testing Centre services not to appreciate the need to validate the test battery in the local organization. He wrote, most of the users want the Testing Centre to develop and administer tests to prospective employees or students and to rank the individuals according to their scores. The need for follow up validation studies to determine the effectiveness of the test battery and to revise it when appropriate, and the need to develop sound criterion measures have not been appreciated.

The use of tests in the selection of military in Africa was studied by McDonald (cited in Wober, 1975) who was working with British forces in East Africa. In this study scores from nearly 2,000 men on 14 tests were examined alongside ratings on the men's military performance. Although success in military performance is hard to define and express numerically, McDonald found a correlation of 0.43 between his tests and the criterion (Wober, 1975).

Levine, Abram and Ash (1977) conducted two studies on the usefulness of self-assessment in personnel selection. In study 1, it was hypothesized that self assessed levels of clerical ability would be significantly correlated with test measures and supervisory ratings of corresponding abilities. In this study, 73 clerical state government employees rated on 5-point scale their current abilities in several areas including spelling, reading and comprehension, Grammar, word meaning, composing business letters, speed and accuracy of proof reading, speed and accuracy of simple arithmetic concentration on accuracy of details, speed and accuracy of alphabetizing, and speed and accuracy of comparing names and numbers of exact likeness. Written test entitled the Intergovernmental Clerical Exam designed to measure the following clerical abilities: spelling, reading comprehension, English usage, word meaning, comparing copy, arithmetic and filing was administered.

Concurrently in study 1, supervisors were asked to fill performance ratings form on 14 performance dimensions of clerical jobs and two overall performance items on scales ranging from poor(1) to superior(5). The performance dimensions of clerical jobs included: checking and sorting, typing final copy, composing letters, computing figures, filing, posting records, repetitive routine work and typing

statistics. The correlations between self-ratings and written tests were between 0.30 and 0.74, self-ratings and supervisory ratings were between -0.21 and 0.58 and written test and supervisory ratings were between 0,28 and 0.67.

Significant positive correlations, in study 1 (Levine, Abram and Ash, 1977), between self-assessments and written test scores were found for spelling, grammar, word meaning, reading and arithmetic. Self-assessments of spelling, reading, grammar, speed and accuracy of proof-reading, comparing names and numbers and concentration on accuracy of details correlated significantly with supervisory ratings on corresponding performance dimensions. Written tests of spelling, reading, English usage, word meaning, comparing names and numbers correlated significantly with supervisory rating of corresponding job dimensions.

In study 1 (Levine, Abram and Ash, 1977) concluded that individuals can validly assess themselves on abilities, skills and knowledge as measured both by written test scores and supervisory ratings. Self-assessments may provide as a valid set of performance predictors similar to written test measures.

than the other tests.

In the light of the absence of past studies on the predictive validity of the secretarial tests given by TMU of the IER, the findings of the present study are highly tentative. From the wide use of these secretarial tests by AAU and other government organizations as a tool for the selection of secretarial candidates the study could be a starting point. Further exploration could be made by using the same or different criterion measures such as personality inventories, Co-workers' ratings, self-rating supervisory ratings of different sort (for instance, performance ratings prepared by a researcher and to be filled by the supervisors).

In general, item analysis of the G-III, G-II and G-I secretarial tests showed that the tests are useful as an instrument. The reason for obtaining not significant validity coefficients for most of the tests could be these tests were not developed on the basis of job analysis. Moreover, the validity and reliability of the performance evaluation form were not determined. Further investigation is necessary to improve these tests through job analysis and determine the validity and reliability of the performance evaluation form.

REFERENCES

Annastasi, A. Psychological Testing (4th ed.) N.Y.; MacMillan Publishing Co. Inc., 1976

Baker, F. B. "Advances in Item Analysis." Review of Educational Research, Vol. 47, 1977, pp. 151-178.

Borman, W.C. "Validity of Behavioral Assessments for Predicting Military Recruiter Performance," Journal of applied Psychology, vol. 67, 1982, pp 3-9.

Bray, D.W; & Campbell, R.J. "Selection of Salesman by Means of an Assessment Center." Journal of Applied Psychology, vol. 52, 1968, pp. 36-41.

Brown, F. Principle of Educational And Psychological Testing (3d Ed.). N.Y.: Holt, Rirehart and Wiriston, 1983.

Certo, S.C., J. & Appelbaum, S.H. Principles of Modern Management: A canadian Perspective (2nd Ed.). Iowa: Wm.C. Brown Publishers, 1985.

Cronbach, L.J. Essential of Psychological Testing (4th Ed.). N.Y.; Harper and Row Publishers, 1984.

Dunnettee, M.D. "A Note on the Criterion." Journal of Applied Psychology, Vol. 47, 1963, pp. 251-254.

Ebel, R.L. Essential of Educational Measurement (3rd Ed.). Englewood: Prentice-Hall, Inc., 1979.

Freeman, F.S. Theory and Practice of Psychological Testing (3d Ed.). New Delhi: Oxford & IBH Publishing Co., 1962.

Ghiselli, E.E., & Barthol, R.P. "The Validity of Personality Inventories in the selection of Employees." Journal of Applied Psychology, Vol. 37, 1953, pp, 18-20.

Glueck, W.F. Personnel: A Diagnostic Approach (3d Ed.). Plano, Texas: Business Publications, Inc., 1982

Grant, D.N., & Bray, D. W. "Validation of Employment Tests for Telephone Company Installation and Repair occupation." Journal of Applied Psychology, Vol. 54, 1970, PP. 7-14.

Gronlund, N.E. Measurement and Evaluation in Teaching (3rd Ed.). N.Y., MacMillan Publishing Co. Inc., 1976.

Henrysson,S. " Gathering, Analyzing and Using Data in Test Items." In Thorndike, R.L.(Ed.) Educational Measurement (2nd Ed.). Washington D.C.: American Council on Education, 1970.

Hopkins, K.D,. & Stanely, J.C. Educational and Psychological Measurement and Evaluation (6th Ed.). Englewood Cliffs: Prentice - Hall, Inc., 1981.

Jones, R.L. "Personnel Selection in Ethiopia: Principles and Problems." Technical Report, University Testing Center, 1973-2.

_____ " Services and Activities of the University Testing Center." Technical Report, University Testing Center, 1973.

Lawshe, C.H. Inferences from Personnel Tests and Their validity." Journal of Applied Psychology, Vol. 70, 1985, pp. 237-238.

Levine, E. L., Abram, F., & Ash, R.A. "Self-Assessment in personnel Selection." Journal of Applied Psychology, Vol. 62, 1977, pp. 428-435.

Lindeman, R. H. and Merenda, P.F. Educational Measurement (2nd Ed.). Illinois: Scott Foresman and Company, 1979.

Lord, F., Cowles, J.T., & Cynamon,M. " The Pre-Engineering Inventory as a predictor of Success in Engineering Colleges." Journal of Applied Psychology, Vol.34, 1950, pp. 30-39.

Mehrens, W. A., & Lehmann, I.J. Measurement and Evaluation in Education and Psychology (3rd Ed.). N.Y.: Holt, Rinehart and Winiston, 1984.

Meyer, H.H. " Predicting Supervisory Ratings Versus Promotional Progress in Test Validation Studies." Journal of Applied Psychology, Vol. 72, 1987, pp. 696-697.

Nitko, A.J. Educational Tests and Measurement: An Introduction. N.Y.: Hart Court Brace Jovanovich,Inc., 1983.

Orpen, C. An Empirical Assessment of the Job performance of High Level Executives by Means of Multitrait-Multimatrix Method." Psychologia Africana, Vol. 15, 1973, pp. 7-14.

Pearlman, K., Schmidt, F.L., & Hunter, J.E. "Validity Generalization Results form Tests Used to Predict Job Proficiency and Training Success in Clerical Occupations." Journal of Applied Psychology, Vol. 65, 1980, pp. 373-406.

Popham, W. J. Modern Educational Measurement. Englewood Cliffs: Prentice-Hall, Inc., 1981.

Shoenfeldt, L.F. "The Status of Test Validation." In Plake., B.S. (Ed.). Social and Technical Issues in Testing: Implications of Test Construction and Usage, Hillsdale, N.J.: Lawrence Erlbaum Associates Publishers, Vol. 1, 1984, pp. 61-86.

Shefritz, J. M. The Facts on File: Dictionary of Personnel Management and Labour Relations, 1985.

Schrader, W.B. "The Predictive Validity of College Board Admission Tests. In Angoff, W.H. (Ed.). The College Board Admission Testing Program: A Technical Report on Research and Development Activities Relating to Scholastic Aptitude Test and Achievement Tests. N.Y.: College Entrance Examination Board, 1971.

Seashore, H. H. "Reliability and Confidence." Test Service Bulletin, No. 44, 1952, pp. 2-7.

Shelley D., & Cohen, D. Testing Psychological Tests. London & Sydney: Croom Helm Ltd., 1986.

Springer, D. "Ratings of Candidates for Promotion by Co-workers and Supervisors." Journal of Applied Psychology,l. 37, 1953, pp. 347-351.

Sussman, M., & Roerston, D.U. "The Validity of Validity: An Analysis of Validation Study Designs." Journal of Applied Psychology, Vol. 71, 1986, pp. 461-468.

Thorndike, R. L. Personnel Selection. N. Y.: John Wiley & Sons, Inc., 1949.

_____ " Reliability." Journal of Counselling and Development, Vol. 63, 1985, pp. 528-530.

Thorndike, R. L., & Hagen, E,P. Measurement and Evaluation in Psychology and Education (4th Ed.). New Delhi: Wiley Eastern Ltd., 1977.

Tuckman, B. W. Measuring Educational Outcomes: Fundamentals of Testing. N.Y.: Harcourt Brace Jovanovich, 1975.

Turnage, J.J., & Muchinisky, P.M. "A Comparison of the Predictive Validity of Assessment Center Evaluations Versus Traditional Measures in Forecasting Supervisory Job Performance: Interpretative Implications of criterion Distortion for the Assessment Paradigm." Journal of Applied Psychology, Vol.69, 1984, pp. 595-602.

Wober, M. Psychology in Africa. London : International African Institute, 1975.

Yusuf, O. M. "Historic Merger of Former University Testing Center with the Institute of Educational Research (IER). "The Ethiopian Journal of Education, Vol. 10, 1989, pp.75-79.

Zippo. M. "Employee Selection Tests: Upping the Odds for Success." Personnel, Vol.57, 1980, pp. 48-49.

nonplausible destructors.

Naturally the variance of the items depends on difficulty values. Easy or Difficult items have low variances, whereas an item of average difficulty has high variances. Hence 76% of the items has variances between 0.15 and 0.25 which could be referred as moderates. It was found that almost 81% of the items were with significant positive correlations at 0.05 (N=100, two-tailed test). The reliability of the items revealed that 24% of the items were with reliability values of 0.01 to 0.05. These items contributed little (only 0.14) to the total test variance of 7.78. Hence, these items with low reliability values appears to need revision.

The KR-20 for G-II and G-I English comprehension test was 0.55 which indicates the test items had relatively low internal consistency.

The analysis of the discrimination and difficulty indices of the English and Amharic typing and shorthand test for G-III, G-II and G-I secretaries by Whitney and Sabers method showed (see Appendix-I) that the relatively difficult tests were G-III Amharic typing straight copy, G-II English typing-letters, tables and Amharic typing - revised

manuscript, and G-I English typing letters and Amharic typing - revised manuscript. The difficulty values for these test were from 0.41 to 0.46. The relatively easy tests were G-III English typing-letters, G-II English typing revised manuscript, and G-I English typing-tables and shorthand letter I with difficulty values form 0.60 to 0.66.

G-III Amharic typing-revised manuscript, G-II English typing-tables, G-I English typing-revised manuscript and tables were with relatively high discrimination power. The discrimination indices for these tests were between 0.62 to 0.65. Relatively low discriminating tests from each group were: G-III English typing - tables, G-II English typing straight copy, letters and revised manuscript and G-I Amharic typing - revised manuscript with discrimination values from 0.41 to 0.43.

The obtained discrimination and difficulty indices for the English and Amharic typing and shorthand tests in general were not either very low or very high. The discrimination indices for these tests were found to be between 0.41 to 0.65 while the difficulty indices were between 0.41 to 0.66.

5. DISCUSSION AND CONCLUSION

5.1. Discussion

The findings from the present study suggest that the validity coefficients between G-III secretarial tests and the supervisory ratings were ranging from -.32 to .28 (Table 1). Among the G-III typing tests Amharic Revised Manuscript correlated significantly with the last supervisory ratings before and the first supervisory ratings after promotion. Although English Typing -Straight Copy resulted in significant validity coefficients with the second supervisory ratings, it was not selected as a variable that adds significantly to the prediction when the stepwise procedure is applied.

From these findings it appears that among the G-III secretarial tests Amharic Typing - Revised Manuscript was relatively an efficient test in selecting G-III secretarial candidates.

The findings for G-II secretarial tests were more promising than that of either G-III or G-I secretarial tests. The validity coefficients for G-II secretarial tests

were between -.11 to .26 (Table 2). English typing-straight copy, Amharic typing-straight copy and English Comprehension were selected by the use of Stepwise procedure as variables which significantly contribute to the prediction of success in G-II secretarial work when the criterion measure was the first supervisory rating after promotion (Table 4). English typing- straight copy was also found to add significantly to the prediction when the criterion measure was the last supervisory ratings before testing. From these evidences it seems that English Comprehension, English typing-straight copy and Amharic typing straight copy could be of practical use for selecting applicants for G-II secretarial position.

The G-I secretarial tests except shorthand - Letter II did not provide dependable results that could predict success for G-I secretarial position. The validity coefficients for G-I secretarial tests with the supervisory ratings ranged from -.32 to .27 (Table 3). Significant negative correlation was obtained between shorthand - letter II and first supervisory ratings after promotion ($P < .05$). The remaining G-I secretarial tests correlated not significantly with each of the three supervisory ratings. From this it appears that the G-I secretarial tests except shorthand - letter II were not useful for the prediction of success in G-I secretarial work.

From the item analysis of the objective tests it could be observed that most of the items of G-III, G-II and G-I secretarial tests were found to have discrimination and difficulty indices between .20 to .80, point beserial correlations between 0.20 and 0.50 and reliability indices between .15 to .25 which could be considered as reasonably good because a good rule of thump would be to require r to be 0.20 or higher (Brown , 1983).

The internal consistencies of the objective tests using KR-20 were: .69, .67. .26, .87, .55, for G-III Arithmetic, English usage, English Comprehension, English Grammar and G-II and G-I English Comprehension tests, respectively. Discrimination and difficulty indices ranging from 0.41 to 0.66 were obtained for the typing and shorthand tests of G-III, G-II and G-I. The evidence from the item analysis of the G-III, G-II and G-I secretarial tests suggests that most of the tests have consistent items with the total test. The findings indicated the need to revise few items. The reliabilities of the objective tests were relatively moderate for G-III Arithmetic, English usage and G-II and G-I English comprehension tests, high for G-II English Grammar and low for G-III English comprehension. This low reliability for G-III English comprehension seems to be

attributed to the length of the test which was short containing only 15 items.

The results of the present study did not agree with past similar studies dealing with validates of selection tests. For instance, Grant and Bray (1970) obtained significant correlations ranging form o.16 to 0.50 between five aptitude tests and proficiency measures for telephone company installation and repair occupations. Levine, Abram and Ash (1977) found high correlations ranging from 0.28 to 0.67 between clerical ability test and supervisory ratings. Borman (1982) reported a validity coefficient close to 0.50 between assessment ratings and performance ratings for the selection of army recruiters.

In the present study, however, correlation of -.32 to .28 were obtained between the secretarial tests and the supervisory ratings. Significant correlations were obtained only for few tests. From each group not more than two tests were correlated significantly with at least one of the criterion measures. As a result, most of the tests were not found to be valid predictors of success in secretarial work.

The probable reason for the differences between the past and the present studies could be due to differences in the

development and standardization of the tests. In the present study evidences were not available about the development and standardization of the tests. In the process of developing a test, if it was desired to have a test that provides strong correlation with the criterion measures, as a result a highly valid test, those items which are related strongly with the external criterion measures should be constructed. Furthermore, the construction of each item should be based on job analysis. Important testing procedures such as adminstration, scoring and interpretation should be standardized to have a valid test.

The findings of the present study revealed that there was not any test that was significantly correlated with the three criterion measures. The study also showed that no test was significantly correlated to both the first and the second supervisory ratings after promotion. One possible explantation for this is that the first ratings after promotion could be low because the secretarial works in the new position may be new demanding a longer time of acquaintance and hence the supervisors may assign low ratings to the new secretary. Another possible explanation is that the first ratings after promotion could be high due to high dedication to work at the beginning after promotion

with subsequent decline in the amount of dedication to work which could lower the second ratings after promotion.

The effect of taking the same test repeatedly by the secretarial candidates on the validity of the secretarial tests was not investigated in the present study. A candidate with failing score could take secretarial tests repeatedly for unlimited number of times, at least every three months, until the Personnel Department of the AAU declares that the candidate has scored a passing score in a given test. The present study considered the best scores of the candidates available in the records kept by TMU of the IER. It is felt that the result could be different if scores form the first adminstration of the tests were used in the validation of the secretarial tests instead of the best available scores obtained by taking the tests repeatedly.

5.2. Conclusion

The purpose of the present study was to determine the predictive validity of G-III, G-II and G-I secretarial tests given by TUM of the IER. Along with this predictive study,

the items of the objective type tests for G-III, G-II and G-I were analysed using the criterion of internal consistency. The discrimination and difficultly values of the typing and shorthand tests were also estimated.

With the use of simple and stepwise regression analysis, significant positive validity coefficients ($P < .05$) were obtained for G-III Amharic Typing Revised Manuscript, G-II English and Amharic typing -straight copies and English Comprehension and significant negative validity coefficient for G-I shorthand - Letter II. No other significant validity coefficient was obtained. From this it appears that the secretarial tests with significant validity coefficients could be used for selecting secretarial candidates.

Item analysis data for the objective tests indicated few items from each test that could be revised. Estimation of the reliabilities of the secretarial tests with KR-20 resulted with reliability indices of 0.69, 0.67, 0.26, 0.87 and 0.55 for G-III Arithmetic, G-III English usage, G-III English comprehension, G-III English Grammar and G-II and G-I English comprehension tests, respectively. The G-III English Comprehension test has low internal consistency which indicates that this test is relatively less dependable

For the tests of aptitude, administered to examinees who are unfamiliar with its content, items with difficulty values 0.80 and above could be referred to as easy items while items with difficulty values of 0.20 and below could be referred to as difficult items (Popham, 1981).

The item discrimination values for G-III arithmetic test (see Appendix-D) were all positive, Items 1,12 and 21 were poor discriminators with discrimination values of 0.02, 0.08 and 0.18, respectively.

A close look at the items with low discrimination values shows that item 1 is relatively easy demanding knowledge of simple multiplication and items 12 and 21 are word problems which demand the candidates understanding and translation to numerical values to arrive at the answers.

Items 9 and 15 are marginal items with discrimination values of 0.28 and 0.20, respectively. Six items with discrimination values between 0.30-0.33 could be referred as reasonably good items while the remaining eleven (50%) items with discrimination values above 0.40 could be consider as very good items.

The contribution of each items to the variance of the

test as a whole seems to be satisfactory since the variances for most of the items were greater than 0.20 Eight-two percent of the point biserial correlations were statistically significant from 0 (N=100,r≥0.20, two tailed t-test) at 0.05 level . So, most of the items are strongly related to scores on the test.

A reliability index combines measures of difficulty and discrimination indices. It indicates the contribution of each item to the total variance. Items, 1 12, 21 and 22 of the G-III arithmetic test contributed only 0.13 to the total variance of the test (13.40). In the final analysis these items need revision.

The KR-20 is a measure of internal consistency of a test. Since the objective type tests were dichotomously scored coefficient alpha gives the same result as KR-20. The KR-20 for the G-III arithmetic test was 0.69 which is relatively moderate.

Item analysis for G-III English usage test (see Appendix-E) showed 14% of the items as poor discriminators with discrimination values ranging from 0.00 to 0.18. An investigation of these items revealed that the items evaluate knowledge of proposition and vocabulary which are

usually difficult to be answered unless the candidate is absolutely sure of the answer. The remaining 17% of the items were marginal items, 17% were reasonably good items and 52% were very good items. About 18% of the items were with difficulty values below 0.20 and about 5% of the items were with difficulty values above 0.70. These figures indicate that most of the items were difficult.

The correlation of 64% of the G-III English usage test with the total test scores showed statistically significant correlations at 0.05 level (N=100, r≥20 and two-tailed t-test). The reliability of 40% of the items were less or equal to 0.09 contributing only 0.74 to the total test variance of 24.30. These items appear to need revision. The test has a moderate reliability index of KR-20 = 0.67 which suggests that the test is useful.

The G-III English Comprehension test (see Appendix-F) has items with discrimination power of 0.25 to 0.82. Only items 2, 6 and 11 were marginal items. According to Ebel's (1979) criteria the remaining items are reasonably good or very good items. 27% of the items were with difficulty values of 0.05 to 0.11 and almost for all items the difficulty values were below 0.50 indicating the items were difficult. Two-third of the items showed significant

correlation at 0.05 level (N=100; two-tailed t-test). This indicates the existence of substantial relationship between most of the items and the total test scores. 27.6% of the items contributed only 0.14 to the total test variance of 4.00. It appears that these items need revision. The internal consistency of the test was low (KR-20=0.26). One reason for this is that the test contained only 15 items which is short. Usually tests with few items have low internal consistency.

The item analysis for G-III English Grammar test (see Appendix-G) indicated that 12 items (32%) were found to be perfect discriminators (discrimination value = 1.00). In general, most of the items were with discrimination values greater than 0.40, which could be referred to as very good items. The maximum difficulty value was 0.72 and the minimum was 0.03. Further more, 24% of the items were with difficulty values, below 0.20 and these items appear to be difficult. One possible reason for this is that the test contains three sets of items demanding the candidates to provide the past tense of irregular verbs, plurals for singular names and feminine for masculines. These demand good command of past tense of English irregular verbs, plurals of English nouns and masculines of English feminine which could not be expected from the subject of this study.

None of these items were with difficulty values greater that 0.80 which indicates that the test did not contain easy items.

Almost 95% of the English Grammar test of G-III showed significant positive correlation at 0.05 level (N=100, two-tailed t-test). 5 items were found to be with reliability values of 0.03 to 0.09 which is very low. These items contributed only 0.35 tothe total variance of the test. At least these items seem to need revision. The test had a high reliability index of KR-20 = 0.87.

The English comprehension test of G-II and G-I was analysed on the basis of the response patterns of G-II and G-I groups. separate item analysis for these group was not made because the number of available answer sheets was found to be small for each group. On the basis of G-II and G-I response patterns, the item analysis of English Comprehension test for G-II and G-I (see Appendix-H) showed low discriminations (33% of the items were with discrimination values between 0.04 to 0.17). The difficulty values of 33% of the items were above 0.80 indicating that the items appear to be easy. It was observed that most of the stems for the English comprehension test for G-II and G-I were direct statements from five passages with

English typing-straight-copy results in a relatively high precision in the prediction. However, the use of G-II Amharic typing-straight-copy as a predictor of first supervisory ratings after promotion results in a relatively low precision in the prediction (Table 2).

In table 2, English comprehension, English typing-straight copy Amharic typing-straight copy correlated significantly with the first supervisory ratings and not correlated significantly with the second supervisory ratings. This may be due to the inconsistency of supervisors in evaluating secretaries. In addition, as time increases the relationship between the supervisors and the secretaries could be worsen or strengthen. As a result supervisory ratings may depend on the personal relations between the supervisors and the secretaries (Cronbach, 1984).

The predictive validity coefficients of G-II secretarial tests relative to the second supervisory ratings after promotion (Table 2) were generally low and statistically not significant.

Work experience correlated negatively with all the

variables in Table 2. This indicates that supervisors tend to evaluate low those secretaries with long work experience. One possible explanation for this could be secretaries at a lower position with long work experience, if left unprompted for a longer time, tend to fail to fulfil their duties.

With regard to G-I secretary, Table 3 presents means, standard deviations, inter-correlations, b-weights and standard errors of each of the independent and dependent variables.

The inter-correlations among the independent variables (Table 3) revealed the absence of strong associations between English comprehension (x_1) and each of the typing and shorthand tests (x_2 to x_7) except for shorthand letter I (x_8) which showed significant correlation ($P < 0.05$). This is due to the fact that English comprehension is relatively a measure of understanding whereas typing tests are relatively a measure of skills. Shorthand tests partly measure listening which could be a factor to result in relatively closer association with English comprehension.

In many instances of Table 3 substantial positive associations were obtained among the G-I typing tests. Table 3 also shows a considerable overlap between shorthand-

letter I and shorthand-letter II. Work experiences and Educational Status related not significantly with most of the secretarial test scores of G-I. Substantial correlations were obtained between work experience, and English-typing-revised manuscript and English typing-tables, Educational Status and English typing-revised manuscript (Table 3).

With regard to the correlations among the criterion measures table 3 indicated relatively strong positive relations which implies the existence of consistency among the ratings.

Analysis of the validity coefficients of G-I secretarial tests (Table 3) revealed the absence of any systematic positive relationships among the predictors and the criterion measures. The first supervisory ratings after promotion and shorthand-letter-II of G-I secretarial test showed negatively significant correlations ($P < 0.05$). On the basis of this data it appears that G-I secretarial candidates with low scores in shorthand-letter II are to be promoted or selected rather than those with high scores in shorthand-letter II, since G-I secretarial candidates with the low scores in shorthand-letter II tend to get high ratings before and after promotion.

Table 3. Means, Standard Deviations, Inter-correlations, b-Weights and Standard Errors of the predictors and Criterion Data for G-I Secretary (N=42)

Variables	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Last Ratings Before Test (Y_1) r	b	SE	First Ratings After promotion. (Y_2) r	b	SE	Second Ratings After promotion (Y_3) r	b	SE
Eng. Comprehension (X_1)		-.03	-.06	-.00	0.06	.17	.07	.30*	.24	-.06	.03	.12	.02	-.00	.17	.00	.11	.18	.01	.05	.19
Eng.Typing-Straight Copy (X_2)			.40*	.28*	-.02	.25	.31*	.28*	.19	-.03	.06	-.42*	-.19	-.44*	.20	.01	.01	.22	.09	.06	.23
Eng. Typing - Letters (X_3)				.47*	.18	.36*	.43*	.04	.10	.03	-.13	-.12	-.09	.03	.20	-09	.07	.21	-.07	-.00	.22
Eng. Typing - Revised Manuscript(X_4)					.13	.41*	.49*	.26*	.19	.29*	-.26*	-.12	.03	-.08	.21	.04	.08	.23	-.05	-.13	.24
Eng. Typing - Tables (X_5)						.10	.17	.04	-.11	-.31*	.22	.01	.16	.26	.18	.09	.14	.19	.01	.09	.20
Amh. Typing-Straight Copy (X_6)							.64*	.42*	.38*	.20	-.19	-.20	.09	-.11	.24	-.23	-.36	.25	-.11	-.25	.27
Amh. Typing-Revised Manuscript (X_7)								.23	.19	.03	-.14	-.15	-.01	.04	.22	-.04	.10	.24	-.06	.08	.24
Shorthand - Letter I (X_8)									.41*	.06	.13	-.16	.18	.18	.20	-.01	.15	.21	.12	.21	.23
Shorthand - Letter II (X_9)										.10	-.10	-.00	.09	.14	.18	-.32*	-.30	.19	-.14	-.16	.20
Work Experience (X_{10})											-.12	.19	.18	.38*	.19	.04	.19	.20	.14	.30	.21
Educational Status (X_{11})												.30*	.01	.14	.19	.01	-.03	.21	.11	.06	.22
Number of Ratings from Test Date to Promotion (X_{12})													-.21	-.50*	.16	-.12	-.18	.21	-.09	-.17	.22
Y_1																.43*			.46*		
Y_2																			.78*		
Mean	51.5	53.2	53.9	55.9	64.3	49.4	55.1	48.5	53.6	17.2	1.76	2.81		4.48			4.49			4.49	
S.D	7.72	3.33	6.74	7.96	9.54	7.34	7.57	7.40	10.9	5.97	.88	1.29		.39			.36			.38	

*P<.05

Shorthand-letter II is correlated significantly with Shorthand-letter I and the first supervisory ratings after promotion. But Shorthand-letter I is correlated near zero with the first supervisory ratings (r=-0.01). With the use of partial correlation the removal of the effect of Shorthand-letter I from the correlation between Shorthand-letter II and the first supervisory ratings resulted in not significant increase (increase in r=0.06) in the correlation between Shorthand-letter II and the first supervisory ratings after promotion. Hence, the significant relationship between shorthand-letter I and Shorthand-letter II does not significantly affect the correlation between Shorthand-letter II and the first supervisory ratings after promotion.

Stepwise regression procedure was applied to identify those variables which could be greatly attributed to the prediction achieved. Table 4 presents the result.

Table 4 Results of Stepwise Regression Analysis for G-III, G-II and G-I Secretaries

Group	step	Predictor Variable	Criterion Variable	b	se	R^2	F
G-III Secretary (N=76)	1	Amharic Typing Revised Manuscript (X_9)	Last Rating Before test (Y_1)	.29	.11	.06	7.12*
	1	Amharic Typing Revised Manuscript(X_9)	First Rating After Promotion (Y_2)	.24	.10	.06	6.38*
G-II Secretary (N=71)	1	English Typing Straight copy (X_2)	Last Rating Before Test (Y_1)	.26	.12	.07	4.98*
	1	English Typing Straight Copy(X_2)	First Ratings After Promotion (Y_2)	.29	.11		
	2	Amharic Typing Straight Copy(X_6)		.30	.11	.18	4.96*
	3	English Comprehension (X_1)		.22	.11		
G-I Secretary (N=42)	1	Shorthand Letter II (X_9)	First Ratings After Promotion (Y_2)	.32	.15	.10	4.54*

*P<0.05

Application of the stepwise procedure to G-III secretarial test scores (Table 4) resulted in the selection of Amharic Typing- revised manuscript (x_9) as a sole predictor of success in secretarial jobs when the last supervisory ratings before testing and the first supervisory ratings after promotion were used as a criterion measures. The contribution made by the Amharic typing-revised manuscript to the prediction was about 6% in both criterion measures. It was found that the inclusion of additional variables adds nothing of much significance to the multiple correlation coefficient. It appears that the G-III secretarial tests except Amharic Typing-revised manuscript were not valid predictors of success in secretarial work. Thus, this finding indicants all the G-III secretarial tests should not be given much importance in selecting applicants. Instead, importance could be given to Amharic Typing-revised manuscript which could predict that could be predicted by all the G-III secretarial tests.

The stepwise procedure with G-II secretarial test scores (Table 4) showed English Typing straight copy (x_2) as the sole significant predictor of last supervisory ratings before test. It accounted for 7% of the variance in the last supervisory ratings before testing. With the first

supervisory ratings after promotion the stepwise procedure selected three secretarial test scores. It resulted in the initial selection of English Typing-Straight Copy (x_2) followed by the selection of Amharic Typing Straight Copy (x_6) and in the final selection of English Comprehension (x_1). These tests explain 18% of the variance in the first supervisory ratings after promotion.

Table 4 also shows results of stepwise regression result for G-I secretarial test scores. The analysis selected Shorthand-Letter II (x_9) as a variable that explains significantly the first supervisory ratings after promotion. Table 3 showed a negative strong relationship between Shorthand-Letter II and the first supervisory ratings after promotion. Shorthand-Letter II attributed by 10% to the variance of the first supervisory ratings after promotion. The existence of negative relationship indicates that applicants for G-I secretarial position with low scores in shorthand-Letter II should be selected. From these findings it appears that among G-I secretarial tests shorthand-Letter II is valid predictor of success in secretarial work.

The comparison of Table 1, 2 and 3 in terms of the inter-correlations among the independent variables indicates that, as a whole, the pattern of relationship is stronger

among each of the written and each of the typing tests than among the written and typing tests. Within the typing tests in all cases the English typing tests generally are more related to each other than to the Amharic typing or shorthand tests. On the whole, work experience and educational status showed inappreciable association with the secretarial tests. In the three tables the relationship between the number of supervisory ratings from test date to promotion and the last supervisory ratings before testing were negative.

In most of the cases in each secretarial group comparison of validity coefficients across each criterion measures generally showed similar results.

4.2. Analysis and Interpretation of Item Analysis.

In the following the findings based on the item analysis of the objective, typing and shorthand tests are presented.

In the item analysis section, the item discrimination and difficulty values of the objective tests of G-III (Arithmetic-form 77B, English usage Form 77A), and G-III and G-I (English comprehension Part VII) were analysed on the

basis of the criteria suggested by Ebel (1979) and Popham (1981).

Using the upper and lower 27% to calculate the discrimination index Ebel (1979) offered the following experience based guidelines to evaluate the quality of test items.

Discrimination index	Item Evaluation
0.40 and up	Very good items
0.30 - 0.39	Reasonably good but possibly subject to improvement.
0.20 - 0.29	Marginal items, being subject to improvement
below 0.20	Poor items, to be rejected or improved by revision.

In study 2 (Levine, Abram amd Ash, 1977) 569 applicants for clerical jobs rated their typing speed and were later given a standardized typing test. Typing test scores correlated 0.62 or better with self assessments. From this study, it was concluded that applicants assessed themselves accurately and validly on typing ability.

Pearlman, Schmidt and Hunter (1980) applied the Bayesian validity generalization procedure to 56 distribution of validity coefficients drawn from 698 published and unpublished studies representing five clerical job families, 10 test types and two classes of criteria (job proficiency and training success). Results indicated that a serious doubt on the traditional belief that employment test validates are situationaly specific. It was concluded, in most cases generalization of validity to similar clerical jobs or new setting was justified.

The article by Bormans (1982) described results from an evaluation experiment to determine the validity of an assessment center designed to help the selection of US Army recruiters. Sixteen experienced and successful army recruiters assessed 57 soldiers entering the army recruiter school. Validity of the assessment ratings was evaluated by correlating the assessment ratings with subsequent

performance in training. A composite of assessment ratings yielded correlated validates close to 0.50 which was significant at 0.05 level. It was concluded that the assessment program is reasonably successful in predicting recruiter training performance.

Turnage and Muchinisky (1984) examined the ability of assessment center evaluations to predict actual job performance criteria. The authors also compared the predicability of an Assessment Center evaluations with traditional measures in fore casting job success. The subject were 799 (92%) males) employees who were promoted to supervisory position. The predictor variables were Assessment Center Data and tests of arithmetic and verbal measures of general ability. The criterion data included: Standardized appraisal ratings for performance, promotions past the foreman level, transfers and reductions, quits and discharges and salary progress. The average correlation between the predictor variables and the criterion data was 0.40. All of the correlations were positive and significant. The results revealed that both assessment center evaluations and traditional predictors, predicated job performance criteria and success in promotion.

In this review of literature various research articles dealing with the selection of personnel in various job categories such as telephone company installations and repair occupations clerical jobs, army recruiters were reviewed. The review also dealt with instruments used in personnel selection including personality inventories, self-assessments, Assessment Center, written tests, supervisors and co-workers' ratings. Over all the various research literatures indicate that personnel selection tests are valid in selecting employees for the various job categories.

4. ANALYSIS AND INTERPRETATION OF DATA

In this section analysis of the predictors and criterion, and item analysis data are presented. First descriptive data on the predictors and criterion variables, results of simple regression and stepwise regression analysis related to the G-III, G-II and G-I secretarial tests are presented. Second, data related to item analysis of the objective, typing and shorthand tests of the G-III, G-II and G-I secretaries are presented.

In the analysis the three groups were treated individually. No attempt was made to compare the groups, because data from the predictor variables were different for each group.

4.1 Descriptive Data, Simple and Stepwise Regression Analysis.

Table 1 below presents the means, standard deviation inter-correlations, b-weights and standard errors of estimates of the predictor (independent) variables with respect to the criterion measures (dependent variables) of G-III secretary.

Inspection of the inter-correlations among the independent variables (Table 1) reveals significant correlations ($P < 0.05$) among:

1. Arithmetic test and each of the English usage, English Grammar and Educational Status,

2. English usage and each of the English Grammar and English Comprehension, and

3. The typing tests (English and Amharic) except English-typing tables and Amharic-typing straight copy which were correlated not significantly.

The pattern of inter - correlation in Table 1 shows that the written tests are more associated with each other than with the typing tests and the typing tests are also more related with each other than with the written tests. The existence of such relationships seems due to the fact that the written tests which measure verbal and mathematical abilities are more related to knowledge where as the typing tests which measure speed, accuracy and finger dexterity are more related to skills.

Work experience is found to have weak relation with many of the G-III secretarial tests (Table 1) and educational status is found to have strong relations with arithmetic, English Grammar and English typing - straight copy. The weak relation between work experience and G-III secretarial test implies that the G-III secretarial tests demand new knowledge and skills which may not be directly related to the experiences of the candidates in the secretarial field. The negative substantial correlation between the number of supervisory ratings from test date to promotion and last supervisory ratings before test indicates that the G-III secretaries with lower supervisory rating wait for longer time to be promoted than those with higher supervisory ratings.

The simple regression analysis for G-III secretary (Table 1) indicated that Amharic typing-revised manuscript predicted individually about 8% of the variances in the last supervisory ratings before testing with significant b-weights, although its precision was relatively low. With second supervisory ratings after promotion English typing - straight copy is found to be significant predictor of success in G-III secretarial work explaining a little more the 4% of the variances of the second supervisory ratings after promotion.

Strong significant relationship ($P < 0.05$) was obtained among the three supervisory ratings. This indicates that the ratings were relatively consistent.

With regard to G-II secretaries, descriptive data and simple regression analysis afforded the results in Table 2 below.

From the inter-correlations of the predictor variables (Table 2) it could be seen that English comprehension (x_2) was negatively correlated with typing tests where two of them are significant ($P < 0.05$). This indicates that candidates who performed better in English comprehension performed low in typing tests because English comprehension measures understanding whereas the typing tests measure skills. Among the English typing tests Significant positive correlation ($P < 0.05$) was obtained between straight copy (x_2) and letters (x_3). The Amharic typing tests were also considerably associated with each other.

Table 1 Means, Standard Deviations, Inter-Corelations, b-weights and Standard Errors of the predictors and Criterion Data for G-III Secretary (N=76)

Variables	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Last Ratings Before Test (Y_1) r	b	SE	First Ratings After promotion. (Y_2) r	b	SE	Second Ratings After promotion (Y_3) r	b	SE
Arithmetic (X_1)		0.39*	.12	.39*	-.03	.02	-.08	.16	.02	-.11	.31*	-.09	-.17	-.24	.13	-.10	.05	.12	.06	.01	.12
Eng. Usage (X_2)			.19*	.35*	-.17	-.24*	-.19*	-.07	-.21*	-.17	.07	.10	-.14	-.02	.13	-.03	-.03	.13	.07	.05	.13
English Comprehension (X_3)				.14	.02	.03	.24*	-.12	-.07	-.12	.10	.01	.10	.13	.11	.00	-.05	.11	.01	-.15	.11
English Grammer					-.09	-.22*	-.08	-.05	-.18	-.29*	.24*	.06	-.12	-.07	.13	-.02	-.16	.12	.10	.03	.12
English Typing St. Copy (X_5)						.44*	.20*	.24*	.39*	.04	.20*	-.16	.13	-.03	.13	.04	-.05	.12	.21*	.04	.12
English Typing Rev. Manuscript (X_6)							.36*	.27*	.51*	-.13	.09	-.09	.15	-.01	.14	.04	-.07	.13	.15	-.01	.13
English Typing Tables (X_7)								-.12	.33*	-.17	.13	-.08	.16	-.10	.13	.08	-.06	.11	-.04	-.10	.11
Amharic Typing St. Copy (X_8)									.40*	.00	.06	-.06	-.06	-.17	.12	-.00	-.03	.12	.13	.08	.13
Amharic Typing Revised Manuscript (X_9)										-.15	.06	-.10	.28*	.32*	.14	.11	.28*	.14	.14	.11	.14
Work Experience (X_{10})											-.11	-.12	-.15	-.18	.12	-.17	-.12	.11	-.06	-.05	.11
Educational Status (X_{11})												-.09	.11	.15	.12	.09	.06	.11	.12	.12	.11
No. of Ratings from Test Date to prom: (X_{12})													-.32*	-.34*	.11	-.08	.04	.10	-.23*	-.12	.10
Y_1																.55*			.37*		
Y_2																			.51*		
Mean	55.5	50.3	51.1	55.1	57.2	59.2	58.2	57.5	61.8	7.66	1.45	1.29	4.09			4.08			4.08		
S.D.	7.12	6.67	7.56	9.72	7.23	8.17	7.16	7.13	6.83	2.64	.77	.56	.39			.51			.48		

*P < .05

From the fact that at least speed and accuracy are common to all the typing tests it would be strange to get significant negative correlation between English typing-letters and Amharic Typing straight copy. One possible explanation for this negative result would be the variation in the amount of daily practice attached to English and Amharic typing.

In Table 2, work experience was related substantially to English comprehension and English typing-tables. The correlations between Educational Status and all the G-II secretarial tests were found to be not significant. Significant int-correlations were observed among the supervisory ratings ($P < 0.05$).

With the last supervisory ratings before testing, among the G-II secretarial test scores (Table 2), English Typing-straight copy was found to be the only significant predictor contributing individually about 7% to the prediction. With respect to the first supervisory ratings after promotion English and Amharic typing straight copies were found to be significant predictors of success in G-II secretarial work, contributing individually 4% and 7% respectively to the prediction. Moreover, the standard errors showed the use of English typing-straight-copy results in a relatively high

Table 2. Means, Standard Deviations, Inter-correlations, B-weights and Standard Errors of the Preictors and Criterion Data for G-II Secretary (N=71)

Variables	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Last Ratings Before Test (Y_1) r	b	SE	First Ratings After Promotion (Y_2) r	b	SE	Second Ratings After Promotion (Y_3) r	b	SE
Eng. Comprehension (X_1)		.06	-.14	-.11	-.27*	-.18	-.34*	.20*	.12	.14	.12	.06	.14	.19	.26	.13	.10	.20	.13
Eng. Typing St. Copy (X_2)			.34*	-.12	-.17	-.18	-.05	.00	.16	-.20*	.26*	.18	.14	.26*	.16	.13	.16	.31*	.13
English Typing - Letters (X_3)				-.14	.02	-.30*	.05	.03	-.16	-.02	.17	.05	.15	-.06	.18	.14	.01	-.06	.14
English Typing - Revised Manuscript (X_4)					.19	.36*	.31*	-.06	.10	-.19	.03	.13	.14	.13	-.03	.13	.07	.08	.14
English Typing - Tables (X_5)						.20*	.16	-.26*	.06	.16	-.07	-.11	.14	.01	-.02	.13	-.12	.04	.13
Amharic Typing Straight Copy (X_6)							.53*	.04	-.06	.05	.03	.16	.17	.21*	.11	.16	.03	.24	.16
Amharic Typing Revised Manuscript (X_7)								-.14	.03	.13	.10	-.17	.17	.08	.16	.16	-.08	-.01	.16
Work Experience (X_8)									-.12	.06	-.11	-.05	.13	.09	-.14	.13	.02	.06	.13
Educational Status (X_9)										-.01	.05	.02	.14	.10	.02	.13	.03	.01	.13
No. of Ratings from Test Date to prom. (X_{10})											-.22*	.06	.14	.01	-.24	.13	-.05	.05	.13
Y_1														.47*			.37*		
Y_2																	.66*		
Mean	51.1	53.0	50.0	55.0	61.1	54.0	56.9	11.3	1.56	1.69		4.32			4.34			4.34	
S.D	7.46	5.91	7.78	6.13	9.28	4.56	6.17	4.39	.65	.94		.50			.46			.50	

*P < .05

APPENDIX-D

Discrimination, difficulty, correlation and relaibility indices for G-III Arithmetic test

Item	Dis.Value	Diff.Value	Vari	Correlation	Reliability
001	0.02	0.88	0.11	0.01	0.00
002	0.46	0.78	0.17	0.54	0.22
003	0.41	0.24	0.18	0.32	0.14
004	0.57	0.57	0.25	0.42	0.21
005	0.59	0.57	0.25	0.48	0.24
006	0.49	0.67	0.22	0.51	0.24
007	0.66	0.51	0.25	0.52	0.26
008	0.33	0.65	0.23	0.35	0.17
009	0.28	0.66	0.22	0.35	0.16
010	0.33	0.72	0.20	0.37	0.17
011	0.64	0.36	0.23	0.43	0.21
012	0.08	0.49	0.25	0.08	0.04
013	0.30	0.76	0.18	0.46	0.20
014	0.87	0.24	0.18	0.49	0.21
015	0.20	0.71	0.21	0.29	0.13
016	0.31	0.66	0.22	0.37	0.17
017	0.53	0.25	0.19	0.40	0.17
018	0.52	0.54	0.25	0.51	0.25
019	0.67	0.47	0.25	0.56	0.28
020	0.33	0.37	0.23	0.21	0.10
021	0.18	0.34	0.22	0.19	0.09
022	0.33	0.11	0.10	0.01	0.00
KR	0.69				
SX2	13.40				

APPRNDIX-

Discrimination difficulty, correlation and relibility indices for G-III English usage test

Item	Dis.Value	Diff.Value	Vari	Correlation	Reliability
001	0.41	0.27	0.20	0.33	0.15
002	0.45	0.42	0.24	0.33	0.16
003	0.31	0.51	0.25	0.30	0.15
004	0.53	0.28	0.20	0.30	0.13
005	0.00	0.18	0.15	0.04	0.02
006	0.40	0.35	0.23	0.23	0.11
007	0.22	0.34	0.22	0.20	0.09
008	0.40	0.17	0.14	0.16	0.06
009	0.43	0.51	0.25	0.30	0.15
010	0.45	0.39	0.24	0.19	0.09
011	0.38	0.29	0.21	0.11	0.05
012	0.26	0.68	0.22	0.40	0.19
013	0.78	0.29	0.21	0.44	0.20
014	0.21	0.75	0.19	0.35	0.15
015	0.18	0.80	0.16	0.34	0.14
016	0.68	0.59	0.24	0.53	0.26
017	0.59	0.56	0.25	0.47	0.23
018	0.44	0.31	0.21	0.28	0.13
019	0.89	0.29	0.21	0.48	0.22
020	0.45	0.53	0.25	0.35	0.17
021	0.38	0.56	0.25	0.35	0.17
022	0.67	0.20	0.16	0.30	0.12
023	0.00	0.09	0.08	0.02	0.01
024	0.63	0.27	0.20	0.26	0.12
025	0.07	0.28	0.20	0.06	0.03
026	0.63	0.27	0.20	0.41	0.18
027	0.33	0.16	0.13	0.19	0.07
028	0.44	0.38	0.24	0.34	0.17
029	0.33	0.27	0.20	0.12	0.05
030	0.22	0.35	0.23	0.13	0.06
031	0.45	0.25	0.19	0.17	0.07
032	0.09	0.20	0.16	0.08	0.03
033	0.00	0.31	0.21	0.04	0.02
034	0.39	0.62	0.24	0.47	0.23
035	0.25	0.16	0.13	0.06	0.02
036	0.25	0.15	0.13	0.07	0.03
037	0.50	0.32	0.22	0.29	0.14
038	0.45	0.25	0.19	0.16	0.07
039	0.42	0.38	0.24	0.31	0.15
040	0.37	0.37	0.23	0.26	0.12
041	0.26	0.33	0.22	0.20	0.09
042	0.67	0.18	0.15	0.33	0.13
KR	0.67				
SX2	24.30				

APPENDIX-F

Discrimination, difficulty, correlation and reliability indices for G-II English Comprehension test

Item	Dis.Value	Diff.Value	Vari	Correlation	Reliability
001	0.42	0.47	0.25	0.34	0.17
002	0.25	0.38	0.24	0.28	0.14
003	0.35	0.60	0.24	0.29	0.14
004	0.42	0.43	0.25	0.39	0.19
005	0.50	0.05	0.05	0.03	0.01
006	0.25	0.11	0.10	0.13	0.04
007	0.33	0.16	0.13	0.08	0.03
008	0.56	0.50	0.25	0.37	0.18
009	0.40	0.14	0.12	0.17	0.06
010	0.60	0.42	0.24	0.49	0.24
011	0.26	0.53	0.25	0.31	0.15
012	0.50	0.38	0.24	0.33	0.16
013	0.58	0.44	0.25	0.40	0.20
014	0.82	0.20	0.16	0.31	0.12
015	0.47	0.55	0.25	0.44	0.22
KR	0.26				
SX2	4.00				

APPENDIX-G

Discrimination, difficulty, correlation and reliability indices for G-III English Grammar test

Item	Dis. Value	Diff.Value	Vare	Correlation	Reliability
001	0.40	0.55	0.25	0.39	0.19
002	0.64	0.45	0.25	0.52	0.26
003	0.89	0.33	0.22	0.55	0.26
004	0.63	0.63	0.23	0.63	0.30
005	0.89	0.31	0.21	0.54	0.25
006	1.00	0.15	0.13	0.46	0.17
007	1.00	0.19	0.15	0.54	0.21
008	0.54	0.43	0.25	0.46	0.23
009	1.00	0.32	0.22	0.55	0.26
010	0.49	0.72	0.20	0.53	0.24
011	1.00	0.35	0.23	0.65	0.31
012	1.00	0.09	0.08	0.30	0.08
013	0.73	0.44	0.25	0.48	0.24
014	0.35	0.69	0.21	0.44	0.20
015	1.00	0.10	0.09	0.37	0.11
016	1.00	0.22	0.17	0.29	0.12
017	1.00	0.10	0.09	0.43	0.13
018	0.89	0.42	0.24	0.50	0.24
019	1.00	0.03	0.03	0.18	0.03
020	0.83	0.36	0.23	0.60	0.29
021	1.00	0.21	0.17	0.61	0.25
022	0.71	0.10	0.09	0.31	0.09
023	0.41	0.38	0.24	0.38	0.19
024	0.47	0.31	0.21	0.25	0.11
025	0.82	0.16	0.13	0.37	0.13
026	0.41	0.62	0.24	0.48	0.24
027	0.18	0.69	0.21	0.25	0.11
028	0.53	0.37	0.23	0.29	0.14
029	0.48	0.40	0.24	0.32	0.16
030	0.03	0.57	0.25	0.08	0.04
031	0.22	0.26	0.19	0.20	0.09
032	0.52	0.35	0.23	0.37	0.18
033	0.65	0.40	0.24	0.44	0.22
034	1.00	0.21	0.17	0.47	0.19
035	0.86	0.23	0.18	0.43	0.18
036	1.00	0.29	0.21	0.51	0.23
037	0.82	0.21	0.17	0.32	0.13
038	0.69	0.18	0.15	0.38	0.15
KR	0.87				
SX2	48.86				

APPENDIX-H

Discrimination, difficulty, correlation and reliability indices for G-II and G-I English Comprehension test

Item	Dis.Value	Diff. Value	Vare	Correlation	Reliability
001	0.04	0.96	0.04	0.26	0.05
002	0.29	0.82	0.15	0.44	0.17
003	0.06	0.93	0.07	0.14	0.04
004	0.17	0.89	0.10	0.38	0.12
005	0.30	0.78	0.17	0.43	0.18
006	0.21	0.77	0.18	0.30	0.13
007	0.41	0.68	0.22	0.41	0.19
008	0.27	0.80	0.16	0.40	0.16
009	0.22	0.73	0.20	0.31	0.14
010	0.21	0.76	0.18	0.29	0.12
011	0.27	0.83	0.14	0.45	0.17
012	0.15	0.91	0.08	0.44	0.12
013	0.08	0.27	0.20	0.01	0.00
014	0.17	0.28	0.20	0.08	0.04
015	0.44	0.61	0.24	0.42	0.21
016	0.21	0.77	0.18	0.31	0.13
017	0.21	0.69	0.21	0.25	0.11
018	0.13	0.71	0.21	0.03	0.01
019	0.44	0.57	0.25	0.48	0.24
020	0.56	0.47	0.25	0.38	0.19
021	0.61	0.61	0.24	0.61	0.30
KR	0.55				
SX2	7.78				

APPENDIX - I

Discrimination and difficulty indices for G-III, G-II and G-I English and Amharic Typing and Shorthand Tests

Secretarial Position	Indices	English Typing				Amharic Typing		Shorthand	
		Straight Copy	Letters	Revised Manuscript	Tables	Straight Copy	Revised Manuscript	Letter I	Letter II
G - III	Dis.	0.54	-	0.49	0.43	0.45	0.64	-	-
	Diff.	0.46	-	0.59	0.64	0.53	0.59	-	-
G - II	Dis.	0.43	0.41	0.41	0.64	0.44	0.41	-	-
	Diff.	0.53	0.46	0.60	0.46	0.56	0.41	-	-
G - I	Dis.	0.40	0.55	0.65	0.62	0.47	0.48	0.55	0.43
	Diff.	0.51	0.46	0.56	0.64	0.52	0.41	0.52	0.66

DECLARATION

I the undersigned declare that the thesis is my original work, has not been presented for a degree in any other university and that all sources of material used for the thesis have been duly acknowledged.

Name DESALEGN CHALCHISA

Signature Desalegn Chalchisa

Place and Date of Submission School of Graduate Studies, Addis Ababa University, June 25, 1991.

manuscript, and G-I English typing letters and Amharic typing - revised manuscript. The difficulty values for these test were from 0.41 to 0.46. The relatively easy tests were G-III English typing-letters, G-II English typing revised manuscript, and G-I English typing-tables and shorthand letter I with difficulty values form 0.60 to 0.66.

G-III Amharic typing-revised manuscript, G-II English typing-tables, G-I English typing-revised manuscript and tables were with relatively high discrimination power. The discrimination indices for these tests were between 0.62 to 0.65. Relatively low discriminating tests from each group were: G-III English typing - tables, G-II English typing straight copy, letters and revised manuscript and G-I Amharic typing - revised manuscript with discrimination values from 0.41 to 0.43.

The obtained discrimination and difficulty indices for the English and Amharic typing and shorthand tests in general were not either very low or very high. The discrimination indices for these tests were found to be between 0.41 to 0.65 while the difficulty indices were between 0.41 to 0.66.