

T
831

A PRELIMINARY INVESTIGATION FOR THE CONSTRUCTION
OF A TEST OF VERBAL ABILITY FOR LEBANESE
PRIVATE SECONDARY SCHOOL BOYS
BETWEEN AGES 13-17

By

ALI MOHAMMED YAHYA

A Project

Submitted in Partial Fulfillment of the
Requirements For the Degree of Master
of Arts in the Education
Department of the American
University of Beirut, Lebanon

October, 1966

ACKNOWLEDGEMENTS

I wish to express my thanks to Dr. Naim Atiyeh, the Chairman of my project committee, for the help he gave in planning and carrying out this research, and for the personal interest he showed in this project. I am also indebted to Dr. George Za'rour and Dr. Robert Young, who were members of my committee, for their advice and directions without which this project could not have been accomplished. I also acknowledge with gratitude the help given by Dr. Frederick Korf during the statistical analysis of the results.

I wish to thank my fellow students who gave so generously of their time in scoring the test papers and school teachers and administrators for their cooperation in executing this research.

ABSTRACT

Psychological and educational measurement are two interwoven practices that complement each other. The educator is always cognizant of the need for standardized psychological measuring instruments in the school. This concern stems from the fact that psychological factors influence academic achievement. Among the former, the mental factor has the greatest impact. Through the use of a reliable scale for the assessment of mental ability, the educator can make valid judgements as to the extent of educability of a child. Mentally retarded as well as mentally superior children are thus singled out and given special treatments. If this practice is carried out at an early stage in the child's development, he can be educated up to the level of his native potentialities. The frustration and underachievement concomitant with improper placement are thus avoided.

This work was prompted by the growing need for local intelligence scales that are geared to the cultural environment in the Arab world. It serves as a pilot study that paves the way for more comprehensive research to follow. The work was confined to one aspect of intelligence namely that of verbal ability so that more exploratory studies are needed in other areas such as numerical ability and reasoning ability. Two main purposes of this project may be distinguished. First, it aimed at the construction and standardization of an Arabic multiple-choice vocabulary test for private secondary school boy students in Beirut of ages 13 - 17 years inclusive. Secondly, a correlation was run between the standardized vocabulary test scores and achievement scores (year's averages). The possibi-

lity of a relationship between vocabulary and general intelligence was then considered.

The Sample

The population comprised all Lebanese private secondary school boy students of ages 13 to 17 years. A sample of 218 boys was drawn from four secondary schools in Beirut. The inadequate number of the sample is attributed to the fact that it was assembled on an availability basis.

The Tryout

A preliminary multiple-choice test of 44 items was administered to the sample. Four administrations were necessary, one at each school. The results were then subjected to item analysis. To be eligible for inclusion in the final form of the test, an item had to comply with the following five criteria:

- 1) Increase in performance with age.
- 2) Difficulty level.
- 3) Discriminative value.
- 4) Validity.
- 5) Effectiveness of distractors.

Ten items were able to meet the requirements and they were allocated to the appropriate age levels accordingly. The vocabulary test in its final form consisted of these ten items.

Results

The means and standard deviations were calculated for the various age levels. The former showed a definite increase with age. The reliability coefficients of the test were .54, .63, .82, .67 and .82 for ages 13, 14, 15, 16 and 17 respectively. Absolute reliability was determined by finding the standard error of measurement at each age. The values obtained were high in general. Validity coefficients for each age level calculated on the basis of very scanty data ranged between .03 and .72.

The best-fitting curve was then plotted for each distribution under the assumption that vocabulary ability was normally distributed at the age levels concerned. These curves were superimposed on the original frequency polygons. T-score equivalents were calculated for the raw scores.

Vocabulary and Achievement

A simple correlation was run between the vocabulary scores and the year's averages of 84 students provided by the International College. A coefficient of .24 was obtained which is significant at the 5 per cent level.

CONTENTS

<u>Chapter</u>	<u>Page</u>
List of Tables	x
List of Figures	xi
1. The Problem	1
Background	1
The Functional Approach	2
The Factorial Approach	4
Delineation of the Problem	7
Purpose of the Study	9
2. Development and Administration of the Test	10
Format of the Test	10
Rationale for Adopting an age Scale	11
Development of the Preliminary Form	12
Administeration of the Test	13
Instructions	15
Scoring	16
The Sample	16
Representativeness of Sample	17
Randomness of Sample	18
3. Item Analysis	20
Increase in Performance with Age	20
Level of Difficulty	24
Discriminative Value	24
Validity of Items	26
Effectiveness of Distracters	32
The Final Selection	33

4. Results	40
Increase in Performance With Age	40
Reliability of the Instrument	43
Validity of the Instrument	44
Vocabulary and Achievement	46
Normalizing the Distributions	47
5. Conclusion	59
Educational Implications	60
Suggestions and Recommendations For Future Research.....	61
Bibliography	63
Appendix	66

LIST OF TABLES

Table	Page
1 Number of Students From Each School	17
2 Number of Cases Corresponding to Each Age Level	19
3 Item Analysis: Percent Passing According To Age	21
4 Item Analysis: Difficulty, Discrimination, and Validity of Items At Age Level 13	27
5 Item Analysis: Difficulty, Discrimination, and Validity of Item At Age Level 14	28
6 Item Analysis: Difficulty, Discrimination, and Validity of Items At Age Level 15	29
7 Item Analysis: Difficulty, Discrimination, and Validity of Items At Age Level 16	30
8 Item Analysis: Difficulty, Discrimination, and Validity of Items At Age Level 17	31
9 Analysis of Distracters	34
10 Frequency Distribution of Scores	41
11 Validity Coefficients At The Five Age Levels	45
12 Normalized Distribution For Age 13	49
13 Normalized Distribution For Age 14	50
14 Normalized Distribution For Age 15	53
15 Normalized Distribution For Age 16	55
16 Normalized Distribution For Age 17	57

LIST OF FIGURES

Figure		Page
1	The Relationship Between Mean Performance and Age	42
2	Observed and Normalized Distributions for Age Level 13	50
3	Observed and Normalized Distributions for Age Level 14	52
4	Observed and Normalized Distributions for Age Level 15	54
5	Observed and Normalized Distributions for Age Level 16	56
6	Observed and Normalized Distributions for Age Level 17	58

CHAPTER I

THE PROBLEM

Background:

Intelligence is one of the psychological concepts that have given rise to a vast literature of theoretical and empirical research. The nature of intelligence has always constituted the major issue in this connection. Does it have a unitary function or is it a juxtaposition of functions? Could it be broken down into a number of independent abilities or is it the over-all global capacity of the individual? And if it could be broken down, to what extent is each constituent element involved in intelligent behavior? Also, what type of instrument is capable of yielding the best estimate in the measurement of mental ability? These and others are basic questions which lie at the root of modern research.

A review of the literature reveals that a large number of definitions of the term "intelligence" has been offered, each one commensurate with the writer's conception of the term and according to his theoretical position. A brief account of these definitions is now in order. Despite the diversity in the literature, some of the more prominent schools of thought may be distinguished. These fall into two broad classifications, namely, the functional approach and the factorial approach. The former will include unitary, developmental, and behavioral definitions of intelligence while the two-factor, group-factor, and multi-factor theories will appear under the latter.

The Functional Approach:

Alfred Binet merits consideration as the first exponent of the general factor theory. He conceived of intelligence as a single ability that manifests itself in a variety of activities in life. This he called the innate or inborn capacity of the individual. The faculty underlying intelligent action is that of judgement. According to Binet, there are four levels at which intelligence functions, namely, comprehension, invention, direction and censorship. In other words, Binet maintained that intelligence is the capacity to reason well, to judge well, and to be self-critical.¹ It is not surprising then that the Binet scale largely taps the ability to carry on abstract thinking. As such, it is replete with situations that necessitate the effective use of verbal and numerical symbols.

Jean Piaget subscribed to a developmental view of intelligence. Intelligent behavior begins with sensory-motor adaptations at the early stages. These are coordinations of actions linking means and goals. When fully developed, intelligence is the coordination of operations. The latter are internalized actions that are characterized by reversibility. Development takes place through three functional invariants, i.e. assimilation, accommodation and organization. This invariant constant is present in every operation that links means and goals. In summary, all the successive adaptations of a sensory-motor and cognitive nature as well as all assimilatory and accommodatory interactions between the organism and the

¹

J.J. Jenkins and D.G. Patterson (eds.), Studies in Individual Differences, (New York: Appleton-Century-Crofts, 1961), p. 93.

environment strive to attain a state of equilibrium. This state of equilibrium¹ is intelligence according to Piaget. Piaget's contribution is that he provided principles that work. However, no intelligence scale has been constructed along these lines so far.

The behavioral approach encompasses cognitive as well as non-cognitive aspects of the individual. Intelligent action becomes a function of both intellectual and non-intellectual considerations such as drives. Prominent among the definitions in this respect are those of Wechsler and Stoddard. The former defined intelligence as the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment.² This definition brings forth the notion of incentive in intelligence. Comprehensive and sound as it is, this definition is not without its limitations. Its applicability is greatly hampered by the practical difficulty of devising an instrument that will take into cognizance temperamental as well as intellectual factors in the appraisal of intelligence.

Stoddard offered a more comprehensive definition of intelligence. He conceived of it as the ability to undertake activities that are characterized by 1) difficulty, 2) complexity, 3) abstractness, 4) economy, 5) adaptiveness to a goal, 6) social value and 7) emergence of originals; and to maintain such activities under conditions that demand a concentration of energy and a resistance to emotional forces.³ This definition is unique

¹ Jean Piaget, The Psychology of Intelligence, (New York, Harcourt, Brace and Co., 1950), p. 11.

² David Wechsler, The Measurement of Adult Intelligence, (Baltimore: Williams and Wilkins Co., 1944), p. 3.

³ George Stoddard, The Meaning of Intelligence, (New York: The Macmillan Co., 1947), p. 4.

in the sense that it touches on the mental, emotional, social and physical environments of the individual. However, it is difficult to combine and control all those elements in the assessment of one situation.

The Factorial Approach:

This is essentially a statistical approach based on factor analysis. It is mainly concerned with the psychological factors underlying the structure of intelligence and relies on the analysis of intercorrelations to achieve its purpose. A number of theories has emerged in this respect to which we now turn.

To begin with, there is Spearman's two-factor theory. He saw all mental activity as dependent primarily on a general factor possessed by all individuals in varying degrees. This "g factor" comes into play when the education of relations and correlates is called for. In view of the imperfect intercorrelations obtained among his subtests, Spearman postulated the presence of a specific or "s factor" which is operative in a particular type of activity.¹ The subtests used by Spearman were analogies, completion, understanding paragraphs, opposites, instructions, resemblances and inferences. It is clear that those tests are highly saturated with the verbal element.

Another factorial approach is that of Thurstone. He believed that certain related mental operations which he called a group, have in common a unifying primary factor. He was able to identify six primary factors each corresponding to a group of mental abilities and hence the name group-factor theory. The six factors were number factor, verbal

1

C. Spearman, The Nature of Intelligence and the Principles of Cognition, (London: Macmillan and Co., 1927), p: 5.

factor, space factor, word-fluency factor, reasoning factor, and rote memory factor.¹ However, due to the positive and significant intercorrelations among those abilities, Thurstone postulated the presence of a second-order general factor. It is noteworthy that in his Primary Mental Abilities Test, the verbal and reasoning subtests have given the highest correlations with other subtests.

Thorndike introduced an atomistic theory of intelligence known as the multi-factor theory. He broke mental ability down into minute elements. In any mental act, a number of these elements is at work. Further statistical analysis led Thorndike to believe that more than a certain number of these elements is involved in any one mental act. He was able to distinguish four distinct types of activity, namely sentence completion, arithmetical reasoning, verbal ability, and following directions.² It must be remembered that those four acts are not all-inclusive of mental activity since more could still be added on further analysis.

A look into the above definitions and interpretations of the term intelligence is enough to convince us that there is evidence of some unity-within-diversity despite the apparent lack of agreement. The first point which is of paramount significance to this research is that almost all intelligence tests, regardless of the psychological constructs they claim, are to some extent loaded with the verbal element. Even in the so-called performance tests, instructions and explanations have to be given in words. Hence, ability to deal effectively with verbal symbols becomes a pre-requisite for

¹ Frank S. Freeman, Theory and Practice of Psychological Testing (New York: Holt, Rinehart and Winston, 1965), pp. 164-65.

² Ibid., pp. 160-61.

a reliable performance in those tests. Most tests now recognize the importance of verbal ability and subtests measuring this function yield the highest correlations with other subtests and with the tests as wholes. The Stanford-Binet and the Wechsler scales are a case in point.

The second point is that they all seem to converge towards a general factor which underlies intelligent action. Except for the differences in nomenclature, Binet's general inborn capacity, Piaget's invariant constant, Spearman's general factor, Thurstone's reasoning factor, Wechsler's global capacity and Stoddard's general ability - all appear to be different ways of designating the same thing. This conclusion is also supported by the fact that tests designed to measure intelligence on the basis of different psychological assumptions have been found to be highly intercorrelated. It is logical to ask what common ground do those diverse instruments share? Whatever factors may be active to produce this relationship, it must not be forgotten that those tests have much in common by virtue of their verbal composition. It would seem that each item on those tests measures verbal ability besides its proposed function.

Two questions are now raised. Since verbal ability plays such a major role in intelligence, could it yield by itself a reliable index of the latter? The second question is what are the criteria of a good test of verbal ability that could be used in this connection? These two issues provide the backbone of this work and an attempt to answer them has been made.

Delineation of the Problem:

Whether intelligence is defined as adaptation to the environment or as the ability to learn or as the ability to abstract, the role of communication cannot be overlooked. It is through symbols and in particular verbal symbols that ideas and concepts are conveyed. Only to the extent that those symbols are clear in our minds can we translate them into behavioral terms. Success in any field in our complex age entails mastery of symbols. Until proficiency in their use is acquired, no step can be taken in the direction of their application in practice. If the wrong concepts are formed, then the wrong performance is bound to follow. Consequently, any training which is not well grounded in a clear, correct conceptualization of symbols is futile.

These statements are not without their implications for educational practice. Ebel sees the development of a student's command of useful verbal knowledge as the principal function of formal education. Direct experience is essential as a foundation, but it provides little more than the foundation. It is the verbal knowledge, through which those experiences are integrated, interpreted and communicated that makes possible the intelligent behavior of a human being and the culture of a human society.¹ Hence, the need becomes obvious for good instruments that measure this function adequately. This need is keenly felt here in the Arab World where few attempts have been made in this direction.

Current intelligence scales employ a number of subtests such as analogies, classifications, sentence completion, vocabulary, similarities,

¹

Robert L. Ebel, Measuring Educational Achievement, (New Jersey: Prentice-Hall, 1965), pp. 40-41.

and opposites in the measurement of verbal ability. Among these, vocabulary tests rank highest. Their peculiar merit lies in the fact that they have proved to be indicative of the level of intelligence of the testee. This is not surprising since the high correlations vocabulary tests show with intelligence scales attest to the significance of their role in the mental process.

These conclusions are supported by experimental evidence. On the basis of twenty-one studies utilizing samples covering a wide range of ability and employing vocabulary tests of intermediate length, Miner has found the median correlation between those tests and well-known measures of intelligence to be ¹.83. This result indicates that a vocabulary test is as good a predictor of general intelligence as any other full scale. In the 1937 revision of the Stanford-Binet, Terman and Merrill describe the vocabulary test as the most valuable single test in the scale. Besides its interest value, it presents a familiar task to the subject. More important still is the fact that it begins with words in common use and increases rapidly in difficulty, thus giving the examiner a quick survey method of estimating the subject's ability.² Wechsler also considers the size of a man's vocabulary to be an excellent measure of his general intelligence and believes that apart from its value as a measure of intelligence, the vocabulary test is desirable because of its qualitative possibilities.³ It would seem then that the vocabulary test has the double advantage of measuring intelligence as well as verbal ability.

1

John B. Miner, "On the use of a short vocabulary test to measure general intelligence", *J. of Educ. Psychol.*, LII (June, 1961), 157.

2

L.M. Terman & M.A. Merrill, Measuring Intelligence, (Boston: Houghton Mifflin, 1937), p. 302.

3

David Wechsler, The Measurement of Adult Intelligence, (Baltimore: Williams and Wilkins, 1944), p. 258.

Purpose of the Study:

This research was urged by the acute need for standardized intelligence tests in the Arab World that are geared to the cultural environment. In this sense, this work paves the way for further research and it serves as a pilot study that lays the foundations for the development of a local intelligence scale in Lebanon. More exploratory work is still needed on other aspects of intelligence such as numerical and reasoning abilities. This study was also directed towards one facet of verbal ability, namely, that of vocabulary so that other constituents of this ability were not included. This limitation, though regrettable was inevitable due to time pressure.

In sum, the purpose of this work was two-fold:

- 1) In the first place, it aimed at the construction, administration, and standardization of a verbal ability(vocabulary) test in Arabic for Lebanese private school boy students from the age of 13 through 17. Words and phrases in the test were allocated to appropriate age-levels on the basis of their difficulty, discrimination indices and their correlation with total test scores. Items that did not meet the necessary conditions were deleted from the test in the statistical analysis. The reliability and validity of the test in its final form were then determined. This, in brief, was the primary objective of this study.
- 2) A second and minor purpose of the study was to examine if a positive relationship existed between verbal ability, as measured by the vocabulary test, and academic achievement. Achievement records were only available at one school namely the International College. A simple correlation was run between test scores and the year's averages of the students who took the test in this school only.

CHAPTER II

DEVELOPMENT AND ADMINISTRATION

OF THE TEST

This chapter depicts the steps and procedures that were followed in the construction and administration of the vocabulary test. An account is also given of the method employed in the selection of subjects and of the limitations of the sample. Statistical analysis of the results will be deferred to later chapters. This will center around item analysis together with the reliability and validity of the test in its final form.

Format of the Test:

It seems to be the general agreement of test specialists that vocabulary questions give the best index of vocabulary ability when they are cast into the multiple-choice form. Outstanding among the works in this respect is an early experimental study conducted by Kelley. On using various techniques for testing word meaning and then subjecting the results to statistical treatment, Kelley reported, that multiple-choice, together with matching questions, proved to be the best measures of this function. Also, group tests of intelligence such as the Chicago Primary Mental Abilities Test and the Lorge-Thornlike use multiple-choice questions in their vocabulary subtests. Therefore, it was decided to follow suit in the construction of this test for the gain in objectivity and reliability inherent in this method.

¹ V.H. Kelley, "An Experimental Study of Certain Techniques for Testing Word Meaning", J. Educ. Res., XXVII (December, 1933), 277-282.

As already stated, the test was in Arabic and it comprised 44 items. Each word constituted an item on the test and five options were included to minimize the possibility of chance success. The distracters were phrased so that they were equally attractive and plausible to the inept student. Homogeneity and avoidance of specific determiners were observed in writing the options. However, two errors were committed in typing the preliminary form of the test. The first was that the stem of each item should have been in question form so that a problem was raised. The second error was that the responses should not have been arranged in tandem. This would make it tiring for the eye to follow through and hence to single out the correct answer. Ideally, each option should have occupied a separate line.

Rationale for Adopting an Age Scale:

Studies in the relationship between vocabulary and chronological age show that the former increases steadily until maturity is reached and then levels out. Once its peak level is attained, it tends to be stable. Lewinski found that there were no well-defined increments or decrements in vocabulary ability with increasing chronological age when adulthood was reached.¹ This stability in vocabulary is believed to take place roughly between ages 17 and 18. Since the scale of this study extended for ages 13 through 17, it was well within the growth interval of this ability. During this period, vocabulary ability could serve as a predictor of mental

1

Robert J. Lewinski, "Vocabulary & Mental Measurement", Journal of Genetic Psychology, LXXII (June, 1948), 270.

ability which is a correlate of chronological age.

In addition, the adoption of an age scale facilitated the establishment of norms for the ages already indicated. The items allocated to each age level formed a subtest. The mean and standard deviation were computed for each age level. This had the obvious advantage of inter-age comparison in addition to providing a rapid estimate of ability.

Development of the Preliminary Form:

In view of the fact that current vocabulary tests were standardized for different cultural environments, a fresh start was necessary if a local scale was to be developed in Lebanon. Any attempt at borrowing or duplication was predestined to failure because it overlooked the needs of the cultural milieu. Since those tests are mostly verbal, the language difficulty constitutes a formidable obstacle to effective performance on them. With these points in mind, the vocabulary test was constructed.

A number of Arabic readers and textbooks assigned this year to the first through the sixth grades (ages 12 to 18 years) were scanned by the examiner. Words were randomly selected for inclusion in the test. One criterion for selection was that the rarity of the concept rather than the difficulty of the word was sought. Another criterion was that words should be in current use rather than **obsolete**. Two hundred words were chosen in this way. These were of graded difficulty so that all age levels were taken into consideration. Further scrutiny and consultation with experts in psychological testing led to the elimination of words that were not rationally valid with the result that only forty eight items were retained.

The next step was to construct the test. Words were looked up in Arabic dictionaries to ascertain their correct meaning. This meaning constituted the correct response and it was randomly placed among the distracters. At this stage, the test was administered to five Arab university students with a view to evaluating the validity of the responses. Their scores ranged between forty-three and forty-five. When a discussion of the test was held soon afterwards, they unanimously agreed that four of the words were controversial since they could lend themselves to different meanings. Those words had to be deleted. They also made valuable suggestions for the improvement of some of the distracters. The test was then ready for the try-out.¹

Administration of the Test:

The test was administered to a sample of 218 students drawn from four secondary schools in Beirut. In order to establish rapport with the subjects and to remove fear and anxiety, the purpose of the test and its non-evaluative nature were made clear to the testees at the outset. Some of the anticipated outcomes of the study were made known to them to stimulate their interests and to enlist their cooperation. This preparatory step was usually made one or two days before each administration.

The role of extraneous factors in test performance is not to be underestimated. The moderate correlation between intelligence and achievement attests to the influence of those factors in learning. As such, a perfect estimate of the ability measured is rendered difficult. It would seem that three capacities function within the same individual; the innate

¹ A copy of the preliminary form of the test is included in the appendix.

capacity, the actualized capacity and the measured capacity. It is with this last capacity that testing is concerned and it is clear that if it is not supplemented by native and actualized potentials, errors are bound to emerge in the process of measurement.

Factors that affect test performance may be broadly classified into physical and psychological factors. The former may arise on the one hand from some organic defect or physical handicap within the individual himself, and on the other hand the environmental physical conditions may not be favorable to test administration. Therefore, it is incumbent on the examiner to create a climate in which the impact of those factors is brought down to a minimum if not totally eliminated.

Before administering the test, arrangements were made to secure well-ventilated and large rooms with well-spaced seats. The rooms were to be away from any source of distraction. This request was kindly granted by all school authorities. Three of the schools where the test was given did not keep health records but it was clear that no student was suffering from a serious illness.

Unlike the physical factor, the psychological factor is not easy to control especially in group tests. The preparatory steps and introductory written statements made before the test were intended to ease emotional tension accruing from test anxiety, but some psychological disorders are so deeply embedded in personality dynamics that their origins have to be traced to the home or elsewhere. Such trends cannot be easily detected from group tests where the examiner's attention is divided among so many individuals and where his role is reduced to that of maintaining surveillance over testees. In such cases, individual tests are preferred where referral

to specialists, e.g. psychologists, can be made on the basis of a thorough clinical diagnosis. Therefore, the psychological factor is apt to be reflected to some extent in test performance.

Instructions:

When instructions are not clearly and lucidly stated, ambiguities and misconceptions are bound to ensue. Such a state of affairs is detrimental to effective performance on any test. For the sake of the validity of the results, the test maker should take pains to ascertain that every examinee understands what is expected of him.

Students in three of the four schools in which the test was conducted were not familiar with the multiple-choice type of questions. Hence it was thought necessary to illustrate with an example on the board before each administration. Some time was also spent on explaining how to use the separate answer sheets. The examinees were encouraged to raise questions whenever they were in doubt.

Some other important points were brought to the students' attention. First, it was emphasized that to every question there was only one correct answer. Secondly, students were urged to attempt all questions, Thirdly, their attention was drawn to some misprints that appeared under questions 14, 17, 34 and 35 in the preliminary form of the test.

The time allotted for the test (1 1/4 hours) may seem to be more than is necessary for a test of 44 multiple-choice items. However, since the results of this preliminary testing were to be a basis for item analysis, the test should be given with quite ample time limits so that most individuals would have a chance to try all items.¹ It was observed that

¹
R.L. Thorndike, Personnel Selection, (New York: John Willey & Sons, 1949), p. 53.

the slowest student was able to finish in one hour and that all the items were tried by every examinee. The time limits of the test in its final form were set in the light of these observations.

Scoring:

Examinees were instructed to do all work on the separate answer sheets. They were to respond by crossing the letter which preceded the answer they believed to be correct. ¹ A key for scoring was prepared by punching holes on an answer sheet. Each punch corresponded to a correct answer and an answer sheet was marked by placing it against this key and counting the number of correct responses.

The separate answer sheet proved to be useful in many respects. It had the advantage of being scored quickly, and helped to facilitate the analysis of test items. Also, the same question papers could be reused with no need for further duplications.

The criterion for scoring was that each response was awarded a score of one or zero depending on whether it was right or wrong so that the maximum score on the test was 44. The same test was machine scored later and a coefficient of correlation of .98 was found which was high enough.

The Sample:

For the purpose of this research, the population comprised all Lebanese private secondary school boys students in Beirut of ages 13 through 17 years. This restriction should be observed so that the results of this study are not extended to government schools and to rural and suburban areas in Lebanon. Exploratory work along similar lines needs to be carried out in those parts of the country. Government and girl schools must also be taken into account in any future research.

1

A copy of the separate answer sheet is included in the appendix.

The sample was drawn out of four secondary schools in Beirut and it totalled a number of 218 students as seen in table 1.

TABLE 1

Number of Students From Each School

School	No. of Students
International College	110
Beirut Al-Watania	48
Ibn Sina	45
Ras Al-Nab'	15
Total	218

This distribution was not planned ahead of time but it was imposed by some administrative considerations as will be seen later.

Representativeness of Sample:

The inadequacy of the sample was evident as far as representativeness was concerned. Private girl-schools and sectarian schools as well as others should have been taken into account. Clearly, the sample was biased towards certain private secondary schools in Beirut. While representativeness was acceptable as a principle of sampling, it was difficult to implement in practice. This was because it presupposed a knowledge of the char-

acteristics of the population. Therefore, the principle of randomness had to be used instead.

Randomness of Sample:

A random sample is defined as one in which every member of the population has an equal chance of being included. Such a sample falls within the range of random sampling errors of being representative with respect to the trait under investigation. However, the effective implementation of this principle was impaired by some practical and administrative difficulties.

First, schools were chosen on the basis of availability rather than on a random basis. This line of action was dictated by the fact that some of the schools contacted were antagonistic to the idea of the test. Hence, the consent of a school was the sole basis for its inclusion in the sample. Out of seven schools only four showed signs of cooperation.

Even in those four schools, the plan of sampling proposed by the writer proved to be in conflict with the school schedules. This plan consisted in preparing lists of the names of the students according to chronological age, and then selecting every third name on each of those lists, but school authorities were in favor of whole class administrations for convenience sake. They also set the number of students who were to participate in each administration of the test. As a result, the sampling adequacy suffered appreciably at some age levels particularly 12 and 18 which had to be discarded. The following table shows the number of cases corresponding to each age level in the interval 13 - 17 years.

TABLE 2
Number of Cases Corresponding
To Each Age Level

Age in Years to the Nearest Birthday	No. of Cases
13	53
14	41
15	32
16	36
17	42
Total	204

A sample of 400 or more was planned for this study in the first place. It was to encompass ages 12 through 18 years. This was not possible to accomplish for two reasons. The first was that some schools were not willing to provide more than a specified number of students. The second reason was that part of the testing was carried out during summer when the size of classes is small. In view of these considerations, it was thought more appropriate to concentrate on a smaller age interval, namely, 13 - 17 years.

CHAPTER III

ITEM ANALYSIS

The process of item analysis is to be considered as an integral part of the reliability and validity of the test. It was conducted with a view to examining the degree of compatibility of each of the items with the purpose of the test. The aim of the analysis was to allocate items at the various age levels on the basis of (1) the level of difficulty of each item, (2) the discriminative value of each item; and (3) the effectiveness of the distracters. Items that fell short of meeting the necessary standards were subject to deletion from the final form of the test. The items in the preliminary form of the test will now be studied in the light of the three criteria above.

Increase in Performance with Age:

If the proposed scale is to possess practical utility, it should be able to differentiate among the different age levels. A higher level of performance was to be expected of the 17 year old rather than of the 13 year old. Therefore, priority was given to items that possessed this characteristic. Items that revealed fluctuations in performance from age to age were thus discarded.

Table (3) shows the percent passing each item at each age level. It will be observed that less than half the items in the test registered

TABLE 3

Item Analysis: Percent Passing According To Age

	Item	Age 13	Age 14	Age 15	Age 16	Age 17
	1	98	95	88	89	98
	2	94	98	88	89	98
	3	90	93	81	89	95
	4	94	93	94	94	100
	5	90	90	91	81	100
	6	87	90	94	97	95
	7	83	88	91	89	93
	8	85	93	91	92	86
	9	42	30	41	31	60
	10	71	83	69	92	93
	11	100	93	91	97	95
x	12	69	90	88	81	93
x	13	10	13	16	28	26
	14	73	73	72	83	79
	15	90	83	100	92	88
x	16	38	40	47	53	72
x	17	33	35	50	56	71
	18	90	90	88	89	93
	19	90	90	88	92	93
	20	90	83	81	97	90

TABLE 3 (Continued)

Item Analysis: Percent Passing According To Age

	Item	Age 13	Age 14	Age 15	Age 16	Age 17
	21	87	85	88	86	90
	22	75	83	53	56	52
	23	81	85	69	81	83
	24	48	53	41	44	52
	25	23	28	28	28	17
	26	33	60	59	56	55
x	27	27	30	56	50	65
x	28	13	40	38	72	74
x	29	56	55	53	64	71
	30	33	35	28	44	31
x	31	35	45	47	39	55
x	32	42	60	63	86	76
x	33	33	48	53	44	57
x	34	57	60	63	78	81
	35	13	30	9	36	45
x	36	27	33	47	50	62
x	37	33	45	56	58	74
x	38	13	20	19	33	38
	39	63	60	50	64	62
	40	12	28	9	47	49
x	41	50	53	63	86	81

TABLE 3 (Continued)

Item Analysis: Percent Passing According To Age

	Item	Age 13	Age 14	Age 15	Age 16	Age 17
x	42	31	43	44	53	71
x	43	48	60	55	58	83
	44	38	35	13	56	50

x Items marked with a cross show increase in performance with age. A difficulty index for the total group was found for each item by averaging the difficulty values obtained at the five age levels. The difficulty index was 84.2 for item 12, 18.6 for item 13, 50 for item 16, 49 for item 17, 45.6 for item 27, 47.4 for item 28, 59.8 for item 29, 44 for item 31, 65.4 for item 32, 47 for item 33, 47.8 for item 34, 43.8 for item 36, 47.2 for item 37, 24.6 for item 38, 66.6 for item 41, 48.4 for item 42, 60.8 for item 43.

steady increases in performance with age. The discrepancy in the rest of the items may be attributed to the inadequacy of the words sampled and also to some extent to the shortcomings of the population sample. Items that failed to differentiate among ages were excluded from further statistical analysis.

Level of Difficulty:

The difficulty level of an item for a specific age group was determined by the percent of that age group passing the item. The number of discriminations among examinees is believed to be greatest when an item of fifty per cent difficulty is used. This is because the variance is maximized at this level. The test-score reliability is also increased to its maximum value when this condition is fulfilled.¹ In view of these considerations, items of fifty per cent difficulty for a certain age were considered eligible for inclusion in the final form of the test. However, this was difficult to accomplish in practice; so the range was extended to encompass items of seventy percent difficulty. The selection so far was only tentative since other requirements had to be satisfied before the final decision was made.

Discriminative Value:

Besides having a reasonable difficulty level, an item should be able to discriminate satisfactorily between able and weak subjects on the criterion variable. Unless this is achieved, a test can hardly be said to have practical value. Such a defect is detrimental to any selection program. It is to be noted that an item of fifty percent difficulty may be answered correctly by a majority of inept testees. Such an item should be eliminat-

1

E.F. Lindquist (ed.), Educational Measurement (Washington: American Council on Education, 1951), p. 292.

ed if the instrument is to function properly. On the other hand, an item that bears no relationship to the test as a whole is not a good item regardless of its difficulty level or discriminative index. It should measure what the test is measuring.

The discriminative value of an item was investigated with reference to three criteria, namely, (1) its discriminative index, the total test scores being taken as the external criterion; (2) its correlation with the total test scores; and (3) its standard deviation. These three procedures were carried at each of the five age levels separately. They will now be discussed in detail in their order of presentation.

The discriminative index of an item at a certain age level was calculated by using the Phi coefficient (Guilford's formula*) on the basis of a fifty-fifty dichotomy. This was done in view of the inadequate number of cases at the various age levels. The use of the top and bottom 27 percent would have been feasible had a larger sample been assembled. The significance of the Phi coefficient at the different ages was then obtained. For significant discriminations at the 5 percent level, Phi coefficients were found to be .27, .31, .35, .33 and .30 for ages 13, 14, 15, 16 and 17 respectively.

The values of Phi appear under column (3) on each of tables (4) through (8). Each table represents an age level. The coefficients were only computed for items that indicated progressive increase in performance with age. An inspection of the tables reveals that some of the selected items failed to discriminate between subjects within the various age levels in spite of the fact that they complied with the criterion of increase with age.

¹
* The formula used was $X_1 = N\phi^2$

A list of the formulas used is included in the Appendix.

A point biserial correlation was then run for the scores on each item reduced into a dichotomy of rights and wrongs with the total test scores at each age level. This was done to ensure that the item was measuring the same function as the whole test. The higher the correlation coefficient, the more consistent the item was with the test. This process contributes to the validity and reliability of the whole test in the sense that the correlation coefficients may be thought of as dependable estimates of item reliability and internal validity of the test.

The point biserial correlation coefficients are shown under column(4) of tables (4) through (8). On the whole, it would seem they were not high enough due to the errors inherent in both the test and the sample. No rule of thumb was followed in the interpretation of those coefficients. The higher the coefficient, the more suitable the item was for placement at a certain age level.

The standard deviation of each item was then calculated at each age level with the purpose of examining the degree to which the scores were spread out for the particular item at the different age levels. The greater the value of this statistic, the more indicative it was of variability. The values of the standard deviations appear under column(5) of tables (4) through (8).

Validity of Items:

An attempt was made to establish the validity of each item on the basis of very limited data. These data were available only at the International College and for a limited group of subjects numbering 84. Since the test was in Arabic, grades in Arabic of the examinees constituted a reasonable external criterion for item validity purposes. Those students who did

TABLE 4

Item Analysis: Difficulty, Discrimination, and Validity
of Items at Age Level 13

Item	Difficulty Level	Discriminative Index	Correlation With Test Scores	Standard Deviation	Validity Coefficient
12	69	.17	.30	.47	.10
13	10	.33	.27	.29	.00
16	38	.12	.28	.50	.25
17	33	.45	.40	.47	.21
27	27	.09	.10	.45	-.15
28	14	.06	.08	.36	-.17
29	56	-.19	.03	.50	-.09
31	35	.16	.35	.47	.14
32	42	.31	.42	.49	.16
33	33	.20	.26	.47	.26
34	57	.45	.48	.47	.49
36	27	.26	.18	.45	.02
37	33	.45	.40	.47	.22
38	13	.28	.22	.34	.04
41	50	.62	.65	.50	.21
42	31	.08	.04	.47	.28
43	48	.42	.33	.50	.01

TABLE 5

Item Analysis: Difficulty, Discrimination, and Validity
of Items at Age Level 14

Item	Difficulty Level	Discriminative Index	Correlation With Test Scores	Standard Deviation	Validity Coefficient
12	90	-.17	.06	.33	-.32
13	13	-.08	.07	.33	.10
16	40	.10	.20	.50	.21
17	35	.42	.46	.48	.35
27	30	.22	.22	.45	.10
28	40	.20	.46	.49	.13
29	55	.11	.16	.50	-.12
31	45	.30	.44	.50	.09
32	60	.41	.59	.49	.20
33	48	.25	.32	.50	.19
34	60	.61	.71	.49	.33
36	40	.27	.10	.47	.02
37	45	.52	.52	.48	.11
38	20	.38	.19	.40	.10
41	53	.36	.38	.49	.05
42	43	.25	.44	.49	.35
43	60	.20	.35	.49	.31

TABLE 6

Item Analysis: Difficulty, Discrimination, and Validity
of Items at Age Level 15

Item	Difficulty Level	Discriminative Index	Correlation With Test Scores	Standard Deviation	Validity Coefficient
12	88	.19	.25	.33	-.02
13	16	.43	.57	.36	.25
16	47	.44	.47	.50	.07
17	50	.50	.52	.50	.32
27	56	.17	.25	.50	-.10
28	37	.26	.15	.48	.03
29	53	.19	.11	.50	.29
31	47	.06	.10	.50	.04
32	63	.52	.60	.48	.19
33	53	.19	.32	.50	.02
34	63	.52	.56	.48	.61
36	47	.56	.67	.50	.26
37	56	.50	.63	.50	.19
38	19	.16	.18	.39	-.15
41	63	.69	.68	.50	.27
42	44	.76	.79	.50	.41
43	55	.35	.28	.45	.17

TABLE 7

Item Analysis: Difficulty, Discrimination, and Validity
of Items at Age Level 16

Item	Difficulty Level	Discriminative Index	Correlation With Test Scores	Standard Deviation	Validity Coefficient
12	81	-.07	.17	.40	.00
13	28	.25	.24	.45	-.09
16	53	.39	.40	.50	.37
17	56	.56	.56	.50	.29
27	50	.22	.15	.50	.11
28	72	.62	.55	.45	.23
29	74	.40	.40	.48	.04
31	39	.11	.03	.49	.11
32	86	.40	.44	.35	.10
33	44	.45	.42	.50	.44
34	78	.40	.33	.42	.02
36	50	.22	.11	.50	.27
37	58	.28	.35	.49	.30
38	33	0	.11	.47	-.31
41	86	.24	.41	.35	.08
42	53	.61	.56	.56	.48
43	58	.51	.36	.49	.16

TABLE 8

Item Analysis: Difficulty, Discrimination, and Validity
of Items at Age Level 17

Item	Difficulty Level	Discriminative Index	Correlation With Test Scores	Standard Deviation	Validity Coefficient
12	93	.09	.05	.26	.14
13	26	.38	.43	.44	.32
16	72	.39	.57	.49	.48
17	71	.53	.45	.45	.20
27	65	.24	.54	.50	.03
28	74	.38	.58	.44	.10
29	71	.42	.29	.45	.06
31	55	.05	.19	.50	-.04
32	79	.34	.41	.43	.06
33	57	.48	.59	.49	.23
34	81	.24	.35	.39	.09
36	62	.29	.43	.49	.08
37	74	.38	.62	.44	.10
38	38	.10	.21	.49	.06
41	81	.24	.39	.39	.01
42	71	.42	.58	.45	.16
43	83	.45	.60	.37	.30

well in Arabic courses in school were the ones expected to show a high level of performance on the test and vice versa. Validity coefficients were computed for all the selected items at each age level by using a point biserial correlation and they are shown under column (6) of tables (4) through (8). In general, the coefficients tended to be low. This was expected since they were based on a limited number of subjects; but they indicate that a positive relationship exists between the two variables.

Effectiveness of Distracters:

A good multiple-choice item should favor subjects who are able with respect to the trait in question. With this fact in mind, the distracters must be so phrased as to lead astray the weak subject. A distracter is considered effective in-so-far as it appeals to more of the inept testees. Distracters that do not tally with this principle are evidently misplaced. As such, they hinder the proper functioning of the item. The following analysis is done with a view to locating good and bad distracters with the intention of adapting and modifying the latter for future use.

The subjects at each age level were divided into a high and a low group on the basis of their test scores. A fifty-fifty split was thus made. The responses of the two groups to a particular distracter were counted for each group separately and then added up for all age levels. The omits also received the same treatment. This procedure was repeated for each item and the results were then entered into table (9). The upper row in front of an item in the table gives the number of responses of the high-scoring group to each of the five options while the lower row does the same for the low-scoring group. Correct responses are marked in order to distinguish them from distracters and omits.

Table (9) includes all the items in the preliminary form of the test. It is first noted that some distracters attracted more students of the two groups than the correct responses. This is clear in items 13, 25 and 38. Those distracters must be completely discarded and the correct responses have to be revised. In item No. 40, more students of the low-scoring group responded correctly to the item. Therefore something should be done about the right answer to make it more appealing to the higher group. Some distracters like options 4 & 5 in item No. 27 are attracting more of the good students. Hence, they have to be modified. In item No. 12, options 1 and 4 are examples of non-functioning distracters because they are drawing equal numbers from the two groups. They should be replaced by others. Distracters, such as options 1 and 4 in item 1 with no respondents, should also make way for better ones.

The modification and adaptation of distracters is not without its returns, because the efficiency of distracters enhances the discriminative value and difficulty of an item. So, an item that was not originally good may be rendered desirable by adjusting its distracters.

The Final Selection:

The items included in the final form of the test were the satisfactory ones from the standpoint of increase in performance with age, difficulty level, discriminative value, validity and effectiveness of distracters. To the extent that an item satisfied these criteria, it was allocated to the appropriate age level. Items that were deficient in the characteristics outlined were deleted from the test. It must not be concluded, however, that the chosen items constitute a full scale in the proper sense of the word. They only set the pattern for more elaborate

TABLE 9

Analysis of Distracters

Item	Omits	Option 1	Option 2	Option 3	Option 4	Option 5
1	0	0	2	99 ^x	0	0
	3	0	4	91	0	3
2	0	1	98 ^x	1	1	0
	3	3	91	3	0	1
3	0	1	99 ^x	0	1	0
	2	9	83	3	2	2
4	0	0	0	0	1	100 ^x
	5	1	1	1	1	92
5	0	96 ^x	1	0	4	0
	4	87	0	0	9	1
6	0	2	98 ^x	1	0	0
	2	9	88	1	0	1
7	0	6	0	2	0	93 ^x
	2	11	1	2	0	85
8	0	3	96 ^x	1	0	1
	4	8	85	3	1	2
9	2	37	11	0	50 ^x	1
	7	44	12	3	33	2

TABLE 9 (Continued)

Analysis of Distracters

Item	Omits	Option 1	Option 2	Option 3	Option 4	Option 5
10	0	0	1	0	95 ^x	5
	3	3	1	0	69	25
11	0	1	100 ^x	0	0	0
	0	5	93	0	2	1
12	0	2	86 ^x	3	9	1
	2	2	82	5	9	1
13	1	56	1	6	28 ^x	11
	4	60	7	9	8	13
14	0	2	8	0	1	90 ^x
	2	7	27	1	1	63
15	0	1	3	92 ^x	4	1
	1	0	6	90	2	2
16	1	19	6	8	66 ^x	1
	7	38	9	5	39	3
17	0	9	72 ^x	3	1	16
	10	15	25	15	9	27
18	1	96 ^x	3	0	1	0
	1	86	9	0	2	3

TABLE 9 (Continued)

Analysis of Distracters

Item	Omits	Option 1	Option 2	Option 3	Option 4	Option 5
19	0	2	96 ^x	1	1	0
	0	4	87	2	8	1
20	0	1	0	2	95 ^x	3
	2	7	2	4	84	2
21	0	0	4	94 ^x	0	3
	2	1	11	82	2	3
22	1	5	1	8	14	72 ^x
	3	11	2	13	13	59
23	0	2	94 ^x	3	1	1
	4	15	68	6	3	5
24	0	51 ^x	0	3	43	4
	2	46	2	9	38	4
25	1	66	4	25 ^x	5	0
	3	46	2	24	24	2
26	3	21	10	2	9	58 ^x
	5	26	12	5	7	46
27	4	20	51 ^x	2	15	9
	8	17	34	12	9	21

TABLE 9 (Continued)

Analysis of Distracters

Item	Omits	Option 1	Option 2	Option 3	Option 4	Option 5
28	3	32	0	2	59 ^x	5
	2	24	11	6	30	5
29	2	68 ^x	24	2	4	4
	3	52	35	6	1	3
30	10	3	8	48 ^x	14	17
	10	13	12	21	15	30
31	1	11	1	2	36	50 ^x
	1	15	5	4	38	38
32	0	10	0	9	0	82 ^x
	1	32	3	13	5	52
33	0	5	62 ^x	23	8	3
	4	4	31	46	10	6
34	1	0	0	4	91 ^x	5
	3	2	14	16	50	16
35	2	9	37 ^x	7	22	24
	1	11	17	24	19	29
36	1	5	16	58 ^x	15	6
	5	16	26	28	19	7

TABLE 9 (Continued)

Analysis of Distracter

Item	Omits	Option 1	Option 2	Option 3	Option 4	Option 5
37	0	77 ^x	11	6	0	7
	2	36	27	23	3	10
38	0	25	40	0	32 ^x	4
	1	40	32	5	17	6
39	0	2	0	9	5	75 ^x
	1	6	3	20	14	47
40	4	30	27 ^x	11	11	18
	2	18	29	21	13	20
41	2	0	6	3	3	87 ^x
	2	6	39	12	6	46
42	0	67 ^x	12	4	16	2
	2	29	22	6	32	10
43	0	8	8	1	82 ^x	2
	3	22	21	3	46	6
44	0	17	17	58	5	4
	6	19	47	21	4	4

research to be conducted in future.

Out of the 17 items that showed increase in performance with age, ten only were able to meet the other requirements. The choice was made after a thorough inspection of the available data. The allocation of items to the different age levels was made mainly on the basis of the levels of difficulty of the items, their discriminative reliability and validity indices. The reliability index is defined as the product of the item standard deviation and the coefficient of correlation of the item with the total test scores. Similarly, the validity index is defined as the product of the item standard deviation and the validity coefficient of the item. Item 34 was allocated to age level 14; items 32, 36, 37, and 41 to age level 15; items 17, 42, and 43 to age level 16; and items 16 and 33 to age level 17. Tables (4) through (8) show the items assigned to the different age levels with the data that were used in the selection.

CHAPTER IV

RESULTS

The ten selected items constitute the content of the test in its final form. A few questions are now posed as to the dependability of the new instrument. Does the mean performance on those ten items show a definite increase from age to age? How reliable is the instrument at the various age levels? To what extent does it agree with the validation criteria? Is there any evidence of normal distributions at the different age levels? And last but not least, is vocabulary ability as measured by this instrument a correlate of academic achievement? These questions are now taken up in turn on the following pages of this chapter.

Increase in Performance With Age:

The answer sheets were scored again for the ten items only. A response was awarded a score of one or zero so that the maximum score on the test was ten. Frequency distributions were set up for the five age levels taken separately. Those distributions are shown on table (10). The means and standard deviations were computed from this table.

It may be observed from the table that the means indicate a steady increase from one age to another. The relationship between average performance on the ten selected items and age is linear and positive as represented in fig. (1). As for the standard deviations, they are not in agreement showing that the scores were not well spread out at some age levels. An attempt will be made to normalize the distributions later.

TABLE 10

Frequency Distribution of Scores

Score	F R E Q U E N C Y				
	Age 13	Age 14	Age 15	Age 16	Age 17
10	0	0	3	3	9
9	1	4	5	3	9
8	3	4	2	7	8
7	2	3	2	4	4
6	8	7	4	6	1
5	9	7	5	3	2
4	11	6	2	5	1
3	4	3	1	3	5
2	9	4	3	1	1
1	4	2	4	1	2
0	2	1	1	0	0
Mean	4.11	5.09	5.46	6.22	7.19
S.D	2.12	2.36	3.06	2.33	2.75

THE RELATIONSHIP BETWEEN MEAN PERFORMANCE AND AGE

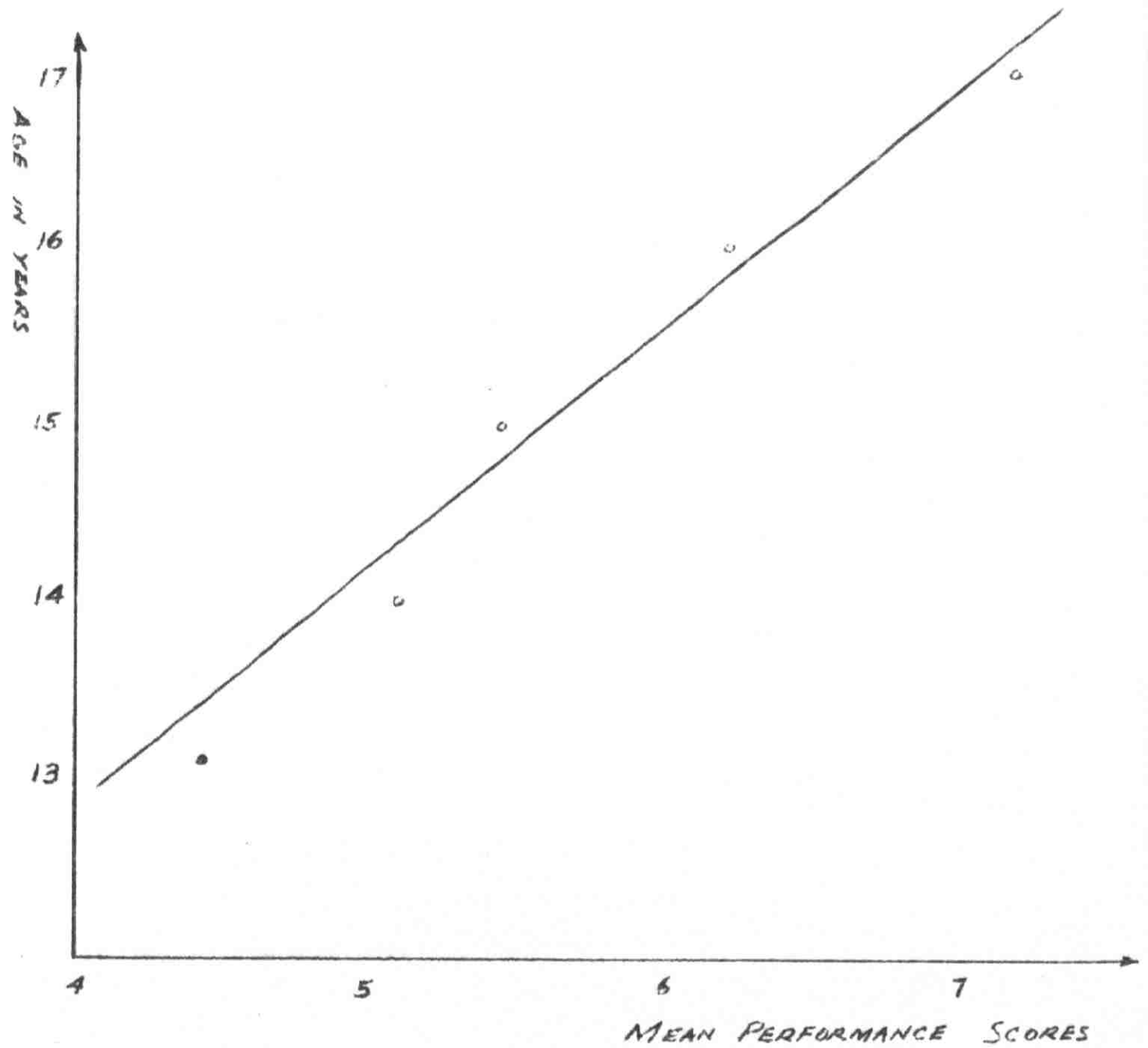


FIG. 1

Reliability of the Instrument:

The reliability of a test is defined as the degree of consistency with which the test measures what it is intended to measure. In operational terms, this means that on a test-retest situation with the same instrument, a testee should rank in the same position relative to his group. Relative reliability is found by correlating the sets of scores obtained on two successive administrations of a test. Besides relative reliability, there is absolute reliability. The latter is concerned with the errors inherent in the instrument itself and the statistic used in this connection is the standard error of measurement. This statistic yields an estimate of the deviation of a set of obtained scores from their true values that will be found when the instrument is free from chance errors and errors of measurement. The two reliabilities will now be considered separately.

Kuder-Richardson formula No. 20 was employed in the calculation of relative reliability. The application of this method was preferred in view of the homogeneity of the test items. The same factor was being tapped by all items in the test. Reliability coefficients were computed for the five age levels and the values were .54, .63, .82, .67 and .82 for ages 13, 14, 15, 16 and 17 respectively. The reliability coefficient for all ages taken together was found to be .75. It will be noticed that, in general, as age increases, reliability coefficients increase. The coefficients are low at some ages showing that the range of variation in the trait was small.

The standard error of measurement was calculated at each of the five age levels. The values obtained using raw scores were ± 1.44 , ± 1.43 , ± 1.30 , ± 1.34 and ± 1.17 for ages 13, 14, 15, 16 and 17 respectively. It

may be noticed that as reliability increases, the standard error of measurement is reduced. Therefore, it follows that the predictive value of the instrument is enhanced at that particular age level. The above results can be interpreted along the following lines, assuming that the distributions are normally distributed. Taking age 13, it can be said that the odds are 68 in 100 that an obtained score is in error by 1.44 points or less; 95 in a 100 that it is in error by 2.88 or less; and 99 in 100 that it is in error by 3.6 or less. The other values can be interpreted in a similar fashion. It is clear that the standard errors of measurement are high. This is to be expected due to the low reliability coefficients. Comparatively, the instrument is more dependable at ages 15 and 17, but its predictive value is generally low.

Validity of the Instrument:

The validity of a test is defined as the accuracy with which it measures what it purports to measure. Two aspects of validity are considered here namely rational and statistical validities. These will now be discussed.

Rational validity is concerned with the content of the test from the standpoint of its relevance to the purpose of the test and to the psychological function under investigation. Words that demanded a good level of conceptualization were given preference. Moreover, they were actually taken out of textbooks used in the schools this year. Also, the internal validity of each item in the test was ascertained by correlating it with the total test scores. This part was already dealt with in detail in previous chapters.

Statistical validity aims at validating the test against an external criterion that is believed to be indicative of success in the trait examined. It was pointed out earlier that courses loaded with the verbal element provided the best predictors of success on the job. The scores on each item were already correlated with school grades in Arabic for the different age levels to establish the validity of each item. Items that showed no correlations or that correlated negatively were discarded.

A Pearson r correlation was then run between scores on the final form of the test and grades in Arabic and English. Validity coefficients were found for the test at the five age levels. The choice of English and Arabic grades as the validating criteria was made on a purely rational judgement. The values of the coefficients appear on table (12).

TABLE 11

Validity Coefficients at the Five Age Levels

Age	Validity Coefficient	
	Arabic	English
13	.08	.09
14	.73	.36
15	.03	.09
16	.37,	.31
17	.64	.38

These results were obtained for a total of 84 subjects taken from the international College. Ages 13 and 15 were not well represented in the sample drawn from this school. Therefore, the coefficients should not be generalized to the whole population particularly age levels 13 and 15. The coefficients at the remaining ages are significant showing that the relationship was not due to chance, but a systematic factor was behind it.

On the whole, there is some evidence that success on the vocabulary test is conducive to success on subjects calling for a good command of verbal ability. This conclusion is made on account of the positive correlations obtained. However, in view of the small group of subjects used and the inaccuracy in grading predominant in schools, these coefficients should not be regarded as reliable estimates of the relationship magnitudes.

Vocabulary and Achievement:

It was already pointed out that vocabulary ability could yield by itself an estimate of general intelligence. This statement was made in the light of research evidence already cited. The data pointed to the fact that a positive relationship exists between verbal ability as measured by vocabulary tests and general intelligence. However, the role of vocabulary ability in academic achievement has not been given due attention and concern. It must not be forgotten that verbal symbols are the cornerstone of any educational endeavor since they are the media of transmission and communication. Therefore, it is logical to conclude that a student with a good command of verbal symbols is apt to show a higher level of achievement than one with a low verbal ability. In other words, the high-scoring students on the vocabulary test are also the high achievers at

school. This hypothesis was then put to the practical test and the procedure followed in its empirical validation is explained below.

The scores of the test in its final form were correlated with the year's averages of 84 students available at the International College. A Pearsonian correlation coefficient of .24 was found. This value is significant at the five percent level. It would seem then that within experimental errors, vocabulary ability is a correlate of achievement. The vocabulary test now adds the new significance of being a predictor of achievement as well as of intelligence.

Normalizing the Distributions:

A trait like vocabulary ability should conform to a normal distribution at the various age levels. However, this was not possible to achieve, as is seen in figures (2) through (6), due to sampling errors. The fact that a large number of the subjects was drawn from summer schools appears to have influenced the shape of the distributions appreciably. This is because summer school students comprise a population of repeaters who seem to lack the ability as well as the motivation. Therefore, the best-fitting curve for each distribution has been constructed in order to render it normal. This may be described as the curve that would have been obtained had adequate sampling techniques been employed.

The steps involved in the normalization process are as follows:

- 1) The scores are entered in the first column of table (13) through (17). They range from 0 to 10.
- 2) The corresponding observed frequencies are placed in the second column.
- 3) The Z-score is then calculated for each raw score by taking

its upper limit. The Z-scores appear in the third column.

4) The area under each Z-score is read from tables and the values entered in the fourth column.

5) The area within two successive Z-scores is obtained by subtracting the area below the lower Z-score from that below the upper one. Those areas are shown in the fifth column.

6) The number of cases at each included area is determined by multiplying that area by the number of subjects at the age level concerned. Those expected frequencies appear in the sixth column.

7) The scores and the expected frequencies are then plotted with the former as the abscissa and the latter as the ordinate. The normalized curves are shown by dotted lines in figures (2) through (6).

8) A seventh column has been added to the tables with the purpose of converting raw scores into T-scores with a mean of 50 and a standard deviation of 10.

The normalizing of curves was carried out with the intention of showing the discrepancy, which may be attributed to errors in sampling, between the observed and expected frequencies of scores. The basic assumption was that vocabulary ability was normally distributed among subjects at each age level. The curves could be used by future research workers in order to avoid the sampling pitfalls encountered. The raw scores were converted into T-scores so as to have a common point of reference. Inter-age comparisons could then be made among the different age levels.

TABLE 12

Normalized Distribution for Age 13

Score	Observed Frequency	Z-Score	Proportion Below	Proportion Within	Expected Frequency	T-Score
10	0	3.01	.9987	.0042	0	80.1
9	1	2.54	.9945	.0137	1	75.4
8	3	2.07	.9808	.0356	2	70.7
7	2	1.60	.9452	.0744	4	66
6	8	1.13	.8708	.1254	7	61.3
5	9	0.66	.7454	.1740	9	56.6
4	11	0.18	.5714	.1855	10	51.8
3	4	-0.29	.3859	.1623	9	47.1
2	9	-0.76	.2236	.1143	6	42.4
1	4	-1.23	.1093	.0647	3	37.7
0	2	-1.70	.0446	.0446	2	33

NORMALIZED DISTRIBUTION FOR AGE 13

— OBSERVED DISTRIBUTION
--- EXPECTED DISTRIBUTION

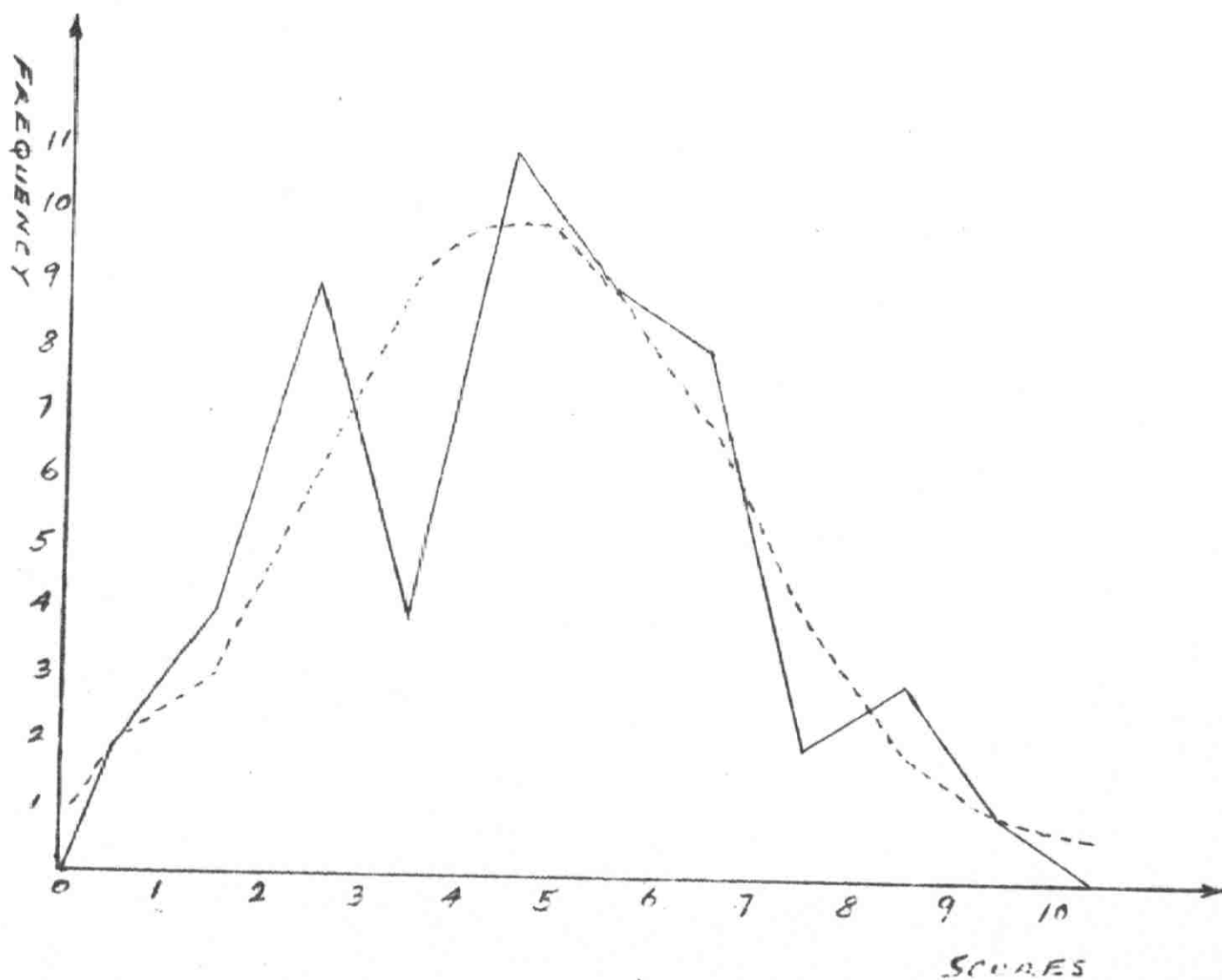


FIG. 2

TABLE 13

Normalized Distribution for Age 14

Score	Observed Frequency	Z-Score	Proportion Below	Proportion Within	Expected Frequency	T-Score
10	0	2.29	.9890	.0204	1	72.9
9	4	1.86	.9686	.0435	2	68.6
8	4	1.44	.9251	.0790	3	64.4
7	3	1.02	.8461	.1170	5	60.2
6	7	0.61	.7291	.1538	6	56.1
5	7	0.19	.5753	.1663	7	51.9
4	6	-0.23	.4090	.1512	6	47.7
3	3	-0.65	.2578	.1155	5	43.5
2	4	-1.07	.1423	.0742	3	39.3
1	2	-1.49	.0681	.0400	2	35.1
0	1	-1.91	.0281	.0281	1	30.9

NORMALIZED DISTRIBUTION FOR AGE 14

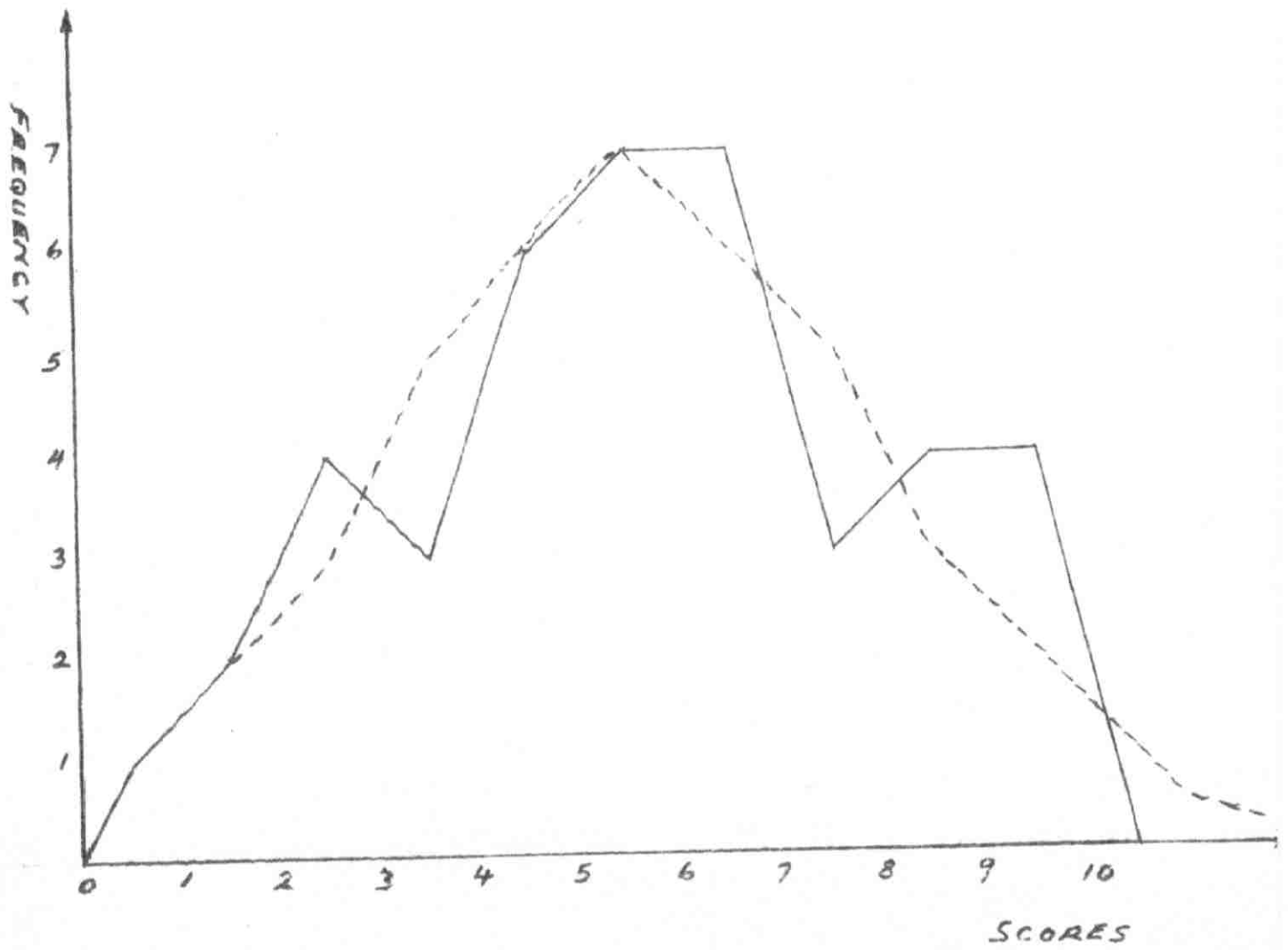


FIG. 3

TABLE 14

Normalized Distribution for Age 15

Score	Observed Frequency	Z-Score	Proportion Below	Proportion Within	Expected Frequency	T-Score
10	3	1.65	.9505	.0439	1	66.5
9	5	1.32	.9066	.0677	2	63.2
8	2	0.99	.8389	.0935	3	59.9
7	2	0.66	.7454	.1161	4	56.6
6	4	0.33	.6293	.1253	4.1	53.3
5	5	0.01	.5040	.1295	4.14	50.1
4	2	-0.32	.3745	.1167	4	46.8
3	1	-0.65	.2578	.0918	3	43.5
2	3	-0.97	.1660	.0709	2	40.3
1	4	-1.31	.0951	.0446	1.4	36.9
0	1	-1.64	.0505	.0505	1.6	33.6

NORMALIZED DISTRIBUTION FOR AGE 15

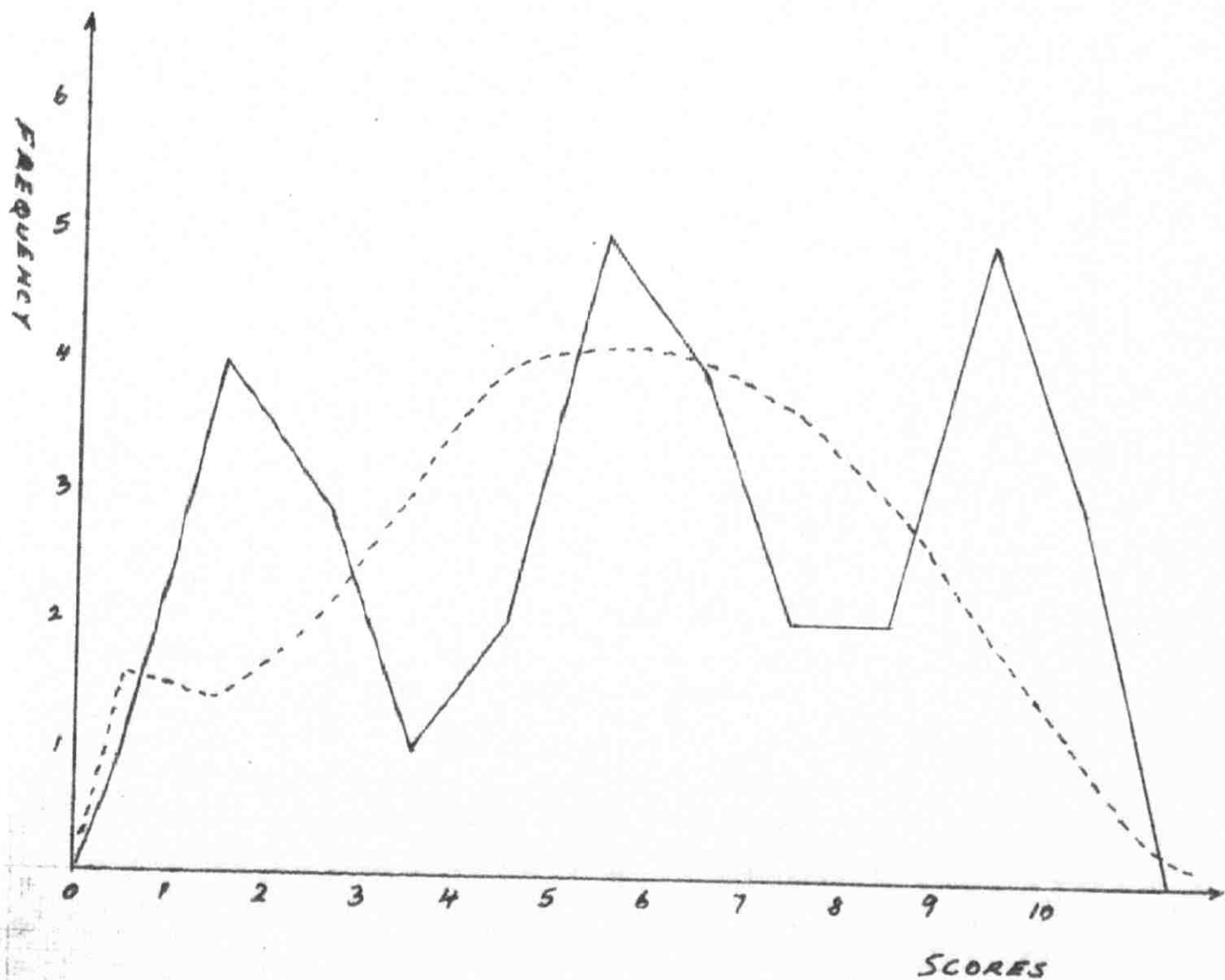


FIG. 4

TABLE 15

Normalized Distribution for Age 16

Score	Observed Frequency	Z-Score	Proportion Below	Proportion Within	Expected frequency	T-Score
10	3	1.83	.9664	.0457	2	68.3
9	3	1.41	.9207	.0842	3	64.1
8	7	0.98	.8365	.1242	4	59.8
7	4	0.56	.7123	.1566	5.6	55.6
6	6	0.14	.5557	.1660	6	51.4
5	3	-0.28	.3897	.1477	5	47.2
4	5	-0.70	.2420	.1106	4	43
3	3	-1.12	.1314	.0696	3	38.8
2	1	-1.54	.0618	.0368	1	34.6
1	1	-1.96	.0250	.0163	.64	30.4
0	0	-2.38	.0087	.0087	0	26.2

NORMALIZED DISTRIBUTION FOR AGE 16

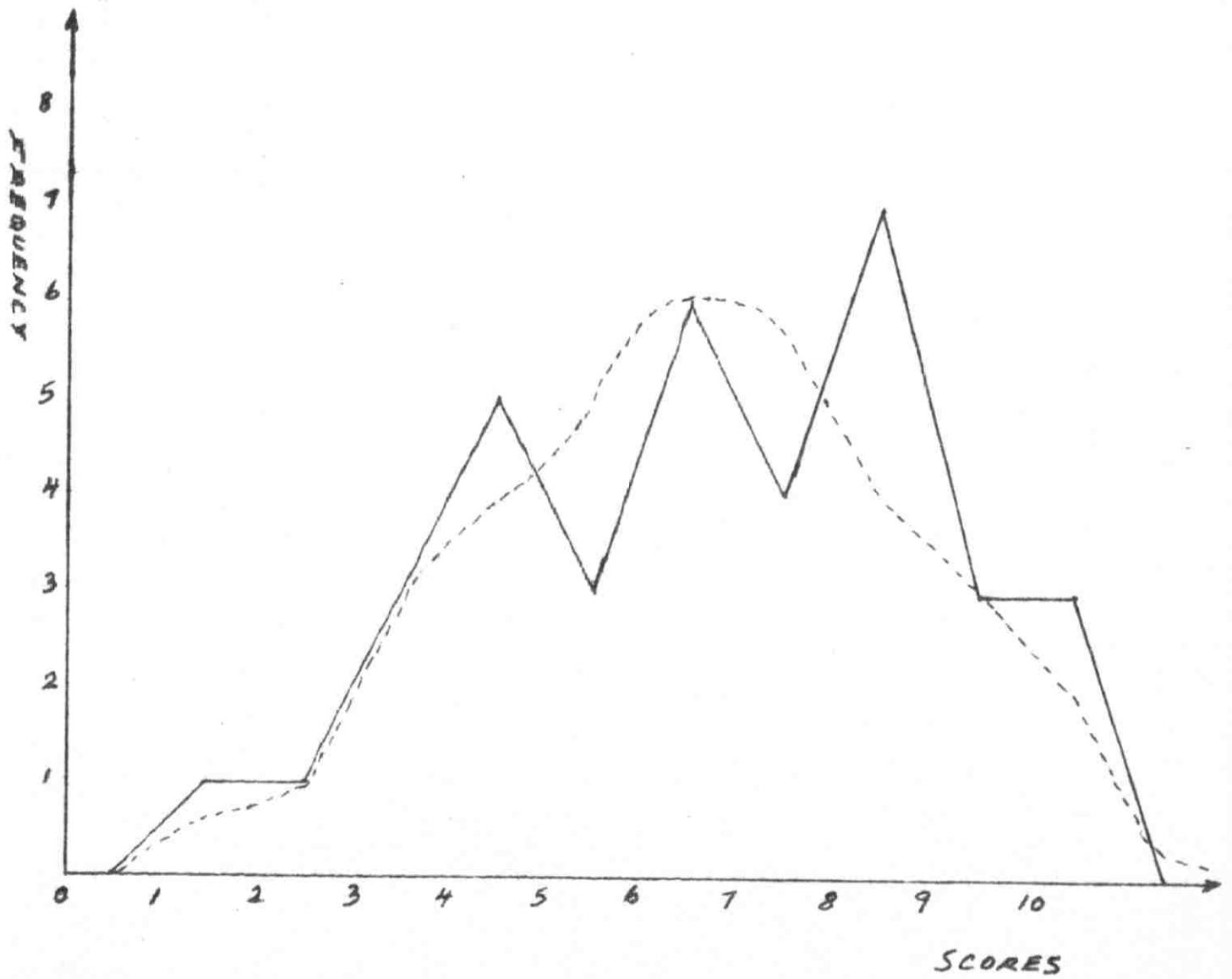


FIG. 5

TABLE 16

Normalized Distribution For Age 17

Score	Observed Frequency	Z-Score	Proportion Below	Proportion Within	Expected Frequency	T-Score
10	9	1.20	.8849	.0854	4	62
9	9	0.84	.7995	.1151	5	58.4
8	8	0.48	.6844	.1366	5.7	54.8
7	4	0.12	.5478	.1426	6	51.2
6	1	-0.24	.4052	.1309	5	47.6
5	2	-0.60	.2743	.1058	4	44
4	1	-0.96	.1685	.0751	3	40.4
3	5	-1.32	.0934	.0469	2	36.8
2	1	-1.68	.0465	.0258	1	33.2
1	2	-2.04	.0207	.0125	0.5	29.6
0	0	-2.40	.0082	.0082	0	26

NORMALIZED DISTRIBUTION FOR AGE 17

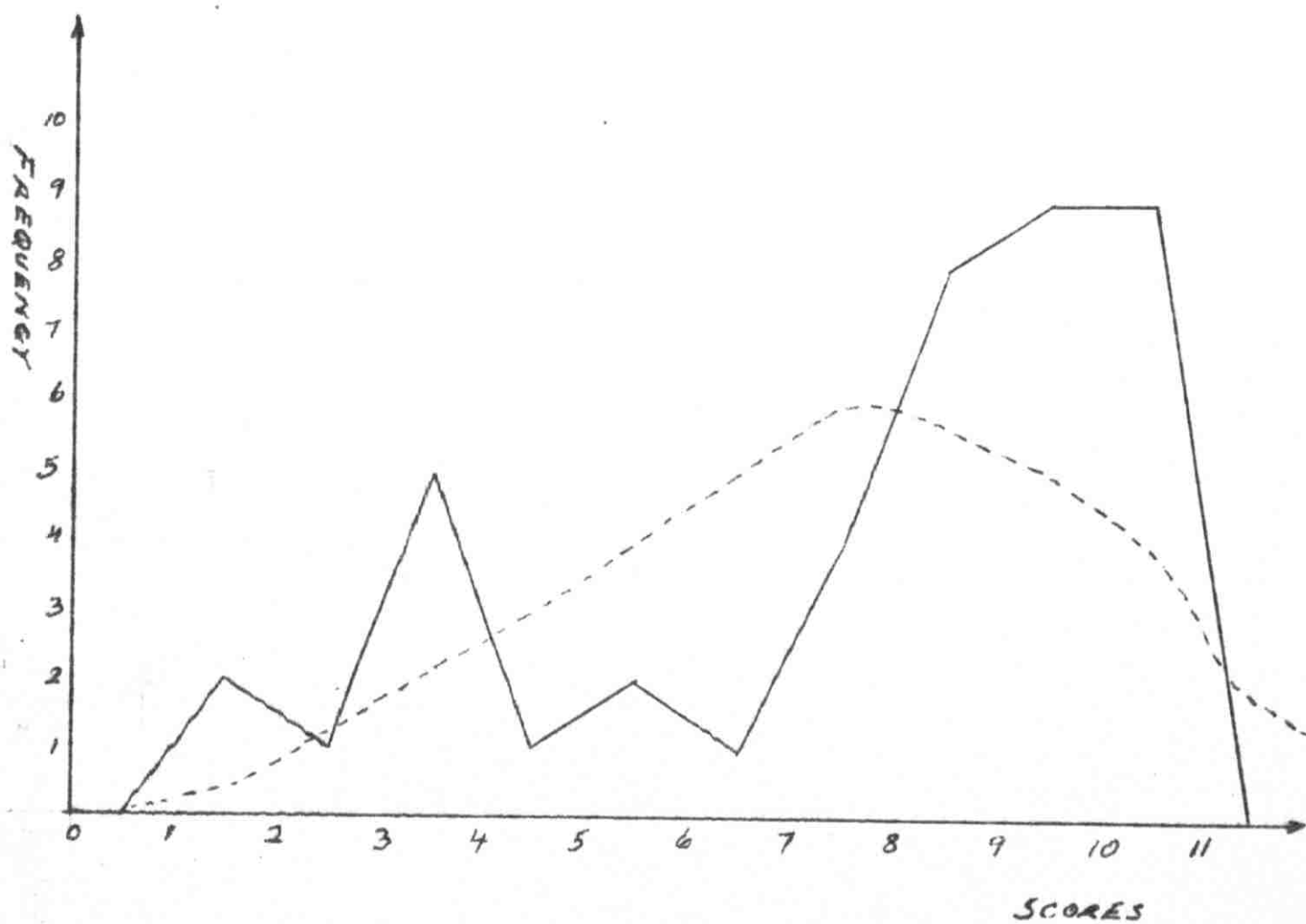


FIG. 6

CHAPTER V

CONCLUSION

Despite its limitations, this study has succeeded in certain areas. Its major contribution lies in two distinct fields:

1) First, through the process of item analysis, suitable items were found which were then allocated to the various age levels. Those items form the nucleus around which a more elaborate scale of vocabulary ability may be designed. The selected expressions also set the pattern for future research in the sense that they indicate the type of words suitable to the abilities of the different age levels. However, the validity of the results obtained should be tested by using a larger and a more representative sample.

2) Secondly, this research has succeeded in finding a positive significant relationship between vocabulary ability and academic achievement. The implications of this finding for educational practice will be discussed later in this chapter.

Among the results of this study, two findings merit special consideration. First, it has been observed that composite expressions rather than single words possess the characteristic of differentiating among the various age groups. Eight out of the ten selected items belong to this category. It would seem then that composite expressions in the Arabic language require a higher level of abstraction than single words. In the latter, the difficulty appears to reside more on the words as such than on the concepts involved. Therefore, through sheer rote memorization or

fortunate guessing, a subject may choose the correct meaning of a word. This accounts partly for the lack of discrimination manifest in the responses to single expressions. On the other hand, composite expressions seem to tap deeper mental processes. Taken separately, the verbal symbols forming such an expression may be well-known to the subject, but it is the totality of the expression that really matters. A good deal of insight and conceptualization is called for if he is to build a meaningful whole of the constituent parts.

The other finding is that the reliability of the ten selected items increases with age in the range 13-17 years. This has already been shown by the reliability coefficients obtained at the different age levels. A possible explanation for this result is as follows: Since vocabulary ability increases with age, it may be assumed that the number of items attempted by a testee is a function of his age level. Therefore, more reliable results are expected at higher age levels due to the fact that reliability increases with increasing test length. However, further research is needed to verify this conclusion.

Educational Implications:

Since vocabulary ability appears to be a correlate of achievement, it becomes the task of the school to foster the environment that will nourish this ability, if the student is to achieve up to his potentialities. The school should place a premium on courses that are loaded with the verbal element. This should be commenced at an early stage during childhood so that optimum development of this ability is reached in adulthood.

Underachievement can no longer be attributed to lack of motivation, talent or interest only, because a verbal handicap contributes equal-

ly as much to the problem. Hence, the need arises for tools that can locate any signs of verbal difficulty in students. The vocabulary test is of use here in the sense that it is a reliable predictor of success in schools. Students with verbal trouble may then be singled out and assisted before any repercussions occur.

The reasons for the relationship between vocabulary ability and general intelligence are not far to seek. It was already pointed out that almost all intelligence scales are verbal in nature. Therefore, their correlation with vocabulary tests is expected in view of this common attribute. More important still, vocabulary tests seem to possess something of their own that enables them to stand on the same footing with other full scales in mental measurement. Hence, the vocabulary test is now regarded as a reliable substitute for an intelligence test in obtaining a quick estimate of mental ability. In the Arab world, where mental measurement is still a virgin field, vocabulary tests can serve as valuable predictors of mental ability until full scales are standardized.

Suggestions and Recommendations for Future Research:

The results of this study should not be regarded as conclusive and final because of the many weaknesses inherent in it. First, the sample used was very inadequate in number and a larger sample should have been used. The age span, 13-17 years, was small. If more ages had been included, some of the items deleted from the test might have proved to be suitable for placement at these age levels.

As far as ability was concerned, the subjects appeared to belong to different populations as seen in figures 2 through 6 in chapter 4. In addition to their insufficient number, subjects drawn from summer schools

were either repeaters or bright students who intended to accelerate their programs of study. They also came from different schools and had different educational backgrounds. The International College group formed a third category. This may account for the three modes in the figures. Therefore, any future study should attempt to improve on the sampling techniques used in this project for more accurate results.

Further research is also needed on the role of education, sex and social class in vocabulary ability. This is because vocabulary ability appears to be the product of these factors and intelligence. Future research should investigate the extent to which each of these factors influences vocabulary ability. Unless this is done, those variables are bound to be reflected in any measurement of vocabulary.

BIBLIOGRAPHY

Books

- Cronbach, Lee J. Essentials of Psychological Testing.
New York: Harper & Bros., 1960.
- Dearborn, Walter. Intelligence Tests.
New York: Houghton Mifflin Co., 1928.
- Downie, N.M., and Heath, R.W. Basic Statistical Methods.
New York and London: Harper & Row, 1959.
- Ebel, Robert L. Measuring Educational Achievement.
Englewood Cliffs, New Jersey: Prentice-Hall, 1965.
- Eells, Kenneth, et al. Intelligence and Cultural Differences.
Chicago: The University of Chicago Press, 1951.
- Flavel, John H. The Developmental Psychology of Jean Piaget.
Toronto, New York and London: D. Van Nostrand Co., 1963.
- Freeman, Frank N. Mental Tests.
Boston: Houghton Mifflin Co., 1939.
- Freeman, F.S. Theory and Practice of Psychological Testing.
New York: Holt, Rinehart and Winston Co., 1965.
- Guilford, J.P. Fundamental Statistics in Psychology and Education.
New York and Tokyo: McGraw-Hill Book Co., 1956.
- Hebb, D.O. The Organization of Behavior.
New York: Wiley & Sons, 1949.
- Heim, A.W. The Appraisal of Intelligence.
London: Methuen & Co., 1954.
- Jenkins, J.J., and Patterson, D.G. (eds.) Studies in Individual Differences.
New York: Appleton-Century Crofts, 1961.

- Lindquist, E.F. (ed.) Educational Measurement.
Washington: American Council on Education, 1951.
- Miner, John B. Intelligence in the United States:
New York: Springer Publishing Co., 1956.
- Monroe, W.S., et al. Educational Tests and Measurements.
Boston: Houghton Mifflin Co., 1924.
- Piaget, Jean The Psychology of Intelligence.
New York: Harcourt, Brace & Co., 1950.
- Pintner, Rudolf. Intelligence Testing.
New York: Henry Holt & Co., 1923.
- Spearman, C. The Nature of Intelligence and the Principles of Cognition.
London: Macmillan & Co., 1927.
- Stoddard, George. The Meaning of Intelligence.
New York: The Macmillan Co., 1947.
- Terman, L.M., and Merrill, M.A. Measuring Intelligence.
Boston: Houghton Mifflin Co., 1937.
- Thorndike, E.L., Bregman, E.O., and Cobb, M.V. The Measurement of Intelligence.
New York: Teachers College, Columbia University, 1925.
- Thorndike, Robert L. Personnel Selection.
New York: John Wiley & Sons, 1959.
- Thurstone, L.L. Primary Mental Abilities.
Chicago: University of Chicago Press, 1938.
- Wechsler, David. The Measurement of Adult Intelligence.
Baltimore: Williams and Wilkins, 1944.

Periodicals

- Elwood, Mary I. "A Preliminary Note on the Vocabulary Test in the Revised Stanford-Binet Scale," Journal of Educational Psychology, XXX, No. 8 (November, 1939), 632-634.
- Fox, Charlotte. "Vocabulary Ability in Later Maturity," Journal of Educational Psychology, XXXVIII, No. 8 (December, 1947), 482-492.
- Garfield, S.L., and Blek, Libby. "Age, Vocabulary Level and Mental Impairment," Journal of Consulting Psychology, XVI, No. 5 (October, 1952), 395-398.
- Kelley, V.H. "An Experimental Study of Certain Techniques for Testing Word Meanings," Journal of Educational Research, XXVII, No. 4 (December, 1933), 277-282.
- Lewinski, R.J. "Vocabulary and Mental Measurement: a Quantitative Investigation and Review of Research," Journal of Genetic Psychology, LXXII (June, 1948), 247-281.
- Miner, John B. "On the Use of a Short Vocabulary Test to Measure General Intelligence," Journal of Educational Psychology, LII, No. 3 (June, 1961), 157-160.
- Terman, L.M. "The Vocabulary Test as a Measure of Intelligence," Journal of Educational Psychology, IX (October, 1918), 452-466.

APPENDIX

FORMULAS

1. Phi Coefficient (Guilford's formula):

$$\phi = \frac{P_u - P_l}{2\sqrt{Pq}}$$

where

P_u = Proportion in the upper group answering the item correctly

P_l = Proportion in the lower group answering the item correctly

$q = 1 - p$

2. Point - biserial correlation coefficient:

$$r_{pb} = \frac{\bar{X}_p - \bar{X}_t}{s_t} \sqrt{\frac{p}{q}}$$

where

\bar{X}_p = the mean score of those answering the item correctly

\bar{X}_t = the mean of the total test scores

s_t = the standard deviation of the test

p = the proportion of the total group answering the item correctly

$q = 1 - p$

3. Kuder-Richardson reliability formula No. 20:

$$r_{tt} = \frac{n}{n-1} \left(1 - \frac{pq}{s^2} \right)$$

where

n = number of items on the test

s = standard deviation of the test

p = the proportion of those responding correctly to an item

$q = 1 - p$

4. Standard error of measurement:

$$S_e = S \sqrt{1 - r_{tt}}$$

where

S = standard deviation of the test

r_{tt} = the reliability coefficient of the test

5. Pearson r computed from raw scores:

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2] [N \sum Y^2 - (\sum Y)^2]}}$$

where

N = number of subjects

X = a raw score on one variable

Y = a raw score on the other variable

هذا الامتحان جزء من بحث يقوم به واحد من خريجي الجامعة الامريكية وليس الغرض منه اختبار معلوماتك او تقييمها فلا داعي لالارتباك والخوف كما ان النتائج ستعاط بسرية تامة ولكن باستجابتك الكاملة وببذل اقصى جهدك تكون قد ساهمت في بحث يجعل من لغتك العربية موضوعه ويهدف الى دراستها من زاوية جديدة تعود عليها بالفائدة.

فيما يلي مجموعة من المفردات والتماير المرقمة يلي كل واحد منها خمس اجوبة مسبقة بحروف - عليك ان تختار الجواب الصحيح لكل واحد منها من بين الاجوبة المعطاة وان تقوم بعد ذلك بوضع علامة x في المربع الممين عليه الحرف الذي ترى انه يسبق الجواب الصحيح وذلك تحت الرقم المناسب للسؤال في ورقة الاجوبة المرفقة - حاول ان تجيب على كل الاسئلة .

(١) جزيرة :

- أ - اعشاب فوق البحر ب - صخرة عالية تطل على البحر ج - قطعة ارض حولها ماء
د - حديقة بها نافورة ه - اشجار تحيط ببحيرة ماء .

(٢) يركض :

- أ - يتنفس بصعوبة ب - يجري بسرعة ج - يسابق الزمن
د - يتمجل الامور ه - يهرب من المصائب .

(٣) يهمس :

- أ - يخاطب نفسه ب - يتكلم بصوت خافت ج - يهذى من الحمى
د - يتكلم وهو نائم ه - يفنئ لنفسه .

(٤) الوسادة :

- أ - القطعة الممزقة من الثياب ب - المقعد الذي نجلس عليه ج - المعطف الذي يقينا
البرد د - السرير الذي ننام عليه ليلا ه - ما يوضع تحت الرأس عند الرقاد .

(٥) الغابة :

- أ - المكان الذي تكثر فيه الاشجار ب - المدينة القديمة المهجورة ج - الحديقة
اليانعة الازهار (د) المكان الذي تسكنه الوحوش ه - الارض الجبلية الجرداء .

(٦) يدنو :

- أ - يطلب المغفرة ب - يقترب من الشيء ج - تسوء اخلاقه د - يفشل في مطلبه
ه - يبحث عن شيء مفقود .

(٧) الصلاح :

- أ - الرجل الذي يصيد الاسماك ب - سائق سيارة الاجرة ج - الدليل الذي يقود القوافل
د - القائد الاعلى للجيش ه - الشخص الذي يعمل في سفينة .

(٨) يستفيث :

- أ - يطلب ماء ب - يريد عونا ج - يصن غاضبا د - يبكي حزنا
ه - يهرب خوفا .

(٩) الاتراب :

- أ - الاصدقاء الاوفياء ب - ابناء البلد الواحد ج - زملاء الكفاح
د - اصحاب الاعمار المتقاربة ه - طلاب المدرسة الواحدة .

(١٠) شفاف :

- أ - ضعيف البناء ب - قليل الفائدة ج - واسع الارجا د - كاشف لما تحته
ه - قابل للكسر السريع .

(١١) الظمأ :

- أ - الحزن الشديد ب - الشوق للماء ج - الرحيل المفاجيء د - السير
دون هدف ه - المرض الخبيث .

(١٢) الابل كم :

أ - الطفل الذي فقد امه ب - الشخص العاجز عن الكلام ج - الرجل البطي الفهم
د - الشخص الذي لا يسمع ه - الشخص المتهور في اعماله .

(١٣) يتغاضل :

أ - تنهار قواه ب - يصاب بمرض مفاجي ج - يعتمد عن المصائب
د - يتوانى عن المشاركة ه - يترث ويتنظر .

(١٤) الحافة :

أ - مكان تباع فيه لعب الاطفال ب - محل لبيع المرطبات ج - المكتبة الكبيرة
د - حديقة غزيرة الاشجار ه - مكان تباع فيه الخمر .

(١٥) الماصفة :

أ - الموج العالي ب - المطر المتواصل ج - الريح العنيفة د - السفينة المحطمة
ه - البرق اللامع .

(١٦) الغمبول :

أ - الاكثار من النوم ب - عدم القيام بالواجب ج - شروذ الذهن
د - انعدام النشاط ه - المبالغة في الاكل .

(١٧) الشراب :

أ - الظل الوارف ب - الشي الخيالي ج - الماء العذب د - المطر الرزان
ه - الطائر المفترس .

(١٨) الكهف :

أ - البيت المحفور في الجبل ب - المكان المظلم ج - الطريق المهجور
د - البنيان المعطم ه - المقبرة الكبيرة .

(١٩) الصدى :

أ - الصرخة العالية ب - انعكاس الصوت ج - صوت الماء المنحدر د - الفناء القديم
ه - تفريد البلبل .

(٢٠) طارى :

أ - نبأ سار ب - موعد هام ج - مصاب اليم د - حدث مفاجي ه - اشتباك مسلح

(٢١) الرابية :

أ - الريح العاتية ب - الصخرة الشامخة ج - الارض العالية د - الموجة القوية
ه - القلعة الحصينة .

(٢٢) الشذا :

أ - الحب الزائد ب - اسم لفاكهة ج - النسيم العليل د - الصوت العذب
ه - الرائحة الذكية .

(٢٣) الهوة :

أ - الطريق الوعر ب - الحفرة العميقة ج - العذب الذي لا يفتقر د - زلة اللسان
ه - الارض المنبسطة .

(٢٤) الكبرياء :

أ - رفعة النفس ب - كبر السن ج - علو الشأن د - احتقار الآخرين
ه - مركب النقص .

(٢٥) المثيرة :

أ - السقوط على الارض ب - البئر المهجورة ج - الوقوع في الخطأ
د - الحصول على الشي ه - القول الذميم .

(٢٦) السقام :

أ - التعب الشديد ب - الشوق للوطن ج - البكاء والمويل د - الصدود والهجران
ه - المرض الشديد .

(٢٧) الاسـمـال :

أ - الاعمال النبيلة ب - الملابس البالية ج - الكتب القديمة د - الفنائـم السلوية هـ - الاخلاق السيئة .

(٢٨) يطـاـطـل :

أ - يبـالـغ في المـجـامـلة ب - يـكـثـر من الشـكـوى ج - يـفـش في الامـتـحـان د - يـكـثـر ارجاء الوعود هـ - يـثـير المشـاـكل .

(٢٩) الزمـهـريـر :

أ - البـرد الشـديـد ب - المـاصـفة الهـوجـاء ج - الطـر المـنـهـمـر د - السـيـل الجـارـف هـ - الحـر الـلاـفـح .

(٣٠) الـارـوصـة :

أ - البـلد الـذي تـعـيـش فـيـه ب - الشـيـء الـذي نـسـمـى اليـه ج - الـاصـل الـذي نـتـمـي اليـه د - الشـعـور القـوى بالـوطـنـية هـ - الثـقـة الشـديـدة بالنـفـس .

(٣١) يـمـشـي الـهـوينـا :

أ - يـغـتـال تـيـها ب - يـسـرـع في المـشـي ج - يـقـفـز من الفـرح د - يـتـهـاـون في المـشـي هـ - يـتـهـادى في المـشـي .

(٣٢) وـاسـع الـصـدر :

أ - كـبـير الجـسـم ب - قـلـيل الـادـراك ج - قـوى البـنـية د - عـدـيم الصـبر هـ - وافر الحـم .

(٣٣) يـيـسـط يـدـه :

أ - يـثـرى بـعد فـقر بـد يـجـزـل المـطـاء ج - يـطـلب المـسـطـدة د - يـحـسن المـعـامـلة هـ - يـكـثـر الدـعـاء .

(٣٤) حـالـك الـسـواـد :

أ - عـظـيم الشـأن ب - سـرـيـع الغـضـب ج - خـفـيف الـسـواـف د - ^دعـراـكـن الـلـون هـ - سـيء الحـظ .

(٣٥) يـضـيق زـرعـا :

أ - يـفـشل في مـطـلبـه ب - يـخـرج عن طـاقـته ج - يـشـكو من الفـقر د - يـيـأس من العـلاج هـ - ~~لا يـسـتـطـيع~~ من يـسـاعـده .

(٣٦) رابـط الجـأش :

أ - قـلـيل الحـياء ب - دـائـم المـبـوس ج - هـادئ النفس د - قـوى البـدن هـ - حـسن السلـوك .

(٣٧) يـهـيـم عـلى وـجـهـه :

أ - يـسـير بـدون هـدـف ب - يـسـقـط عـلى وـجـهـه ج - يـبـكي و يـلـطـم وـجـهـه د - يـشـكو من الحـب والـهـيـام هـ - يـندـم عـلى ما فـعـله .

(٣٨) صـمـب المـراس :

أ - غـرـيب الطـبـاع ب - حـاد المـزاج ج - قـلـيل الكـرم د - عـسـير القـيـاد هـ - كـثـير التـجـارـب .

(٣٩) عـنـفـوان الشـباب :

أ - عـهـد الطـفـولة ب - بـوادـر الشـيـخـوخة ج - طـيـش الشـباب د - فـتـرة المـراهـقة هـ - قـوة الشـباب .

(٤٠) كـثـير الرـمـاد :

أ - قـلـيل الفـائـدة ب - جـزـيل المـطـاء ج - مـتـسـخ الشـباب د - عـدـيم الـاهـتـمـام هـ - شـديـد البـخل .

(٤١) يشفي غليله :

أ- يندم ويمتذر ب- يتماثل للشفاء ج- يتعاطى الخمر د- يؤدي الواجب ه- يشبع رغبته .

(٤٢) ليل بطيء الكواكب :

أ- ليل طويل ب- ليل شديد الظلمة ج- ليل هادئة د- ليل باهت النجوم ه- ليل حزين .

(٤٣) يجهز عليه :

أ- يقوم باعداده لامر ما ب- يعتمد عليه اعتمادا كاملا ج- يلومه على فعلته د- يقضي على حياته ه- يستولي على ممتلكاته .

(٤٤) زوال الرماد في الصيون :

أ- الحاق الأذى بالناس ب- ضح الناس عن المشاهدة ج- تضليل الناس عن الحق د- الصاق الذنب بآخر ه- الهرب من الناس .

