AGE STUDY OF TIMED TESTS FOR CREDENTIAL CANDIDATES
AT SACRAMENTO STATE COLLEGE

by

Lucille Blackwell Colby
A. B. (University of California at Berkeley) 1934

THESIS

Submitted in partial satisfaction of
the requirements for the degree of

MASTER OF ARTS
AT THE
SACRAMENTO STATE COLLEGE

Approved:

Ed L. Klingelholfer, Chair
James M. Bradfield
Emmett C. Thompson

Advisory Committee
Date July 24, 1950
### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION TO THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>The Problem</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>Origin of the Problem</td>
<td>2</td>
</tr>
<tr>
<td>Delimitations</td>
<td>4</td>
</tr>
<tr>
<td>Importance of the Study</td>
<td>8</td>
</tr>
<tr>
<td>The Local Contribution</td>
<td>8</td>
</tr>
<tr>
<td>The General Contribution</td>
<td>9</td>
</tr>
<tr>
<td>II. REVIEW OF THE LITERATURE</td>
<td>11</td>
</tr>
<tr>
<td>Age and Decreasing Intelligence</td>
<td>15</td>
</tr>
<tr>
<td>Age and Increasing Intelligence</td>
<td>21</td>
</tr>
<tr>
<td>Age and Constancy of Vocabulary</td>
<td>25</td>
</tr>
<tr>
<td>Age and Increasing Vocabulary</td>
<td>34</td>
</tr>
<tr>
<td>Age and Arithmetic</td>
<td>38</td>
</tr>
<tr>
<td>Age and Speed</td>
<td>44</td>
</tr>
<tr>
<td>Summary and Discussion</td>
<td>50</td>
</tr>
<tr>
<td>III. THE PRELIMINARY STUDY</td>
<td>58</td>
</tr>
<tr>
<td>Description of Tests</td>
<td>58</td>
</tr>
<tr>
<td>Rationale for Test Selection</td>
<td>59</td>
</tr>
<tr>
<td>The Population</td>
<td>61</td>
</tr>
<tr>
<td>The Method of Procedure</td>
<td>62</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Results of the Preliminary Study</td>
<td>63</td>
</tr>
<tr>
<td>Vocabulary and Age</td>
<td>63</td>
</tr>
<tr>
<td>Linguistic Test and Age</td>
<td>68</td>
</tr>
<tr>
<td>Quantitative Test and Age</td>
<td>72</td>
</tr>
<tr>
<td>Summary of Preliminary Study</td>
<td>76</td>
</tr>
<tr>
<td>Conclusions from the Preliminary Study</td>
<td>78</td>
</tr>
<tr>
<td>Discussion</td>
<td>79</td>
</tr>
<tr>
<td>IV. THE EXPERIMENT AND RESULTS</td>
<td>81</td>
</tr>
<tr>
<td>Design of the Experiment</td>
<td>81</td>
</tr>
<tr>
<td>The Test Battery</td>
<td>81</td>
</tr>
<tr>
<td>CEZ Vocabulary, Untimed</td>
<td>82</td>
</tr>
<tr>
<td>ACE Arithmetic, Untimed</td>
<td>83</td>
</tr>
<tr>
<td>The Marking Test</td>
<td>84</td>
</tr>
<tr>
<td>Description of the Population</td>
<td>87</td>
</tr>
<tr>
<td>Test Procedure</td>
<td>87</td>
</tr>
<tr>
<td>Results of the Experiment</td>
<td>89</td>
</tr>
<tr>
<td>Marking Test Comparisons</td>
<td>90</td>
</tr>
<tr>
<td>Marking Test and Age</td>
<td>91</td>
</tr>
<tr>
<td>Marking Test and Timed Vocabulary</td>
<td>95</td>
</tr>
<tr>
<td>Marking Test and Untimed Vocabulary</td>
<td>98</td>
</tr>
<tr>
<td>Marking Test and Linguistic Scale</td>
<td>101</td>
</tr>
<tr>
<td>Marking Test and Timed Arithmetic</td>
<td>104</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Marking Test and Untimed Arithmetic</td>
<td>106</td>
</tr>
<tr>
<td>Marking Test and Quantitative Scale</td>
<td>109</td>
</tr>
<tr>
<td>Age Comparisons</td>
<td>112</td>
</tr>
<tr>
<td>Age and Timed Vocabulary</td>
<td>113</td>
</tr>
<tr>
<td>Age and Untimed Vocabulary</td>
<td>115</td>
</tr>
<tr>
<td>Vocabulary Timed and Untimed</td>
<td>118</td>
</tr>
<tr>
<td>Age and Linguistic Scale</td>
<td>120</td>
</tr>
<tr>
<td>Linguistic Scale and Timed Vocabulary</td>
<td>124</td>
</tr>
<tr>
<td>The Linguistic Scale: Partial Correlations</td>
<td>130</td>
</tr>
<tr>
<td>Age and Timed Arithmetic</td>
<td>131</td>
</tr>
<tr>
<td>Age and Untimed Arithmetic</td>
<td>131</td>
</tr>
<tr>
<td>Arithmetic, Timed and Untimed</td>
<td>136</td>
</tr>
<tr>
<td>Age and Quantitative Scale</td>
<td>140</td>
</tr>
<tr>
<td>Quantitative Scale and Timed Arithmetic</td>
<td>144</td>
</tr>
<tr>
<td>The Quantitative Scale: Partial Correlations</td>
<td>145</td>
</tr>
<tr>
<td>Vocabulary Groups within the Experimental Population</td>
<td>146</td>
</tr>
<tr>
<td>Comparison of the Preliminary and Experimental Groups</td>
<td>152</td>
</tr>
<tr>
<td>CEE Timed Vocabulary</td>
<td>153</td>
</tr>
<tr>
<td>ACE Linguistic Scale</td>
<td>155</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>ACE Quantitative Scale</td>
<td>187</td>
</tr>
<tr>
<td>Percentile and Decile Comparisons between Age Groups and the Vocabulary,</td>
<td></td>
</tr>
<tr>
<td>Linguistic and Quantitative Tests</td>
<td>159</td>
</tr>
<tr>
<td>V. SUMMARY, CONCLUSIONS AND DISCUSSION</td>
<td>166</td>
</tr>
<tr>
<td>Summary of the Preliminary Study</td>
<td>166</td>
</tr>
<tr>
<td>Summary of the Experiment</td>
<td>169</td>
</tr>
<tr>
<td>Conclusions</td>
<td>179</td>
</tr>
<tr>
<td>Discussion</td>
<td>183</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>189</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of CEE Vocabulary Scaled Scores, Form Z, on Age for the Preliminary Study (N = 509)</td>
<td>64</td>
</tr>
<tr>
<td>II. Table of Medians, Means and Standard Deviations for CEE Vocabulary Scaled Scores, Form Z, on Age for the Preliminary Study</td>
<td>67</td>
</tr>
<tr>
<td>III. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of ACE Linguistic Raw Scores, 1958 Edition, on Age for the Preliminary Study (N = 585)</td>
<td>69</td>
</tr>
<tr>
<td>IV. Table of Medians, Means and Standard Deviations for ACE Linguistic Raw Scores, 1958 Edition, on Age for the Preliminary Study</td>
<td>71</td>
</tr>
<tr>
<td>V. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of ACE Quantitative Raw Scores, 1958 Edition, on Age for the Preliminary Study (N = 585)</td>
<td>73</td>
</tr>
<tr>
<td>VI. Table of Medians, Means and Standard Deviations for ACE Quantitative Raw Scores, 1958 Edition, on Age for the Preliminary Study</td>
<td>75</td>
</tr>
<tr>
<td>VII. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of Marking Test Raw Scores on Age for the Experimental Group (N = 193)</td>
<td>92</td>
</tr>
<tr>
<td>VIII. Means for the Regression of Age on the Marking Test for the Experimental Group (N = 193)</td>
<td>96</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>IX. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of TIMED Vocabulary Scaled Scores, Form Z of the CEE, on the Marking Test for the Experimental Group (N = 193)</td>
<td>97</td>
</tr>
<tr>
<td>X. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of UNTIMED Vocabulary Scaled Scores, Form S of the CEE, on the Marking Test for the Experimental Group (N = 193)</td>
<td>99</td>
</tr>
<tr>
<td>XI. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of ACE Linguistic Raw Scores, 1952 Edition, on the Marking Test for the Experimental Group (N = 193)</td>
<td>102</td>
</tr>
<tr>
<td>XII. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of TIMED Arithmetic Raw Scores, 1952 Edition of the ACE, on the Marking Test for the Experimental Group (N = 193)</td>
<td>105</td>
</tr>
<tr>
<td>XIII. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of UNTIMED Arithmetic Raw Scores, 1948 Edition of the ACE, on the Marking Test for the Experimental Group (N = 193)</td>
<td>107</td>
</tr>
<tr>
<td>XIV. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of ACE Quantitative Raw Scores, 1952 Edition, on the Marking Test for the Experimental Group (N = 193)</td>
<td>110</td>
</tr>
<tr>
<td>XV. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of TIMED Vocabulary Scaled Scores, Form Z of the CEE, on Age for the Experimental Group (N = 193)</td>
<td>114</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>XVI. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of UNTIMED Vocabulary Scaled Scores, Form S of the CKB, on Age for the Experimental Group (N = 193)</td>
<td>116</td>
</tr>
<tr>
<td>XVII. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of TIMED Vocabulary Scaled Scores, Form Z of the CKB, on UNTIMED Vocabulary Scaled Scores, Form S of the CKB, for the Experimental Group (N = 193)</td>
<td>119</td>
</tr>
<tr>
<td>XVIII. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of ACE Linguistic Raw Scores, 1952 Edition, on Age for the Experimental Group (N = 193)</td>
<td>123</td>
</tr>
<tr>
<td>XIX. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of ACE Linguistic Raw Scores, 1952 Edition, on TIMED Vocabulary Scaled Scores, Form Z of the CKB, for the Experimental Group (N = 193)</td>
<td>127</td>
</tr>
<tr>
<td>XX. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of TIMED Arithmetic Raw Scores, 1952 Edition of the ACE, on Age for the Experimental Group (N = 193)</td>
<td>132</td>
</tr>
<tr>
<td>XXI. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of UNTIMED Arithmetic Raw Scores, 1948 Edition of the ACE, on Age for the Experimental Group (N = 193)</td>
<td>133</td>
</tr>
<tr>
<td>XXII. Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of UNTIMED Arithmetic Raw Scores, 1952 Edition of the ACE, for the Experimental Group (N = 193)</td>
<td>137</td>
</tr>
</tbody>
</table>
TABLE PAGE


XXIV. Comparison of the Mean Test Scores of the Total Experimental Group with the Mean Test Scores of an Out-of-Pattern High Scoring Vocabulary Group and a Residual Group which together compose the Experimental Population 149

XXV. Mean Percentile and Decile Ranks on the CEE Timed Vocabulary, Form Z, and the ACE Linguistic and Quantitative Scale, 1952 Edition, for Age Groups of the Preliminary and Experimental Studies .............. 160

XXVI. Summary of Results for the Experimental Study: Pearson Product-Moment Correlation Coefficients and Eta Coefficients of Test Comparisons with the Marking Test and with Age for the Experimental Group (N = 193) .......... 172
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>DESCRIPTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The Marking Test</td>
<td>85</td>
</tr>
<tr>
<td>2.</td>
<td>Mean Scores and Standard Deviations for Marking Test Scores on Age for the Experimental Group (N = 195)</td>
<td>94</td>
</tr>
<tr>
<td>3.</td>
<td>Mean Scores and Standard Deviations for Timed and Untimed CEE Vocabulary Subtests (Forms Z and S) on the Marking Test for the Experimental Group (N = 195)</td>
<td>100</td>
</tr>
<tr>
<td>5.</td>
<td>Mean Scores and Standard Deviations for ACE Arithmetic Raw Scores (Untimed and Timed), 1948 and 1952 Editions, on the Marking Test for the Experimental Group (N = 195)</td>
<td>108</td>
</tr>
<tr>
<td>7.</td>
<td>Mean Scores and Standard Deviations for CEE Vocabulary Scaled Scores, (Timed and Untimed), Forms Z and S, on Age for the Experimental Group (N = 195)</td>
<td>117</td>
</tr>
<tr>
<td>8.</td>
<td>Mean Scores and Standard Deviations for TIMED Vocabulary Scaled Scores, Form Z of the CEE, on UNTIMED Vocabulary Scaled Scores, Form S of the CEE, for the Experimental Group (N = 195)</td>
<td>121</td>
</tr>
<tr>
<td>FIGURE</td>
<td>PAGE</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>11. Mean Scores and Standard Deviations for ACE Arithmetic Raw Scores, (Timed and Untimed), 1952 and 1948 Editions, on Age for the Experimental Group (N = 193)</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>14. Histogram comparing the Out-of-Pattern High Vocabulary Group (N = 85) with the Residual Group (N = 110) on Age for the Experimental Group</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>15. Histogram comparing the Out-of-Pattern High Vocabulary Group (N = 85) with the Residual Group (N = 110) on the Quantitative Scale for the Experimental Group</td>
<td>151</td>
<td></td>
</tr>
<tr>
<td>16. Comparison of Means and Standard Deviations for Timed CEE Vocabulary Subtest Scaled Scores, Form Z, on Age for the Preliminary Group (N = 599) and the Experimental Group (N = 193)</td>
<td>154</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION TO THE PROBLEM

California State Colleges are required by law to evaluate the scholastic ability and achievement in English skills of all candidates for public school credentials. The selection of tests to be used for this purpose is left to the discretion of the individual state colleges.

Sacramento State College has been using the American Council on Education Psychological Examination for College Freshmen (ACE) and the Cooperative English Examination, Higher Level (CEE), for this purpose since 1949. Both tests have been used on a nationwide basis for many years and the test authors have developed norms which permit of comparison between individual colleges and similar colleges throughout the country. Both tests are widely accepted as fulfilling the requirements of rigorous standardization.

The national normative standards are homogeneous as regards the class standing of the individuals tested, but little information is available about the homogeneity of age levels in the sample populations.

I. THE PROBLEM

Statement of the Problem. It was the purpose of this study to compare the test performance of various age groups
represented in the population of credential candidates at Sacramento State College in order to discover what relationships exist between these groups both within tests and between tests. A second purpose was to study the performance of age groups on a test designed to isolate a complex of functions judged to be implicit in the completion of standard IBM answer sheets under timed conditions. A third purpose of the study was to evaluate the relevancy of existing normative standards for different age groups.

**Origin of the Problem.** Screening tests are administered to credential candidates at the beginning of each regular semester and summer session by the Test Officer of the College. Test results are submitted to the Committee on Credential Candidates for action. The author of this study assists in that part of the screening process which is conducted by the Test Officer before the graphic data are given to the Committee on Credential Candidates.

Over a period of time it has been observed that when certain relationships occur between the various subtests of the ACE and CEE, one may frequently predict that the individual showing this pattern is older than the average college junior and senior student. Up to the present there has been no analysis of the frequency of these deviations in test performance.
Specifically, more older credential candidates have been observed to show an outstanding peak on the Vocabulary subtest of the CEE than younger students. Also, it has been noted that the ACE Quantitative score is often depressed both in relation to total individual test performance and to the performance of the group as a whole for the older students.

Each student must meet approved standards of scholastic ability and English achievement before application for credential candidacy may be approved. Since the limitations of group test measurement for all individuals are well known, those students who fall below approved test score standards may be referred by the Credential Committee to the Test Officer for further test evaluation. The Ohio State Psychological Examination for College Freshmen is frequently used for this purpose. Older students have often demonstrated a much higher level of proficiency on this untimed power-type task than on the timed tests of the ACE, CEE battery.

These observations led to a decision to make a preliminary study of credential candidates who were tested in the regular and summer sessions of 1954. Results of the study, given in Chapter III, confirmed the impression that there are significant differences in test performance for different age groups. This further evidence of the influence of age initiated the experiment with which this study is concerned.
Delimitations. A study which is limited to making comparisons between test scores effects little except to reveal that differences exist. Nevertheless, this is the first step in the delineation of a problem. For this purpose, a preliminary study of test data for a large group of credential candidates extending over a wide age range was conducted to discover what differences exist between age groups on the Linguistic and Quantitative sections of the AE Psychological Examination and on the Vocabulary subtest of the Cooperative English Examination (CEE). When significant relationships with age were found within and between these parts of the total screening battery, the decision was made to conduct an experiment on a smaller group of candidates taking courses in the summer session of 1955. The observations made in the preliminary study were extended by giving the CEE Vocabulary and ACE Arithmetic subtests on an untimed basis in addition to the usual screening tests.

In addition, a short test was devised for the purpose of isolating a complex of functions necessary to the use of IBM answer sheets under timed conditions. The purpose of including this latter test was to determine whether variables are being measured which are unrelated to the skills and abilities which are intended to be measured by the content of the various parts of the battery.
This part of the study establishes the differences in test performance that exist throughout the age range of the population studied. Interpretation of the data is more difficult. The problem may be delineated by posing the following questions. How does the test performance of credential candidates at Sacramento State College compare with that of other populations? What is the fate of other measurements of intelligence, vocabulary and arithmetic over the span of adult years? Is there any test criterion against which mental functioning in a variety of situations may be evaluated? Are individual differences in test scores an adequate basis for assuming that real differences in mental efficiency exist?

Tentative answers to these questions, in so far as they relate to the present study, were sought in the literature.

Answers to the first two questions concerning population comparisons and the fate of test scores with age are given in terms of general trends which appear to be supported by numerous investigations.

To answer the question of whether there is a stable test criterion of mental ability against which mental efficiency in a variety of situations may be evaluated, the position of vocabulary tests was given specific attention. (The term mental ability as used in this study refers to the
highest level of mental functioning attained by an individual in his life span. The term mental efficiency refers to the present level of mental functioning whatever the previous level may have been. This is an important distinction in studies concerning the growth and decline of psychological variables in the human organism.)

The third question of whether individual differences in test scores reflect real differences in mental efficiency can be established only by going beyond the purpose of the present study. One step in this direction, for example, would be an analysis of grade point averages to discover whether the various elements in the screening battery are as predictive of college achievement for older adults as for younger students. The above question challenges the effectiveness of present tests for adult measurement both as these relate to internal test variables such as content and timing, and as they relate to human variables which may undergo change, with or without affecting efficiency in life situations, as a function of age.

A limited attempt was made to answer this question. First, studies of the effect of timing on test scores in adult populations were reviewed. Accumulated evidence that speed and intelligence are positively correlated has led to the general and economical practice of incorporating a time element into many paper and pencil tests for group measurement.
This practice has been justified on the basis of studies of school age subjects with a tendency to generalize to adult groups. There is general agreement among investigators that physiological functions, including speed of motor reaction, decline with age, but with impressive individual differences at every age level.

Secondly, the fate of arithmetic test scores with age were reviewed and compared with the results of Arithmetic timed and untimed in the present study. In this comparison an assumption is made that arithmetical ability depends in large degree on learned functions. To the extent that these functions are not continuously used in the life situations of maturity, they may be said to show the effects of disuse. These skills may be more readily available for reuse in the young college age student than in the adult whose interests and occupations have not required the application of arithmetical procedures.

Third, the problem of test content as it relates to adult measurement and the problem of the concomitance of psychological and physiological decline with age were sufficiently reviewed to frame the limited context within which test results can be interpreted.
II. IMPORTANCE OF THE STUDY

The Local Contribution. Heretofore, the ACE and CEE test scores of the credential candidate population have been converted to percentiles on the basis of a single set of norms for each test. The class standing of the population has ranged from high Sophomore through Graduate status and the age range has become extreme, varying from 18 years to as high as seventy years of age. There are recognized limitations in the present uniform method of predicting scholastic ability and English achievement for this diverse age group.

Until recently this has involved a relatively small number of individuals and when test scores have been in question, students have had recourse to further evaluation. In a majority of instances the retest results have led to a reconsideration of the candidate. Two factors of local interest suggest that this study is timely. The recent implementation of legislation pertaining to the renewal of provisional credentials initially granted after July 1, 1954 has resulted in a marked increase in the number of older students who must take the screening battery of tests at Sacramento State College.

Secondly, the population increase in the State, the serious need for teachers and the anticipated expansion of
of the college to meet this need, suggest that it is appropriate to evaluate the influence of age on test performance in the present battery. If this is found to be a real source of variance, the study may provide evidence to support either a change in test content or in predictive standards for age groups. It is desirable and economical to have a program which does not require an unduly elaborate, time-consuming and costly retest procedure, and equally desirable from the viewpoint of the individual candidate to challenge his or her effectiveness with a minimum of trauma.

The General Contribution. In recent decades there has been a decided increase in the number of studies related to the many aspects and problems of aging. Psychologists who have developed tests for the measurement of abilities and skills in young people have had ready access to large school populations. Older individuals are readily accessible for measurement and study only in times of war or other unusual circumstances. With a few noteworthy exceptions, investigators of the adult and aged have had to rely on the more or less unusual and selected populations found in mental institutions, hospitals, clinics, industry and the like. So far as can be determined from a review of the field, the college population of the present study is nearly unique in respect to the wide age range of individuals seeking a common goal.
The study is limited by being cross-sectional in nature, by the small number of measurements involved, the nature of the tests used and the imposed procedural limitations. Nevertheless, it may be expected to be a small contribution to the problem of aging in general and of particular interest to other colleges with a similar population and problem.
CHAPTER II

REVIEW OF THE LITERATURE

The psychometric performance of adults has been purposefully studied only within the last three decades. Impetus for investigation of this area resulted from the extensive use of the Army Alpha in World War I which gave the first quantitative estimate of declining abilities with age (83). Studies now existent in the literature are numerous and beyond the purpose of this chapter to recount in full. Several historical and current reviews are available.1

The tests of the present study have not been used in the measurement of a wide span of adult years with the exception of a study by Sorenson (67) in which an early form of the American Council on Education Psychological Examination was analyzed with respect to adult performance. Therefore, in exploring the literature it was necessary to depend on the results of a variety of tests to establish general trends in relation to mental functions with increasing age. This is justified to the extent that test contents appear to be similar and to measure similar functions. Intertest

---
1Brody (12), Cattell (14), Eysenck (19), Novland (28), Jones & Bayley (30), Kaplan (32), Lawton (34), Lewinski (35), Lorge & Rush (36), Lorge (37, 38), W. R. Miles (48, 49), Shock (64, 65), Vincent (76).
Correlations of mental tests are generally high enough to give credence to the assumption of relative similarity of functions measured.

A parallel problem exists in relation to the population of the present study. Again, Sorensen (67, 68, 69) is the only investigator found to work with a similar population sample. However, it may be assumed that when test findings in relation to age are similar in a variety of populations, these results may be used to establish the general trend or expectancy.

Much of the evidence on the relationship of test scores to age bears a marked resemblance to findings for the credential candidates at Sacramento State College with a few recent and important exceptions. However, the danger inherent in accepting general trends as corroborative of present results is exemplified in the controversial problem of whether tests designed for the young are appropriate for the measurement of adults. The distinction between the maturation of intellectual functions in the young and application of these in maturity is not yet clearly defined.

There is also question as to whether psychological decline parallels physiological decline or whether physiological variables are relatively extraneous factors measured incidentally by many tests in present use. Differentiation needs to be made between the effects of physiological decline
on psychological efficiency in lifelike rather than in ir-
realistic and forced test situations. The young may meet
such demands more effectively because of higher physiolog-
ical ceilings, but do these demands have their counterpart
in the majority of life situations?

In the review which follows, it is important to re-
member that results will vary depending on test content,
methodology and definition, on physiological variables, on
the populations used, education and many other factors. It
is also well to keep in mind that even though there is evi-
dence of general trends as a function of age, individual
differences within age groups tend to exceed differences be-
tween age groups.

Measurements of intelligence (scholastic ability),
vocabulary, arithmetic and effect of timing for adult groups
are of specific interest to this study. No direct evidence
has been found pertaining to the interpretation of directions
as this affects speed and precision of marking an IBM answer
sheet under standard conditions.

Investigations comparing intelligence test scores
with age will afford general comparisons with the ACE test
of scholastic ability used in the present study.

A purpose of this review is to establish that vocabu-
lar y tests are a relatively stable and independent criterion
of mental ability for a wide range. If there is evidence
that this is true, then the CBE Vocabulary subtest of the college screening battery may be used as a clue for judging whether or not an individual candidate has performed effectively on other parts of the battery. When subjects demonstrate a wide difference between vocabulary and other test scores, this may represent a true loss of efficiency or factors which are not relevant to the testing purpose.

ACE Arithmetic scores may be influenced by a number of variables related to age. The test is rigorously timed and performance may be affected by speed factors. Older students may also be more distant from formal training in arithmetical procedures than young students and, hence, forgetting and delayed recall may effect age differences in performance. One purpose of this review will be to gather evidence concerning the relationship of arithmetic with age in other investigations.

Since the effects of timing on adult test performance are a consideration in the study, this will be a fourth area of research in the literature.

Investigations reported here are limited to those which are most clearly related and pertinent to the present study and which have not been subjected to vitiating criticism by later authors in terms of population, method and the like. The purpose is to outline general trends rather than to undertake a critical review.
Age and Decreasing Intelligence. The reviews which follow report negative correlations between age and intelligence test scores for adult groups. This is in general agreement with findings on Army Alpha test results for World War I (83). However, the majority of investigators take the viewpoint, like Weisenburg, Roe & McBride (80, p. 11) that the question as to the course of different test performances with age cannot be solved on the basis of army findings, for at different ages there were probably differing selective factors which made the various age groups incomparable.

World War II presented a second opportunity to obtain test measurements on a large population sample. Tuddenham (1948: 78) reports a comparison between 48,102 World War I white enlisted soldiers on the original Alpha and a representative sample of 769 subjects on the Wells Revision of the Army Alpha. Tuddenham found that the average score for the later group is 33 percentile points higher than that of the World War I group. The median raw score of 62 for the latter group reaches only the twenty-second percentile of the World War II distribution. Conversely, the median raw score of 104 for the World War II group falls at the eighty-third percentile of the World War I population, a difference which is highly significant. Tuddenham finds no satisfactory answer concerning sampling differences in the two populations, but appears to take the stand that changes in the American population take place when he suggests that such factors as experience with tests, improved health and nutrition, and
increased education may partially account for the trend found in his study.

These contrasting results, found on very similar tests over a span of time, are offered at the outset because of the possibility that test results characteristic of one era may not be characteristic of later eras.

One of the first studies following World War I was that of Foster & Taylor (1920: 21) who studied the applicability of mental tests to persons over 50 years of age by using the Yerkes-Bridges Point Scale on a group of younger and older normal and psychotic persons. The study showed a general falling off of test scores for the older normal and psychotic groups, although this varied with the nature of the test material. These authors made the important suggestion for future research that the intelligence of older subjects should be evaluated both in terms of their own normal contemporaries and with their own adolescent ability or that of normal young people.

Beginning in 1931, W. R. Miles and his associates published an important series of papers composing the Stanford Later Maturity Studies. G. O. Miles and W. R. Miles (1952: 46) tested over 800 persons from two fair-sized American towns on a 15 minute form of the Otis Self-Administering Test of Mental Ability. The age range was from 7 to 95 years. They found that ability was greatest at 18, remained fairly
constant to 45, then rapidly declined. The differing relationships between age and intelligence are expressed in correlation coefficients of .80 for the age range from 7 to 17 years, -.28 for subjects between 20 and 30 and -.50 for the wide range from 20 through 95 years of age.

C. C. Miles (1934:45) extending this work, found a correlation of approximately -.50 between age and the 15 minute Otis intelligence score at each of three educational levels in about 2000 adults aged 20 to 95, of whom more than 250 were 70 or more years old. She retested 190 persons from this group after a two year interval and found that the decrements associated with increasing age were actually larger than those predicted from the previous study, thus strengthening the significance of the earlier evidence of decline.

Another early study on the relationship of intelligence to age was made by Jones & Conrad (1933: 31) who gathered Army Alpha test data in two New England towns for 547 persons between the ages of 25 and 59 years, and 644 subjects between the ages of 10 and 24. The chief characteristics of their developmental curve for the total Alpha test are summarised as involving a linear growth to about 16 years, with a negative acceleration beyond 16 to a peak between the ages of 18 and 31. A decline follows which is much more gradual than the curve of growth, but which by the age of 35 involves a recession to the 14-year level. (p. 261)
The authors are confident that the decline of adult ability which is evident beyond the age of 21 is not due to sampling, faulty test administration, poor motivation, remoteness of schooling, failing visual acuity or numerous other causes.

Gilbert (1935: 27), using the Babcock test, also demonstrated a loss of efficiency with age. She selected an experimental group of 175 subjects of average age 65.8 and a younger control group of 185, average age 25. The younger and older groups were matched for vocabulary level on the assumption that vocabulary is a relatively stable index of intelligence in the mature years. The groups were assumed to be comparable in occupational status and motivation. She found a decline with age, though in varying degrees, on all subtests.

Weisenburg, Roe & McBride (1936: 30) report results of an extensive and varied battery of tests involving 15 to 20 hours of testing time per person. These were given to a total of 70 men and women ranging in age from 20 to 59 with a mean age of 36.1. The population was derived from the surgical and orthopedic wards of three hospitals and only subjects without disabilities which would affect test performance were included. The authors were assured of a random and unselected sample of the adult population. They found that the greatest extent of mental development is reached before the twenties and that from this decade through
the fifties there is little further gain and comparatively little decline. Like other investigators, they found that different tests make differing contributions at various age levels and that some subtests show increased scores with age.

Irving Lorge (1956: 42), who conducted a study in which he was primarily concerned with the effects of timing on measurements of adult ability, gave a battery of speed and power tests to a group of 132 persons, ages 20 to 70. The Otis S-A and Army Alpha were administered as speed tests and the CAVD as a power test. The correlations between age and score were found to be -.48, -.56 and -.27 respectively. This study is a good example of the differing results that are found with different tests and will be discussed later in the section on age and speed.

Copeland (1956: 16), to eliminate the effects of lowered speed of reaction time with age, administered the Otis on a work-limit rather than a timed basis and found a slight decrease in average mental ability from the age group 20 to 25 up to the ages from 55 to 60 in a sample of 7500 clerical applicants at an employment office, but this decrease was less than that reported by Miles and Miles (48) who used the Otis test with a 15 minute time limit.

Wechsler (1944: 78) was one of the first psychologists to become disatisfied with methods of measuring and computing adult intelligence. In 1959 he published the first scale for
the measurement of adult subjects, developing norms for the various age groups rather than deriving these from extensions of data for the young or from setting a top chronological age to be used in the computation of intelligence. On data obtained with the individually administered Bellevue Intelligence Examination on subjects from ages 7 to 65, Wechsler finds a curve similar to that of Miles (51) and Jones and Conrad (31). Wechsler finds that what they all show is, that beginning at an age varying from 15 to 22, all scores of mental ability, far from remaining constant, start to fall off. The point at which the direction changes and the rate at which it progresses vary from test to test and cannot be fixed.

It is true that for most intelligence scales the differences between age 15 and 25 are for the most part negligible, but that above that age the decline becomes appreciable and increasingly important. (78, p. 30)

Wechsler finds a more or less linear decline between the ages of 30 and 60 on the Bellevue. At 50-59, efficiency is 85 per cent of that at 25-29.

From an observed parallelism he hypothesizes that mental decline is part of the general organic process of physiological decline which begins relatively early in life, and that mental decline occurs at a somewhat more rapid pace than physical decline. As has been found in group type tests, the different subtests of the Bellevue demonstrate different
rates of decline with age, and variability of performance tends to increase with increasing age among adult groups.

To meet existing needs and criticisms of the original Bellevue scale, Wechsler in 1955 (77) made available a revision, the Wechsler Adult Intelligence Scale (WAIS). This was standardized on a nationwide sample of 1,700 adults, prorated according to the 1950 U. S. Census and including a proportionate representation of the non-white public. IQ tables are extended to cover older persons to age 75 years and above with better representation in the older groups than in his first scale. Supplementary tables make possible a comparison between an individual's performance on each subtest and persons in the same age group. Large (40) points out that the norms on this nationwide sample indicate that intelligence does not decline as markedly as Wechsler has previously indicated. In the new data, the age range 26 to 34 years yields the highest raw scores for the verbal sections with very small decrements to the middle fifties.

**Age and Increasing Intelligence.** The preceding investigations support a general conclusion of rapid intellectual growth to the early twenties and very moderate to fairly rapid decline with age beyond this point. Not all of the evidence is in agreement with this conclusion and, strikingly, this evidence is of very recent origin.

In 1942, Lewinski (33) published an analysis of the
performance of 1,000 men in the U. S. Navy on the Wechsler-Bellevue in an age range from 17 to 62 years. He found no marked differences in mean IQ at the various chronological age levels, and no uniform increase or decline with advancing age. A positive correlation of .13 was found between the total scale and age which is inconsistent with the negative relationship usually obtained in this age range. The author assumes a random sample of U. S. white males, but Jones & Bayley (30) think it probable that the test group was more highly selected at the upper ages.

Corsini and Fassett (1953: 17) conducted a study in which the Wechsler-Bellevue scores of 1,072 prison inmates ranging in age from 15 to 75 years were selected from a pool of 4,000 cases. They hypothesize that "the decline noted by other investigators is mainly a function of two conditions: poor sampling, and the loading of non-intellectual factors in the test" (p. 253). They used a "forced" sampling method to obtain 100 cases for each age interval of 5 years from 15 through 59. Fewer cases were available beyond this age range, but the representation was larger than Wechsler's 1939 standardization group. A 9 per cent increase was found on the Verbal Scale of the Bellevue in contrast to the 8 per cent drop in Wechsler's data. Information and Arithmetic were found to rise significantly with age. The rate of decline on the Performance Scale was found to be similar to that of
Wechsler, and the authors consider that this part of the test contains non-intellectual visual and motor factors.

The investigations reported on to this point have been cross-sectional in nature, that is, different individuals have represented each age interval with no certainty that their test scores would have been comparable had they all been tested initially at the same age. This is a serious limitation of the cross-sectional approach.

W. R. Miles (48) was one of the first investigators to recognise that "the object of a genetic psychology of maturity is a longitudinal survey of the human-mental being in our familiar contemporary terrestrial environment." There is recent evidence, based on this approach, which, like the two preceding cross-sectional studies, indicates that intellectual ability may continue to increase beyond the early limits established by previous workers.

Bayley and Oden (1955: 8) have continued the research on the Stanford Study of the Gifted which was first undertaken by L. M. Terman more than thirty years ago. Their latest report is based on test and retest results of 1,103 adults, 768 of whom were selected by Terman as gifted children, and 335 of whom are spouses of these subjects. Two forms of the Concept Mastery test were administered: one in 1939-40 and the second in the 1950-52 retest. The Concept Mastery test is composed of two subtests: I. Synonyms and
Antonyms and II. Analogies, both given without time limit. The authors found a highly significant increase in mean scores at the second testing for both the gifted and their spouses. Increases occurred in all occupational and educational levels, at all levels of ability -- except where test ceilings prevented -- and at all ages from 20 to 50 years. Gains were found in both sections of the test, but were somewhat larger for Synonyms and Antonyms than for Analogies. Gifted subjects show a decrease in variability at the second testing and with increasing age. The spouses' scores were more variable and increased on the retest. The authors conclude that intelligence of the kind measured by the Concept Mastery test continues to increase at least through fifty years of age.

These findings are supported by another recent study. In 1953 Owens (55) retested 127 college men on the Army Alpha after a lapse of 31 years. His original purpose was to discover how much of an expected age decrement in performance was attributable to speed, but no significant decrements were found on any subtest. Instead, significant increases were found on all the verbal tests and insignificant changes in the two numerical type tests. Total scores improved by .555 standard deviations over the 31 year interval. These results are especially interesting since the test instrument, the Army Alpha, is one which in earlier age investigations
has given evidence of decrements in score with increasing age.

On the basis of the data from these two longitudinal studies, Bayley (7) has devised a growth curve extending to fifty years which shows rapid increases in score to about age 20 and beyond this point a slight but steadily upward linear increase to the age of fifty. She points out, however, that a similar growth curve might not be found in a lower ability group.

These few studies showing increases in test scores with age are suggestive in that they permit the speculation that cultural changes may effect alterations in the fate of mental abilities with age.

**Age and Constancy of Vocabulary.** A purpose of this review is to establish that vocabulary tests are a relatively stable and sufficient criterion of mental ability throughout life. To support this hypothesis, two kinds of evidence are required. First, vocabulary level should have high correlations with tests of intelligence. Secondly, vocabulary should demonstrate little or no correlation with age over a wide span of years. This evidence can be found in studies of both normal and abnormal populations. The latter are included here to demonstrate the relative refractoriness of vocabulary even when the life situation is highly distressed.
As early as 1916, Terman (72) contended that the Vocabulary test of his Binet revision is more valuable than any other single test of the scale. Terman and Merrill (73) continued to maintain this view in respect to the 1937 revision. They found a correlation of .91 between mental age and the Vocabulary test for 651 children. A few years later, Spache (1943: 70) obtained a similar correlation of .915 between Vocabulary and the Stanford-Binet in an above average population of 65 children. These two studies indicate that mental ages of children may be predicted with considerable accuracy from the Vocabulary section of the test.

Terman and Merrill found a correlation of .81 between mental age and Vocabulary for 482 "miscellaneous adults." Helen Green, cited in a footnote by the above authors (73, p. 303) studied a group of 110 adults ranging in age from 19 to 84 years. Care was taken to secure a representative sample according to economic and social status. Results indicated that maturity effected no great change in vocabulary ability. The small correlation of .09 was obtained between age and Stanford-Binet Vocabulary.

When Milas (1933: 47) analyzed items on the 15 minute Otis S-A for 666 adults ranging in age from 15 to 94, his results indicated that verbal associations and interpretations of meaning show marked resistance to the influence of age.
Jones & Conrad (1933: 31) in their study of 1,191 subjects from age 10 to 60 on the Army Alpha found that Subtest 4 (Synonyms-antonyms), which is the nearest to a vocabulary ability test in the Alpha, did not show the post-adolescent decline with age found in all other tests except 8, a general information test.

O'Connor (1934: 54) compared the vocabulary scores of two groups of "equally successful" business executives and found that those who held a college diploma did no better than the group who had discontinued their formal education when they were 18 years old.

Shakow & Goldman (1938: 61) tested 343 men and women ranging in age from 18 to 89 on the Stanford-Binet Vocabulary test and found a correlation of -.10 between vocabulary and age. They report a slight increment in vocabulary up to and beyond age 50 and a slight but insignificant decline after age 60. When the age groups were equalized for intellectual level by obtaining representative educational samples for each decade group, vocabulary was not affected by age until the seventh decade. The authors conclude from their own study and a review of other investigations that the vocabulary test appears to be an adequate device for determining adult intelligence. Vocabulary in this study appears to be little affected by age or by increasing separation from formal training.

When Wechsler (1944: 78) constructed and standardized
the Bellevue scale for adults, the Vocabulary subtest was first included only as an alternate test since it was feared that vocabulary level might be unduly influenced by educational and cultural opportunity. However, study indicated that "the size of a man's vocabulary is not only an index of his schooling, but also an excellent measure of his general intelligence" (p. 99). Its value appears to be derived from the fact that "the number of words a man knows is at once a measure of his learning ability, his fund of verbal information and of the general range of his ideas" (p. 99). For the computation of an Efficiency Quotient for individuals, Vocabulary is treated as one of the "hold tests." Wechsler reports an Eta coefficient of .65 between Vocabulary and the total Bellevue scale, but does not report a linear coefficient.

Using the Gallup poll to obtain a representative sample, Thorndike & Gallup (1944: 74) administered a short form of the multiple-choice vocabulary section of the CAVID to 2,974 men and women. They conclude that age differences are insignificant until the 50 to 59 age interval and that only in the over-60 group was any substantial drop found. Median scores for age intervals decrease very slightly from 11.34 for the 21 to 29 age group to 10.28 for the over-60 group.

Fox (1947: 24) administered vocabulary tests to 50 subjects between the ages of 70 and 79 and to 50 subjects between the ages of 40 and 49 who constituted a sample of the general
population at those ages with regard to the amount of formal education and socio-economic status. Both multiple-choice and definition type vocabulary tests were given and no significant differences were found between the two age groups on either of the tests. The definition task was more difficult than the recognition task for both age groups, but there was no significant difference in the quality of the definitions given by the two groups.

Lewinski (1948: 35) in his study of 1000 men in the U. S. Navy, ages 17 to 62 years, on the Wechsler-Bellevue, found Vocabulary to be relatively stable for this population with no well-defined increments or decrements of vocabulary score with advancing age. The correlation between age and vocabulary was .168. Correlations between Vocabulary and the Performance, Verbal and Total Bellevue Scale were .684, .809 and .816 respectively.

Poulds & Raven (1948: 23) and Poulds (1949: 22) investigated age changes in intelligence over a span of sixty years in a population of more than 5,000 unskilled and skilled men, administrative workers and technical and university students. Scores on the Mill Hill Vocabulary Test were compared with scores on the non-verbal Progressive Matrices test. The authors state that "together the tests provide an estimate of the best intellectual level a person has reached relative to other people, and his present capacity for intellectual
activity" (p. 258). Comparing age with test scores, the Matrices test was found to decline from the age of 25 onward, but the vocabulary test scores increased with age up to about 30 years and beyond this point remained relatively constant. Foulds has shown that there is a decrement in vocabulary ability beyond age 30 when just the lower 25th percentile is considered, whereas in the 95th percentile of the same group there was a slight but measurable increment with age up to the 55th year.

Fox & Birren (1949; 25) found no relationship between vocabulary and age in a study of 218 subjects with an average age of 71 in a range from 40 to 99 years. They used the Wechler-Bellevue and the Seashore Bokerson Vocabulary tests on a population which was derived from a home for older persons.

The stability of vocabulary as a measure of intelligence throughout life is given additional credence by investigations of abnormal and brain-damaged subjects.

Babcock (1930; 5) was one of the first clinicians to devise a test for measuring intellectual impairment. Her method is based on the idea that the extent of loss can be evaluated by comparing a patient's subtest scores with the scores made by normals of equal "vocabulary age." This implies that vocabulary is most resistant to encroachments of mental disorder and may be used as an index of original intellectual level. In her Revised Examination for the Measurement of
Efficiency of Mental Functioning (1940: 4), the Terman-Merrill vocabulary test continues to hold the position given it in the earlier test.

Jeanne Gilbert (1935: 26), using a revision of the Babcock test, confirmed the constancy of vocabulary in a comparison of two groups of people, one consisting of 175 persons whose average age was 63.8 years and a second group consisting of 185 subjects, average age 25. One result of her study was the finding that age alone is an insufficient index by which to judge ability to meet the requirements of a particular vocation.

Shipley (1940: 62) devised a self-administering scale of mental deterioration based on the same principles as Babcock's test and when he and Burlingame (1941: 65) conducted a study of 374 mental patients, they concluded that vocabulary level is a useful index of intellectual level.

Hunt (1943: 29), also using a modification of the Babcock technique, devised a test for the detection of organic brain damage. When a battery of tests was administered to a group of 35 brain-damaged subjects and a non-damaged control group of 44, the correlation between the total score on the deterioration tests and vocabulary was .51, between scores and age -.57, and between age and vocabulary scores .07. This the authors state, is consistent with the findings of previous investigations between these several variables.
Roe & Shakow (1942: 59) compared the performance of 887 state hospital patients with varying diagnoses on the 1916 Stanford-Binet with that of normal subjects and found a discrepancy between vocabulary and other scores in all groups. Correlations between vocabulary and mental age ranged from .76 in psychoneuroses to .83 in the normal group to .92 in paranoid patients. Nevertheless, their results clearly indicated that vocabulary is far from an adequate index of earlier ability since vocabulary is far below the level of previous education. They report a coefficient of .16 between chronological age and vocabulary function.

Rabin (1944: 56) criticizes the techniques used above and points out that most studies dealing with the relationship between vocabulary level and general intelligence have used the Binet as a yardstick. This, he says, is not justifiable for adults whose total mental age may drop with age while vocabulary remains constant. To answer the question of what relationship is maintained between vocabulary and general intelligence from adolescence through senility, the Full Scale Wechsler-Bellevue and the 1937 Stanford-Binet Vocabulary test were administered to 868 patients in a state hospital. All but 52 were "without psychosis." The correlation between vocabulary raw score and the Wechsler-Bellevue was .78. Rabin found an increase in vocabulary level with age and concludes that while vocabulary is a fairly good
indicator of an individual's relative position in a group, vocabulary overestimates mental level all along the line. His results show a difference of 10 to 15 IQ points in favor of vocabulary. The difference is not clearly a function of age, although it seems to be smallest in the younger group and largest in the older group.

In another study of the Babcock test, Rabin (1947: 57) analyzed the records of 404 psychiatric patients with an age range of 20 to over 60. He found an obvious decrease in efficiency with age regardless of psychiatric diagnosis and even in groups where deterioration is not considered characteristic. Along with this was a less marked tendency for vocabulary to rise with age. Rabin points out that unless vocabulary norms are established for each age range to make allowances for vocabulary increases, the apparent decline in mental efficiency is a "forced" rather than a true decline. Rabin also noted that while vocabulary increases with age, the mean educational level decreases. Thus, it cannot be said that the older groups show a higher vocabulary level because of better educational opportunities.

Rapaport (1945: 58) conducted an extensive statistical analysis of the performance of 261 subjects on the subtests of the Wechsler-Bellevue. The population was composed of subjects with a wide array of clinical disorders and a relatively normal control group. As part of the method for
determining differences between clinical and control groups, he uses the vocabulary subtest to compute the scatter of other subtests. Rapsaport makes an important contribution to technique in that subtest scatter is computed from the individual's own vocabulary level rather than from the customary group norms. Rapsaport substantiated his own and earlier observations about the nature of vocabulary by evaluating the stability and variability in his own material. He found that vocabulary was one of the three most stable subtests of the Bellevue and, further, that the standard error of the mean of vocabulary was significantly lower than that for all other subtests. He concluded that vocabulary level is relatively refractory to impairment and may be used as a standard from which to estimate impairment of other functions.

**Age and Increasing Vocabulary.** The evidence is conclusive that, as compared to other measures, vocabulary tests are the best indicators of intelligence at all age levels. A number of reviews have indicated slight but insignificant increments in vocabulary level with age in normal populations. Rabin (56), however, suggests that vocabulary norms need to be established for each age range because of the tendency for vocabulary tests to overestimate mental level in relation to intelligence tests.
The evidence of vocabulary stability is based on a variety of tests, both timed and untimed, varying from the definition type found in individual tests to the various multiple-choice types found in group tests. Related verbal tests, e.g., synonyms-antonyms and completion, have also been found to be relatively stable. Since vocabulary scores are likely to be a function of the nature of the test, it is probable that separate age norms are needed for differing tests.

There is evidence in the literature that groups with differing levels of intelligence may vary in the pattern which vocabulary follows with age. For instance, vocabulary may continue to increase to a significant degree in above average intelligence groups.

Somewhat ambiguous results were obtained by Christian & Paterson (1936: 15) who administered a timed paper and pencil vocabulary test to 129 adults who were above average in intelligence in an age range from 40 to 69. A slight but steady decline was noted until the speed factor was removed. This resulted in a definite increase in vocabulary score and the authors conclude that with the speed factor eliminated, vocabulary may increase up to the age of 70.

Sward (1945: 71) gave an elaborate mental test to 45 university professors aged 60 to 80 years and to a comparable
group ages 25 to 35. He found that individual differences far outweighed age differences, but that in word knowledge or vocabulary, the older group were uniformly superior to the younger middle-aged group.

The investigations of Sorensen are of particular interest because of the population studied. In 1933, Sorensen (68) published an analysis of test results of 5,800 college extension students, ages 15 to 74, who were given the Minnesota College Aptitude Test. This examination consists of four sections of vocabulary, each with 180 items. The sample studied was found to be superior to entering freshmen and slightly superior to juniors and seniors at the University of Minnesota. The table of scores reveals a gradual but consistent mean increase from age 15 through 54. The 55-59 group drops slightly, but the 60-64 group achieved the highest scores of the total group. Scores drop beyond this point, but are determined on an extremely small number of cases.

Sorensen (1933: 69) selected 641 of the above subjects, mostly elementary and junior high school teachers, on the basis of years of schooling and occupational status. Up to the age of 55, each five-year interval is represented by about 75 cases. He found that vocabulary ability, with one exception, increased to a rather marked degree. The difference between the 20-24 age group and the 50-54 year old group
in terms of mean score is equivalent to .56 of a standard deviation of the scores of the original group of 5,500 students.

This study was later extended to six other universities (1938: 67). Approximately 9,000 extension students from the late teens to the seventies were given a variety of aptitude tests. In general, total scores showed a slight upward trend with age. Tests of vocabulary or similar to vocabulary tended to show the largest increases and to contribute most to the total score in the older groups.

Increase in vocabulary score with age is not peculiar to above average groups. Weisenburg, Roe & McBride (1936: 810) in their intensive testing of 70 normal hospitalized adults from age 20 to 59 years, found a correlation of .16 between age and Stanford-Binet Vocabulary and a correlation of .33 on the Thorndike Word Knowledge test (multiple-choice). Mean scores for this latter test progress from 53 for the 20-29 group to 73 for the 50-59 group. The population was presumed to be a random and unselected sample.

Studies on vocabulary stability suffer the same limitations as those on intelligence in that the great majority are cross-sectional in nature. The longitudinal studies by Bayley and Oden (8) and Owens (55) have previously been cited. If one considers that the Concept Mastery Test and
particularly the Synonyms and Antonyms section used in the Bayley and Oden research is similar to vocabulary definition tests, then this study gives good evidence that vocabulary is an increasing ability with age -- at least through fifty years in an above average population.

Similarly, the Owens study, in which 127 university men were retested on the Army Alpha after a lapse of 30 years, suggests that vocabulary increases with age. Owens found significantly higher scores on the retest for five of the verbal type tests of which Synonyms-antonyms is the most nearly like a definition-type vocabulary test.

Age and Arithmetic. For the most part, the relationship of arithmetic to age has been incidental to investigations of general intelligence tests of which arithmetic is often a subtest or, as in the case of the Otis, represented by scattered items.

Although the functions measured in tests of number series completion and figure analogies differ from those measured by arithmetic tests, such results as have been found for these types of tasks are of interest here since the ACE Quantitative score is derived from two subtests of this nature as well as from the arithmetic subtest.

Some studies report negative correlations between arithmetic and age. The most general speculation accounting for this relationship is that as distance from formal training
increases, disuse or lack of practice in arithmetic procedures becomes effective. Other studies report little or no change with age, and an occasional investigator reports that arithmetic ability increases with age. But in all these findings, arithmetic tests occupy an intermediate place between the positive position of verbal tests and the negative and decreasing position of performance tests in their contribution to total intelligence score with age.

Miles (1942: 52), in connection with the Stanford Later Maturity Studies, reports that when the Otis 75 item intelligence test, administered without time limit to 400 adults, is analyzed for content, arithmetical problems correlate -.24 and number completion problems -.26 with age.

However, Jones and Conrad (31) in their extensive study of an adult population with the Army Alpha, found no evidence to support the suggestion that arithmetic is especially difficult for adults long out of school. Test 6, Numerical Completions, was among those showing the greatest decline with age, but Test 8, Arithmetic Problems, was found to contribute about 10 per cent to the total Alpha score at each age level. These results are similar to those found by Willoughby (1927: 81) who used the Army Alpha test in a study of family similarities.

In an extensive battery of tests given to Weisenburg, Roe & McBride's normal hospital population of 40 to 80 subjects
in an age range of 20 to 59 (1936: 80), the Stanford Arithmetic Test was included. This is composed of a section of Arithmetic Computation and one of Arithmetic Reasoning, both given without time limits. The correlation of scores with age are -.05 and -.03 respectively. The largest difference between mean scores for the age groups was only three points for either section.

The authors find that the arithmetic test apparently depends most greatly on school training. The Thorndike-McCall Reading Scale and Arithmetic Computation correlate .66, yet paragraph reading and arithmetic computation would hardly be thought to have a great deal in common. The correlation changes to .67 when age is partialed out and to .46 when education is partialed out. Although Arithmetic Computation apparently depends to a considerable degree on school training, it remains on a rather even level between the twenties and fifties. The authors suggest that arithmetic procedures may be well maintained because they are within the natural realm of experience for the older person, although the possibility that they involve abilities which mature late or decline slowly must not be overlooked (p. 107).

Wechsler (1944: 77), in connection with the individually administered Arithmetic subtest of the Bellevue scale, reports that the correlations between this test and the total score tend to hold an intermediate position between correlations
found for the verbal and performance subtests with total scale scores. However, the correlations between arithmetic and total score tend to vary with age and tend to be higher at the upper than the lower ages. For the age group 20-34, the correlation between Arithmetic and total scale score is .65, for the age group 35-49, .67. Although the test versus total scale correlations rise with age, the absolute scores on the test decrease significantly with age. Here again, Arithmetic occupies an intermediate place. It does not fall off as quickly as most performance tests nor hold up as well as the verbal tests. In computing Efficiency Quotients, Arithmetic is designated as one of the "not hold" tests.

Morris (1940: 53) compared the performance of a group of employed adults on a series of achievement tests with the highest grade that they had reached in school. While gains were found in some of the information and vocabulary tests, losses were found in arithmetic skills. However, clerical and technical workers lost less of the arithmetic than other categories of employed men. Thus it appears that both remoteness from schooling and use of learned material on the job affects achievement scores.

In his study of university professors, Sward (1940: 71) found that 80 per cent of the younger group, ages 25 to 35, reached or exceeded the median of the older group, ages 60 to 80 years, on the numerical tests (Ingenuity, Arithmetic
and Number Series). Nearly two-thirds of the older men fall within the lower quartile of the younger group. Sward suggests that probably no single factor explains this apparent decline in arithmetical proficiency in the older years but that the factor of disuse can hardly be ignored. On power tests it was found that the older men require 10 minutes longer on the average to complete the Number Series and Arithmetic tests. Sward finds a correlation of .47 between the age of the older and time devoted to solving problems, while the correlation between age and scores for this group is -.12.

He states that

what the discrepancy between the size of the two coefficients means, apparently, is that one's failure to brush up on arithmetic operations between the ages of 60 and 80 does less to impair accuracy than to interfere with one's speed or dexterity in tasks of this character (p. 470).

An investigation which has bearing on the present problem was that of Sorenson (1938: 67) who studied the test performance of extension students in a number of universities. In two of these, represented by 233 and 242 students in an age range from 15 to over 50, an early form of the ACE was the source of data. Subtest scores were analyzed in relation to the total ACE score, consisting of both verbal and non-verbal material. There was inconsistency in the results of the two youngest groups between the two universities. Arithmetic ability was relatively lower than any other in the youngest group in one university, but next to the highest in the
youngest group in the other university. However, after a point in the twenties, the curves became similar for both groups with a tendency for arithmetic ability to increase with age. The trend was found to be essentially the same as for the verbal completion and opposites tests. Beginning with the late twenties, then, arithmetic contributes a somewhat disproportionate amount to the total score as compared with certain other subtests.

The figure analogies subtest, similar to that of the ACE Quantitative section used in the present study, showed a very marked and consistent decrement with age in both university groups. In the younger groups it contributed a relatively greater per cent to the total score than any other test. At a point in the late twenties, the curve for figure analogies crosses the curve for the total score and from that point on a relatively smaller and smaller contribution to total score is found. Sorensen accounts for these differential trends in terms of practice and experience.

Adults are usually accustomed to think in verbal terms, seldom in terms of geometrical forms. Thus, while the ability to think verbally increases, non-verbal thinking ability decreases with age. Development continues up to the age of mental maturity; then it declines because there is relatively little activity to maintain it (p. 180).

The recent investigation of Corsini and Fassett (1953: 17) of a "forced sample" of 1,072 adults in which all age ranges from 20 to over 60 were represented indicated that, in
contrast to Wechsler’s results, Arithmetic rises significantly with age.

The longitudinal investigation of Owens (1953: 55), in which 127 university men were retested on the Army Alpha after a lapse of 30 years, reveals no significant changes with age on the subtests Arithmetic Problems and Number Series Completion although verbal tests increase significantly on the retest.

It is apparent that some of the differences in results found by various investigators is partly a function of test content, administrative procedures (individual or group, timed or untimed) and the population under consideration. Here again it should be noted that variability of performance tends to increase with age and that individual differences are greater throughout the age ranges studied than are the differences between age groups.

**Age and Speed.** It is not a purpose here to review investigations of physiological functions as they relate to age. There is general unanimity that these decline, but with variations in the ages at which specific functions reach their peak, plateau and gradually begin a downward descent. Investigations or reviews of this area may be found in Ruger (60), Bellis (9), W. R. Miles (48, 50, 51), Wechsler (79) and others. A common finding in this area, as in the measurement of mental abilities,
is that variability or individual differences within age groups exceeds differences between age groups.

The parallelism in curves of growth and decline for measures of physiological and psychological functions has resulted in two approaches to the problem of whether time-limit or amount-limit tests are appropriate for measuring adult performance. Some investigators take the view that declining physiological variables introduce more or less extraneous factors which are incidental and not necessarily parallel to decrements in mental abilities. Others find in this parallel a universal decline affecting psychological as well as physiological aspects of the human being which can not or should not be separated in measures of adult efficiency.

In the majority of studies, "speed" is ill-defined. The practice of investigating "speed" by comparing timed test performance with power performance does little to differentiate what is actually taking place between the time an individual receives a stimulus and gives a response. Physiological variables related to intake of a stimulus, such as vision and hearing, and those related to output, such as motor accuracy and speed, do not account for the speed or slowness of the mental reaction itself. In this in between area, such variables as motivation, tendencies to precision or impulsiveness, practice effects, fluctuations of attention, fatigue and many other variables play a part in determining
the end result. As McFarland (43) points out, speed is not the same as time in the mental reaction. Time is speed plus the interaction of many variables. Thus, in the following reviews it is obvious that "time" rather than "speed" more correctly defines the area of investigation. However, both terms will be used synonymously unless otherwise defined.

The investigations of W. R. Miles (47, 51), C. C. Miles (45) and Copeland (16) are similar in the respect that their studies of speed and power in relation to age are based on large population samples and derive their data from the Otis test of intelligence. Similarly, their results indicate that when the Otis is administered on an untimed basis the mean score performance of adults does not decline as significantly and as rapidly as when the timed test scores are considered. C. C. Miles (45) reports that decline is not significant before the fifth decade. In terms of standard score units, the decline is .76 for the timed test between ages 20 and 40 and .46 for the untimed. But from 60 to 80, it is .37 for the timed test and .98 for the untimed. Thus, in earlier years, speed appears to decline faster than power, whereas in late maturity the decline of power is more apparent. W. R. Miles (51) reports that whereas speed and age coefficients fall between -.3 and -.5 depending on age range and type of test, power tests and age yield coefficients between -.1 and -.3. Copeland (16), on the basis of time required to complete tests,
found that speed declines in a linear fashion. Younger adults take less time to complete the task.

Jones and Conrad (31) found that when the speed handicap in their Army Alpha study was removed by considering only the first two-thirds of the responses, there was no change in the developmental curve. These results are similar to those of the authors of the original Army Alpha (83) who reported that a doubling of time limits did not materially affect the original rankings. Jones and Conrad conclude that the Alpha is not essentially a speed test and also find that adults do not work more accurately than adolescents though answering fewer items. Their analysis of "items attempted" and "items correct" indicates that the cause of declining ability lies in the intellectual difficulty of the task rather than in speed factors or motivation. When Gilbert (28) omitted the timing factor in her study of the Babcock scale, she also found that decline occurred but was less in degree.

In his investigation of young and old groups of university professors, Sward (71) found that while the older subjects were, as a group, slower on all tasks, the differences were negligible on the shorter verbal tasks, but significantly longer than for the young on the numerical tasks. Sward suggests that the lengthened times of the old may indicate that tenacity and blocking played a larger role in this age group.
Brown (13) compared age with speed and power on a non-language analogies test similar to that of the ACE. Using two forms of the test he found that the speed-power correlation for younger subjects (below 50) was .79. This dropped to .71 for a group aged 31 to 50 and to .65 for the above 50 subjects. Speed and age for the total group correlated -.16 and power and age -.07.

To determine the age relationship between speed and power tests of mental ability, Lorge (40, 43) gave a battery of tests to 143 adults in the age ranges 20-25, 27½-37½ and 40-70. In an earlier section the coefficients with age between the timed Otis and Army Alpha and the untimed CAVD have already been cited as -.48, -.36 and -.27 respectively. Since correlations between the tests run as high as .85, Lorge assumes that the smaller coefficient between power and age is accounted for by speed factors being removed. When the older and younger subjects were equated for CAVD performance, the older group still showed significantly lower scores on the speed tests. On the basis of these results, he applied a correction for speed to the data of the Miles' and Jones' studies and found that age differences tended to vanish. Lorge concluded that "the inference of mental decline with age is an unfortunate libel on adults" (42, p. 110). He criticizes present test devices for the fact that they measure undifferentiated mixtures of speed and power, and he
cites the evidence of physiological decline as reason to believe that speed tests underestimate intellectual power in the adult.

Brown and Lorge would agree with Eysenck (19) that it is more practical to give unspeeded tests to older subjects since speed is not of paramount importance in the professional and ordinary life of the individual where most required tasks are of a familiar nature.

Wechsler (79) disagrees with the conclusion of Lorge, both because of the negative CAVD and age correlation and because he doubts that speed can be abstracted from achievement without doing violence to the concept of intellectual virility. He points out that Lorge did not show that older subjects did as well as younger subjects on the more difficult items of the test.

Eysenck (19), however, remarks that while no separate speed factor in intelligence has been found in children, adolescents and young adults, she has, in her own factor analyses, been unable to conclude that no speed factor is active in later years. Eysenck, along with others, does not consider that the problem of age versus speed and power has been solved and, by implication, argues for a longitudinal approach.

To reduce and control the number of variables involved in test performance, Birren, Allen and Landau (1954: 10) made
an analysis of the task of simple addition in relation to the length of the digits added, time, probability of success and age of the subjects. Data were reported on 413 subjects in the age range 16 to 90 years inclusive. The subjects were required to add rapidly single columns of digits ranging in length from 2 to 25 digits. The elderly subjects were slower for all lengths of problems and showed less accuracy with increased length of problem. The time required to complete long problems did not increase as rapidly as for the younger group, but accuracy declined more rapidly. The authors suggest that older people require more time to perceive digits, to add and to write down responses apart from their loss in accuracy. They propose that speed in the old is functionally related to mental performance.

Summary and Discussion. From the review of the literature it is evident that the relationships found between test scores and age are a function of test content, of the population studied, the age level and range, of whether tests are of the individual or group type, and whether they are given on a timed or untimed basis. It is also evident that in certain types of test performance individual differences within a given age range tend to be greater than differences between age groups.

Intelligence tests, since they are composed of a variety of tasks, yield the most conflicting results. Most investigators
find that intellectual abilities reach a peak in the early twenties and after a period of relative stability gradually decline. Verbal abilities tend to decline less rapidly, if at all, than non-verbal skills. These conclusions have been derived from cross-sectional studies which, as Lorge and Ruch (36) point out, are based on "the implicit assumption of homogeneity of background of subjects (which) is undoubtedly false."

Recent longitudinal studies, with some support from cross-sectional investigations, indicate that intellectual capacities may continue to increase at least to the age of fifty for populations with above average intelligence.

These conflicting results have at least two implications. One possibility is that the longitudinal type study, which does not have the sampling problems of the cross-sectional investigation, gives a more accurate picture of the progress of mental functions with age.

A second possibility is that as cultural demands and opportunities increase or broaden, man is more likely to continue growth and development to meet changing self and social needs. From this viewpoint, it is incumbent on the investigator to establish norms that are compatible with the culture and to keep test content current with the era. Bayley (7), Cattell (14), Tuddenham (75) and others, would concur with this view.

The fact that some older individuals do not demonstrate decrements in ability which follow the course of expectancy of
group averages belies the conclusion that decline is a function of age alone. Since non-verbal skills have been shown to decline more rapidly with age than verbal skills, it is important to note that some investigators find that when life situations and interests center in areas of non-verbal activities, such functions continue to be relatively well maintained. Current intelligence tests may measure the potentialities of the young to progress in many directions. These same intelligence tests, applied to adults, may measure the degree to which these potentialities have been realized. The effects of use and disuse can hardly be denied.

The fact that non-verbal tests such as figure analogies show significant mean decreases with age, suggests, as Sorensen (67, p. 180) has already pointed out, that thinking in terms of geometrical forms "continues to the age of mental maturity; then it declines because there is relatively little activity to maintain it."

There is unanimity among investigators that subtests of intelligence tests make differing contributions to total scores at different age levels. Some investigators, like Shock (64), and Balinsky (6), who conducted a factor analysis of test results for Wechsler's subjects, take the viewpoint that mental traits change and undergo reorganization over the years. Balinsky has shown that the subtests of an intelligence scale do not measure the same factors from age to age. This
obviously has implications for interpretation of test results. Lorge & Rush (36) speculate that in the measurement of adult intelligence, the various tests are functionally impure. They suggest that the process of arriving at a response may differ at different ages.

At younger ages, the response may involve abstraction and generalization to arrive at the answer; in mature groups, the answer may be a learned response; and in aged groups the response may involve return to earlier methods of responding.

A few already cited investigators have studied incidentally the influence of years of education on adult intelligence. Studies of Sorenson (68), Weisenburg, Roe & McRide (80) and others have, through diverse approaches, arrived at the conclusion that it is not the number of years that one has been to school but rather intellectual capacity as measured by tests that is the dominant factor in adult achievement. However, Brody (11) demonstrates that their evidence can be treated in such a way as to indicate that schooling is at least of equal importance. This leaves the problem open-ended.

A number of investigators have used forms of the ACE to demonstrate increasing scholastic ability as a function of grade status. For instance, Silvey (1951: 66) recently found a highly significant mean raw score gain of 18.76 points on the total ACE scores for 517 students from the freshman to the sophomore year. Bailey & Brammer (1951: 5)
found in a review of literature that five out of seven studies reported significant gains on the ACE with increases in grade status and two reported gains of marginal significance. Their own investigation of score changes with increases in grade status from junior college freshman status to upper-division status at Sacramento State College also yielded significant ACE gains. It is obvious from the review of the literature that intellectual growth may continue to the middle twenties and that intelligence score increments with age may be somewhat higher in above average populations. Before formal education can be demonstrated to be the dominant factor in such changes, it is necessary to know what changes would occur on the same test at the same age and intelligence levels without formal education. What is needed is a longitudinal study of individuals who at a given age in early youth are comparable in intelligence, but who in later years differ in respect to levels of formal education. Large (41), for instance, studied the intelligence test scores of individuals at age 34 as a function of their intelligence test status at age 14 and the number of years of schooling after the eighth grade. He found that the groups with the greater number of years of schooling completed, though matched for IQ score at age 14, made larger intelligence test scores than did their 14 year peers.
A more conclusive, though qualified, statement can be made about the status of vocabulary tests. It has been demonstrated that vocabulary tests provide the most stable single index of the highest level of intellectual functioning attained by an individual. However, there is also evidence that vocabulary may continue to increase with age independently of other subtests of mental ability. Since vocabulary tests may overestimate intelligence in the middle and older adult years, separate norms for each age level are advisable.

Before leaving this subject, the fact that the stability of vocabulary has been challenged deserves attention. Yacorszynski (1941: 82) questions the Babcock method and suggests that the stable character of vocabulary is an illusion based on the fact that different levels of definitions of the words are scored as acceptable. He emphasizes that qualitative analysis of vocabulary responses may yield different results. Feifel (20), Rapaport (58) and Fox (24) have undertaken investigations in this area, but evidence is still limited.

The relationship of arithmetic to age depends on the nature of the test and testing procedures. In general it may be said that arithmetic scores tend to occupy an intermediate position between verbal tests on the one hand and non-verbal tests on the other. Several authors emphasize
that the relationship of arithmetic with age may depend on
the factors of use and disuse.

Such tests as number series completion and figure
analogies appear to be negatively related to age. Their ap-
propriateness for measuring adult intelligence is questioned
insofar as the kinds of functions which they measure may not
have their counterpart in life situations.

The problem of the relation of speed to adult efficiency
has not been solved. While speed has been found to be highly
correlated with intelligence in children, adolescents and
young adults, there is no conclusive evidence that speed bears
the same relationship to adult ability. Mental speed has not
been sufficiently differentiated from physiological variables
related to intake of stimuli and output of response. Present
timed tests measure undifferentiated amounts of what may be
clearly defined as non-intellectual physical variables which
are known to decline with age.

Untimed tests of intelligence yield scores that favor
a longer period of intellectual stability and less decline in
the older years than do speeded tests. Until such time as
the relationship of physiological to psychological decline
has been established, the advice of Brown (15), Lorge (42)
and Eysenek (19) that unspeeded tests for adults are justi-
fiable is well taken. Since the use of timed tests has many
practical features for the evaluation of large groups, the limitations of the procedure may be partially overcome by selecting tests which are appropriate in content for the adult population to be measured.
CHAPTER III

THE PRELIMINARY STUDY

In discussing the origin of this study, the observation was made that older credential candidates tend to make higher scores on the Vocabulary subtest of the Cooperative English Examination (CEE) than younger students and that older students often have a low Quantitative score on the American Council on Education Psychological Examination (ACE) in relation to individual test performance and to the group as a whole. Consequently it was decided to conduct a preliminary study of the accumulated test data on approximately 600 credential candidates who were tested in the Spring, Summer Session and Fall semesters of 1954, the first year in which there was a decided increase in the number of older individuals taking the screening tests. The implementation of the new state law concerning provisional teachers was responsible for the change in population.

Description of Tests. The three subtests of the ACE, CEE battery selected for scrutiny were the CEE Vocabulary subtest, the ACE Linguistic scale and the ACE Quantitative scale.

The Vocabulary subtest consists of 60 five-choice items administered with a time limit of 15 minutes. It is a recognition type task with gradually increasing work difficulty,
and items requiring increasingly fine discrimination.

The Linguistic score is derived from three subtests, each with a time limit of five minutes. The Completion subtest is composed of 30 five-choice items; the Same-Opposite subtest has 50 five-choice items and the Verbal Analogies subtest has 40 five-choice items.

The Quantitative score is also derived from three subtests, all with five-choice items. The Arithmetic subtest is composed of 20 verbally stated problems administered with a ten minute time limit and is the first test of the ACE, CEE battery. The Figure Analogies subtest has 30 items with a time limit of five minutes. The Number Series subtest has 30 items with a time limit of eight minutes.

Rationale for Test Selection. The three parts of the total screening battery were selected for study to discover to what extent the previously noted observations are based on fact.

The Vocabulary subtest of the CEE was selected for study not only because of the observation that it tends to be high relative to other subjects in older subjects, but also because the review of the literature has established that of the functions considered vocabulary is the least vulnerable to reduced efficiency with age. Maturation, cultural level, motivation, bilingual handicaps and other factors doubtlessly
influence individual vocabulary level so that it cannot be used as a final criterion of intelligence. However, for the purpose of group study, vocabulary may be used as a base for comparison with other test scores.

The CKE Vocabulary subtest is little influenced by timing. This inference is the result of several years of observation and, as will be noted later, is confirmed by an analysis of the experimental group. Far more students are able to complete this test in the time limit than are able to complete the subtests of the ACE Linguistic and Quantitative scales. The test authors confirm the impression that this is a relatively power-type task in the following statement from their test manual (18):

The Vocabulary Score indicates the extensiveness of the individual’s word knowledge. The time limit for this section is long enough so that, except for a few individuals whose mechanics of reading are extraordinarily poor, speed of word recognition plays little part in determining the Vocabulary Score.

The ACE Linguistic section was included in the analysis since the subtests of which it is composed are rigorously timed and speed variables related to age may be reflected in performance. The test authors (1) state that the Linguistic score is a measure of scholastic ability and, therefore, this test is likely to relate closely to tests of verbal intelligence. One would expect the Linguistic scale to show a high correlation with the CKE Vocabulary unless extraneous variables are having a detrimental effect.
The ACE Quantitative section is also derived from three rigorously timed subtests and the assumption may again be made that variables related to speed will affect performance. Since learned arithmetic functions tend to be used less than verbal functions, an additional possibility is that there may be age variations in speed of recall. A third consideration is that the Figure Analogies and Number Series subtests may measure ability to learn and apply new concepts under pressure rather than facility in using already acquired skills in new situations. The review of the literature has established that many investigators question the practice of judging adult efficiency in the same manner as for children or young adults. The ACE Quantitative test was selected for scrutiny to discover whether there are discernible differences in performance with increasing age.

The Population. The population studied included all students working toward public school credentials: teacher, administrative, supervisory, and pupil personnel. Teacher credential candidates contribute by far the largest proportion of cases to the total group.

The age range of the approximately 600 candidates who took the screening battery in 1954 was 19 through 64 years. This range was divided into intervals of five years, except that the youngest group of 19 and less is represented by only 18 and 19 year old students. These intervals represent the
age at the time the tests were taken regardless of accumulated
months. The 20-24 age range has by far the largest representa-
tion, 188 and 189, followed by 120 and 125 in the 25-29 in-
terval. The number of cases in each age range gradually di-
minishes until in the 55-59 group there are only eleven cases,
and in the 60-64 group only four and five cases. The slight
difference in \( N \) for age groups and separate subtest popula-
tions are accounted for by the fact that all test scores were
tallied whether or not the individuals had completed the total
battery. The mean age levels on the Vocabulary, Linguistic
and Quantitative sections are 50.12, 50.14, and 50.16 respec-
tively. The median age levels are 27.40, 27.03 and 26.40.
The slight differences between median and mean age levels re-
fects the larger number of cases in the younger age groups.

The Method of Procedure. Raw score data were used in
the Quantitative and Linguistic comparisons, but for Vocabu-
lary, the scaled scores offered by the test authors for nor-
malizing the score distributions were utilised.

Means, medians and standard deviations (S.D.) were
computed for total test score and age group distributions.
Pearsonian product-moment correlation coefficients and cor-
relation ratios (Eta) were computed between age and each of
the tests selected for study. In addition, an analysis of
variance was made for each subtest on age in order to test
the hypothesis that the means of the test scores of separate
age groups do not differ from zero except by chance. Fisher's z distribution or its non-logarithmic version, Snedecor's F distribution was the model for the test of significance of the data.  

Results of the Preliminary Study

**Vocabulary and Age.** The correlation coefficient, correlation ratio and analysis of variance for Vocabulary and age are given in Table I. The hypothesis being tested is that the means of Vocabulary scores on age groups do not differ from zero except by chance. The correlation coefficient between Vocabulary and age is .57. To test the significance of the linear correlation, the variance estimates (mean squares) for the linear regression and residuals from the line were computed at 498.63 and 5.44 respectively, resulting in an F-ratio of 91.66. For 1 and 597 degrees of freedom (df), this relationship would occur by chance less than one time in one thousand ($P < .001$), and the null hypothesis is rejected. It may be assumed that a significant positive relationship exists between Vocabulary test scores and increases in age.

The review of the literature (Chapter II) has established that the central tendency of some psychological measurements of adults are described by a parabolic curve when

---

2All formulae and methods used in the statistical treatment of the data are according to McNemar, Quinm, *Psychological Statistics* (New York: John Wiley & Sons, Inc., 1949).
TABLE I

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Rta) and Analysis of Variance for the Regression of CEE Vocabulary Scaled Scores, Form Z, on Age for the Preliminary Study (N = 599)

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>498.63</td>
<td>1</td>
<td>498.63</td>
<td>91.66</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>137.17</td>
<td>8</td>
<td>17.17</td>
<td>3.28</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Between-array means</td>
<td>636.08</td>
<td>9</td>
<td>70.67</td>
<td>13.38</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Within arrays</td>
<td>3107.45</td>
<td>589</td>
<td>5.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>3244.88</td>
<td>597</td>
<td>5.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5743.46</td>
<td>598</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the age range is sufficiently wide, hence, the assumption of linearity of regression is not always tenable. The correlation ratio (Eta), a test for the degree to which within-array means fail to fall on a straight line, was computed at .41. To test the significance of curvilinearity, variance estimates for between-array and within-array means were calculated at 70.67 and 5.28 respectively, resulting in an F-ratio of 13.58. For 9 and 589 df a ratio of this magnitude would be observed less than one time in one thousand by chance, so the null hypothesis is rejected. For any given sample, the within-array means will tend to depart in some degree from a straight line fit. This results in the observation that the correlation ratio is always equivalent to or larger than the correlation coefficient. In order to test the hypothesis that the within-array means of Vocabulary on age do not deviate from a straight line except by chance, the test for linearity of regression was applied. Mean squares of 17.17 and 5.28 for the deviation of means from a line and for within-arrays gave an F-ratio of 3.25. For 8 and 589 degrees of freedom, a ratio of this magnitude would occur less than one time in one hundred by chance, or, in other words, is significant at higher than the .01 level of confidence. Thus the departure from linearity of the array means is sufficiently great to lead to a rejection of the hypothesis of linearity and the relationship can be described best by a fitted curve.
A comparison of medians, means and standard deviations for the Vocabulary subtest scaled scores with age are given in Table II. Inspection of the mean performance of each age group reveals an increment in vocabulary level from the youngest through the 40-44 age interval. This is an increase in mean score from 57.64 for the 19 and less age group to 72.41 for the 40-44 group with a substantial number of cases in each interval. Mean scores for the 45-49, 50-54 and 55-59 age groups drop slightly but remain above the mean of 65.18 obtained by the 25-29 group. The last and oldest group, 60-64, surpasses all other groups with a mean performance of 81.00, a difference of over 23 points, almost 2 standard deviations from the youngest group. The mean test score for the total group is 65.18, equivalent to the mean for the 25-29 age group. The older age level is represented by only five cases and since the problem of selection arises, the position of the 60-64 age group in the later comparisons is of high interest. The median Vocabulary test scaled scores approximate the means at all age levels.

Just as the mean performance ascends with increasing age, so also does the variability from the youngest through the 45-49 age group, a different between S.D.s of 8.64 and 14.98. The remaining three older groups are more stable than the total group variance, S.D. 12.50, with the 60-64 group being the most consistent as shown by a standard deviation of 6.53.
<table>
<thead>
<tr>
<th>Age</th>
<th>N</th>
<th>Median</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 - 64.99</td>
<td>5</td>
<td>80.75</td>
<td>81.00</td>
<td>6.65</td>
</tr>
<tr>
<td>55 - 59.99</td>
<td>11</td>
<td>69.65</td>
<td>69.96</td>
<td>11.85</td>
</tr>
<tr>
<td>50 - 54.99</td>
<td>14</td>
<td>65.65</td>
<td>67.00</td>
<td>8.02</td>
</tr>
<tr>
<td>45 - 49.99</td>
<td>40</td>
<td>72.00</td>
<td>71.63</td>
<td>14.98</td>
</tr>
<tr>
<td>40 - 44.99</td>
<td>49</td>
<td>71.65</td>
<td>72.41</td>
<td>12.90</td>
</tr>
<tr>
<td>35 - 39.99</td>
<td>58</td>
<td>70.10</td>
<td>69.33</td>
<td>12.67</td>
</tr>
<tr>
<td>30 - 34.99</td>
<td>70</td>
<td>67.95</td>
<td>69.36</td>
<td>11.87</td>
</tr>
<tr>
<td>25 - 29.99</td>
<td>125</td>
<td>65.85</td>
<td>65.16</td>
<td>11.65</td>
</tr>
<tr>
<td>20 - 24.99</td>
<td>188</td>
<td>57.85</td>
<td>59.71</td>
<td>9.98</td>
</tr>
<tr>
<td>&lt; 19.99</td>
<td>39</td>
<td>56.05</td>
<td>57.64</td>
<td>8.64</td>
</tr>
<tr>
<td><strong>Total Test</strong></td>
<td><strong>599</strong></td>
<td><strong>65.15</strong></td>
<td><strong>12.50</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Total Age</strong></td>
<td><strong>27.40</strong></td>
<td><strong>30.32</strong></td>
<td><strong>10.12</strong></td>
<td></td>
</tr>
</tbody>
</table>
**Linguistic Test and Age.** Correlations and an analysis of variance for the regression of Linguistic scores on age are contained in Table III. The linear coefficient is -.07. An analysis of variance in which the mean squares for the linear regression and residuals from the line were calculated at 54.87 and 17.02 respectively, yielded an F-ratio of 3.22. For 1 and 583 df, this relationship would occur more than five times in one-hundred by chance (F \( > .05 \)) and the null hypothesis of no difference in mean performance except by chance can be tentatively accepted. Linguistic scores and age do not have a significant linear relationship.

A correlation ratio of .27 between the two variables was analyzed for significance. Mean squares for between and within-array means were calculated at 80.92 and 16.09, resulting in an F-ratio of 5.03. For 9 and 575 degrees of freedom, a ratio of this magnitude would occur by chance less than one time in one thousand. Thus, the null hypothesis can be rejected. The data reveal a significant curvilinear relationship between age and Linguistic scores.

The conclusion that a curve will better describe the relationship than a straight line is confirmed by the F test for linearity of regression. Variance estimates of 84.06 and 16.09 for the deviation of means from a line and for within-arrays respectively yield an F-ratio of 5.22 which for 8 and 575 df indicates that the obtained departure of the array
### TABLE III

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of ACE Linguistic Raw Scores, 1958 Edition, on Age for the Preliminary Study ($N = 585$)

<table>
<thead>
<tr>
<th></th>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistic and Age</td>
<td>-.07</td>
<td>.27</td>
</tr>
</tbody>
</table>

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>54.87</td>
<td>1</td>
<td>54.87</td>
<td>3.22</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>672.44</td>
<td>8</td>
<td>84.06</td>
<td>5.22</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Between-array means</td>
<td>726.31</td>
<td>9</td>
<td>80.92</td>
<td>5.03</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Within arrays</td>
<td>9249.55</td>
<td>575</td>
<td>16.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>9321.99</td>
<td>583</td>
<td>17.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>9976.86</strong></td>
<td><strong>584</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
means from a straight line would occur by chance less than once in one thousand times.

Since the correlation ratio is a general measure which does not describe the form of the relationship between two variables, this is best determined by a description of mean score performance at various age levels. The means, medians and standard deviations for each group are given in Table IV.

Inspection of the table reveals an increment in Linguistic mean score from 71.61 for the youngest age group of 19 and less to a peak of 81.20 in the 30-34 age interval. The mean drops to 74.64 for the 35-39 group and continues to decrease slightly until the 45-49 range which has the same mean level as the youngest group, 71.61. The remaining three older age groups drop abruptly to a low mean score of 59.72 for the 55-59 age interval, the other two groups scoring but little higher. This is a difference of more than one population S.D. of 16.91 from the high scoring 30-34 group. Of all groups, the 35-39 and 40-44 age groups most nearly approximate the population mean of 74.36.

In sum, there are definite increases in mean Linguistic test scores from the youngest to the 30-35 age level. Beyond this point there is a definite decrement and a sharp differentiation between the three older groups and the remainder of the population. As in the case of Vocabulary test scores, the median Linguistic scores closely approximate the means at all age levels.
### TABLE IV

Table of Medians, Means and Standard Deviations for Age and ACE Linguistic Raw Scores, 1982 Edition, in the Preliminary Study

<table>
<thead>
<tr>
<th>Age</th>
<th>N</th>
<th>Median</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 - 64.99</td>
<td>4</td>
<td>62.00</td>
<td>60.78</td>
<td>20.73</td>
</tr>
<tr>
<td>55 - 59.99</td>
<td>11</td>
<td>60.75</td>
<td>59.72</td>
<td>11.34</td>
</tr>
<tr>
<td>50 - 54.99</td>
<td>14</td>
<td>62.00</td>
<td>63.79</td>
<td>11.59</td>
</tr>
<tr>
<td>45 - 49.99</td>
<td>38</td>
<td>73.50</td>
<td>71.61</td>
<td>15.79</td>
</tr>
<tr>
<td>40 - 44.99</td>
<td>44</td>
<td>73.50</td>
<td>73.59</td>
<td>15.80</td>
</tr>
<tr>
<td>35 - 39.99</td>
<td>58</td>
<td>73.60</td>
<td>74.84</td>
<td>17.59</td>
</tr>
<tr>
<td>30 - 34.99</td>
<td>69</td>
<td>80.20</td>
<td>81.20</td>
<td>14.03</td>
</tr>
<tr>
<td>25 - 29.99</td>
<td>120</td>
<td>80.95</td>
<td>78.71</td>
<td>16.48</td>
</tr>
<tr>
<td>20 - 24.99</td>
<td>188</td>
<td>72.45</td>
<td>72.24</td>
<td>15.80</td>
</tr>
<tr>
<td>&lt; 19.99</td>
<td>39</td>
<td>70.00</td>
<td>71.61</td>
<td>15.82</td>
</tr>
<tr>
<td>Total Test</td>
<td>585</td>
<td></td>
<td>74.58</td>
<td>16.91</td>
</tr>
<tr>
<td>Total Age</td>
<td>57.03</td>
<td>30.14</td>
<td>10.04</td>
<td></td>
</tr>
</tbody>
</table>
Linguistic scores demonstrate a positive relationship with Vocabulary for the 19 and less through the 30-34 age groups. Beyond this level there is a negative relationship. The comparison is particularly striking for the older, 60-64, age group who, after exceeding the youngest group in Vocabulary score by nearly 2 population S.D.s, drop nearly one S.D. (18.91) below the mean performance of the total group on the Linguistic test. On the Vocabulary test the older subjects presented the most stability of performance, while on the Linguistic section they are the most unstable (S.D.=20.73).

Quantitative Test and Age. Table V presents the correlations and the analysis of variance table for the regression of Quantitative test scores on age. The linear correlation is -.40 and the correlation ratio is .46.

The significance of the correlation coefficient was tested by the F-ratio. Variance estimates for linear regression and residuals from the line were 777.03 and 6.99 respectively, resulting in an F-ratio of 111.16 which for 1 and 584 degrees of freedom is highly significant and would occur by chance less than one time in one thousand. The null hypothesis of no relationship can be rejected and a linear relationship assumed.

However, the correlation ratio is also highly significant at greater than the .001 level of confidence. The variance estimates for between and within-array means were 105.29
TABLE V

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of ACE Quantitative Raw Scores, 1959 Edition, on Age for the Preliminary Study (N = 888)

<table>
<thead>
<tr>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative and Age</td>
<td>-.40</td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>777.06</td>
<td>1</td>
<td>777.06</td>
<td>111.16</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>170.57</td>
<td>8</td>
<td>21.32</td>
<td>3.18</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Between-array means</td>
<td>947.59</td>
<td>9</td>
<td>105.39</td>
<td>15.51</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Within arrays</td>
<td>3911.86</td>
<td>576</td>
<td>6.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>4082.43</td>
<td>584</td>
<td>6.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4889.46</td>
<td>585</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and 6.79, yielding an F-ratio of 15.51. For 9 and 576 degrees of freedom, a ratio of this magnitude would also occur less than one time in one thousand by chance.

To test for linearity of regression, mean squares of 21.32 and 6.75 were found for the deviation of means from a line and the within-array means respectively. The resulting F-ratio of 3.16, for 8 and 576 df, would occur by chance less than one time in one hundred. With this level of confidence, the null hypothesis may be rejected and a curve may be assumed better to describe the relationship between age and Quantitative scores.

Table VI presents the medians, means and standard deviations of Quantitative scores on age. Inspection of the data reveals a pronounced decrement in mean score with increasing age beyond the 30–34 age level. The age groups from the youngest through 30–34 appear to be on a plateau with mean scores between 48.31 and 44.40. The 35–39 subjects drop sharply to a mean score of 36.83 and the downward slope is continuous from 37.88 in the 40–44 interval to a low of 20.50 in the group of oldest subjects. The population mean of 40.23 falls between the 30–34 and 35–39 intervals. As in both preceding comparisons, the median Quantitative scores closely approximate the means at all age levels. The change in score progression from plateau to negative regression from the youngest through the oldest group may account for the greater significance of the correlation ratio.
# TABLE VI

Table for Medians, Means and Standard Deviations for Age and ACE Quantitative Raw Scores, 1968 Edition, in the Preliminary Study

<table>
<thead>
<tr>
<th>Age</th>
<th>N</th>
<th>Median</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 - 64.99</td>
<td>4</td>
<td>17.50</td>
<td>20.50</td>
<td>6.54</td>
</tr>
<tr>
<td>55 - 59.99</td>
<td>11</td>
<td>22.19</td>
<td>23.86</td>
<td>8.44</td>
</tr>
<tr>
<td>50 - 54.99</td>
<td>14</td>
<td>25.50</td>
<td>26.07</td>
<td>9.02</td>
</tr>
<tr>
<td>45 - 49.99</td>
<td>36</td>
<td>32.70</td>
<td>37.56</td>
<td>11.19</td>
</tr>
<tr>
<td>40 - 44.99</td>
<td>46</td>
<td>36.70</td>
<td>34.46</td>
<td>11.61</td>
</tr>
<tr>
<td>35 - 39.99</td>
<td>58</td>
<td>37.50</td>
<td>36.53</td>
<td>10.85</td>
</tr>
<tr>
<td>30 - 34.99</td>
<td>67</td>
<td>42.02</td>
<td>42.31</td>
<td>8.77</td>
</tr>
<tr>
<td>25 - 29.99</td>
<td>120</td>
<td>44.94</td>
<td>44.40</td>
<td>10.94</td>
</tr>
<tr>
<td>20 - 24.99</td>
<td>189</td>
<td>42.70</td>
<td>42.58</td>
<td>10.30</td>
</tr>
<tr>
<td>&lt; 19.99</td>
<td>39</td>
<td>43.70</td>
<td>44.00</td>
<td>8.70</td>
</tr>
<tr>
<td>Total Test</td>
<td>586</td>
<td>40.23</td>
<td></td>
<td>11.52</td>
</tr>
<tr>
<td>Total Age</td>
<td>26.66</td>
<td>30.16</td>
<td></td>
<td>10.08</td>
</tr>
</tbody>
</table>
The 60-64 subjects show an even more striking contrast in performance between Vocabulary and Quantitative scores than between Vocabulary and Linguistic scores. The mean Quantitative score for this older group is less than half the mean score of the youngest group and about one-half the population mean of 40.23. This is a difference of almost 3 sigmas (S.D. = 11.52). Furthermore, the scores of the older group cluster as consistently in the low range on Quantitative as they did in the high range on Vocabulary and have the smallest sub-group variability (S.D.=6.56).

Summary of Preliminary Study. Accumulated data of the approximately 600 credential candidates who were tested at Sacramento State College in the calendar year of 1954 were studied with respect to age and differences in performance on the Vocabulary subtest of the Cooperative English Examination, Higher Level, Form 2 and the Linguistic and the Quantitative sections of the American Council on Education Psychological Examination for Freshmen, 1952 Edition. The population age range of 18 through 64 years was separated into intervals of five years each. The three older groups had the smallest representation with a total of 39 subjects as compared to 548 cases in the three younger groups. The mean age was a little over thirty years.

A positive relationship between Vocabulary and age was
expressed in a product-moment correlation of .37 and a correlation ratio of .41. By an analysis of variance, both measures were found to be highly significant above the .001 level of confidence. When the F-ratio for linearity of regression was calculated, the results signified that a curve would fit the data significantly better than a line \((F < .01)\). It was noted that mean Vocabulary scores increased from 57.64 for the youngest group to 81.00 for the 60-64 subjects, a difference of nearly 2 S.D.s for the total population. The upward progression of mean scores was interrupted by a slight downward trend in the 50-54 and 55-59 age range.

A linear correlation of \(-.07\) between age and Linguistic test scores was found to be not significant. A relationship of this magnitude would occur more than five times in one hundred by chance, so the null hypothesis of no relationship except by chance was accepted. However, a correlation ratio of .37 was found to be highly significant \((F < .001)\), indicating that the progression of mean Linguistic scores with age can be described as curvilinear.

It was observed that Linguistic scores increase from the youngest group through the 30-34 age range, followed by a decrement and a tendency to plateau for the 35 through the 49 year old subjects. This was followed by a much sharper decline in the three older groups. The mean performance of these latter subjects was more than one population S.D. of 16.91 below the high scoring 30-34 group and about three-fourths of a sigma below the total test mean.
A significant negative correlation coefficient of -.40 was obtained between age and Quantitative scores. When the correlation ratio of .44 was also found to be highly significant \( (P < .001) \) the F-ratio for curvilinearity of regression was calculated and found to be significant above the .01 level of confidence. A curve was accepted as giving the line of best fit.

The observation was made that the mean scores of the younger groups form a plateau extending through the 30-34 age range with mean scores from 42.31 to 44.40. Beyond this point, the mean scores decrease for each succeeding age group until a low of 20.50 is reached in the 60-64 interval. The mean Quantitative score for the older group is almost 2 population S.D.'s below the plateau groups and about one-half the total mean score of 40.23.

Conclusions from the Preliminary Study. When the credential candidates in the preliminary study are compared on the basis of age and test scores, significant differences are found between age groups within tests and between age groups and different tests. The observation that vocabulary scores tend to be higher for older credential candidates and that Quantitative scores are frequently depressed in comparison with both their own average level of performance and that of younger students is confirmed.
Analyses of variance revealed significant curvilinear relationships between age and test scores. Vocabulary scores increase with age from the youngest through the oldest group with the 60-64 year old subjects being higher than all other groups. In terms of mean performance, the two youngest groups, age 19 and less and 20-24, are lowest.

When Linguistic scores are distributed by age, the older subjects, 50-54, 55-59, and 60-64, are distinctly lower than all younger groups, having mean scores which fall about three-fourths of a standard deviation below the total group mean. The 25-29 and 30-34 year subjects are relatively highest, and the remaining groups cluster at or slightly below the mean.

On the Quantitative test, which bears a significant negative relationship to age beyond the 30-34 age range, the four younger groups, 19 and less through 30-34, are higher than all other groups. All successive age groups are lower with increasing age.

Discussion. When the results of the preliminary study are considered from the viewpoint of the many problems of adult measurement found in the review of the literature, it is obvious that many speculations may be made to account for the age differences in performance on the CEE Vocabulary and the ACE Linguistic and Quantitative scales.
If the evidence that vocabulary is a relatively stable measure of intelligence at all age levels is accepted, and even making some allowance for increasing vocabulary as a function of age, one may argue that there is considerable reduction in efficiency in whatever mental skills are measured by the Linguistic and Quantitative scales. However, one may assert that because this is a cross-sectional study, the age groups are not homogeneous with respect to intellectual level, cultural variables, motivation, practice and the like. Still apropos the use the of ACE Quantitative and Linguistic scores, and after taking cognizance of studies which show decreasing physiological functions with age, one may be justified in assuming that a power-type task would be more appropriate for the measurement of adult abilities. Moreover, one may question the adequacy of the test instruments for measuring adult ability, both because of the speed factor involved and the nature of the test content.

It is with these questions of the appropriateness and adequacy of the examinations described in the preliminary study, particularly as they affect the educational and vocational plans of more mature students, that the main experiment was undertaken.
CHAPTER IV

THE EXPERIMENT AND RESULTS

A preliminary study of approximately 600 credential candidates tested at Sacramento State College during the three college sessions of 1984 revealed significant differences in test performance for the age groups represented in the population. Following this study, an experiment was designed further to investigate the relationships between age and subtests of the credential candidate screening battery. In addition, an effort was made to study a complex of implicit functions judged to be of importance in following test directions and marking an IBM answer sheet.

I. DESIGN OF THE EXPERIMENT

The Test Battery. A number of considerations were necessary in designing the experiment. Economy required that available materials be utilized as far as possible. To keep student motivation and administrative labor consistent with previous sessions, testing time had to remain the same relative length as usual. The testing plan had to be congruent with the scheduling of interviews regularly conducted by panels of faculty members at each testing session.

With these considerations in mind, a testing procedure was adopted. Content of the regular screening battery was not
changed. This consists of the American Council on Education Psychological Examination for Freshmen, 1952 Edition (ACE), the Cooperative English Examination, Higher Level, Form 2, (CRE), the Minnesota Multiphasic Personality Inventory, the Minnesota Teacher Attitude Inventory and individual student interviews with panels of faculty members. A compensation for the additional time required to effect the experiment was made by giving the short rather than the long form of the personality inventory. The experiment is concerned with the test data of the CRE Vocabulary subtest (henceforth referred to as timed Vocabulary) and the Quantitative and Linguistic sections of the ACE. These have been described in Chapter III in connection with the preliminary study.

Three additions were made to the test battery: the Vocabulary subtest of the Cooperative English Examination, Higher Level, Form 3 (untimed Vocabulary), the Arithmetic subtest of the Quantitative section of the American Council on Education Psychological Examination for Freshmen, 1948 Edition (untimed Arithmetic), and a Marking Test designed by the author. Appropriate headings were mimeographed on standard IBM answer sheets for ease of administering, taking and scoring the tests. A description of the tests and the purpose of including them in the experiment follows:

CRE Vocabulary, Untimed. The two forms of the Vocabulary subtests are similar in every respect except content.
Both consist of 60 five-choice items. The Form Z Vocabulary was administered with the standard time limit of 15 minutes in the regular battery, but the alternate form was administered without time limit. Both are recognition type tests with gradually increasing work difficulty and with items requiring increasingly fine discriminations.

The Vocabulary subtest of the regular battery has been assumed to be a relatively untimed task. The variant form was given on an untimed basis for the purpose of checking this hypothesis. The review of the literature has established that vocabulary tests provide a relatively stable criterion of mental ability over a wide span of years, hence it is proposed to use the present vocabulary tests as a base for comparison with other test scores.

**ACE Arithmetic, Untimed.** Both forms of the ACE Arithmetic test consist of 20 five-choice items of the type that are commonly called problems in arithmetic reasoning. The form used in the regular timed battery is preceded by practice exercises and the test proper has a time limit of 10 minutes. The alternate form was administered without time limit.

The untimed Arithmetic test was included to discover what differences would occur between the timed and untimed versions in relation to the group as a whole, the age groups, and the Marking Test. Extraneous sources of variance leading to differences in performance on the timed Arithmetic could be
expected to be minimized or lacking on the untimed subtest.

The ACE Arithmetic subtest is invariably the first
timed test in the screening battery. Factors which may af-
fect performance are preliminary instructions for using an
IBM answer sheet, speed and precision of recording answers,
speed of recall, and age. A fifth factor which may effect
age differences in performance on the timed test is initial
anxiety and that which is specific to arithmetic. Experience
with the older credential candidates has been that as a group
they verbalize far more lack of confidence than the younger
students and that they react very specifically to arithmetic
content.

The Marking Test. The Marking Test was given with a
one-minute time limit. Directions for taking the test, read
orally by the examiner and silently by the students, were
mimeographed on the back of standard IBM answer sheets. A
sample of this test is on the following page (Figure 1).

The Marking Test was designed to isolate a complex
of implicit functions which are required in using an IBM
answer sheet in a test situation. Variables which may be
assumed to effect differences in performance on this test are
visual acuity, motor reaction time and coordination, spatial
factors, motivation, and interpretation of the instructions
which precede the test proper. The ACE is the first timed
test of the battery. The standard general directions for
We are interested in determining to what extent speed and dexterity are involved in taking a paper and pencil test. For this reason we are asking you to cooperate by taking this test of speed and precision in marking an IBM answer sheet. Since the test has a very short time limit, it is essential that you follow directions exactly. You must not turn over the page until the examiner says "go." You must stop immediately at the signal to stop.

On the opposite side of this page you will find five columns of answer spaces numbered from 1 to 150. Opposite each item number are answer spaces arranged like the sample below.

As fast as you can, you are to mark with your pencil between the rows of double lines, making a solid black pencil mark. First fill in all the spaces opposite the number 1 at the side of the page, then fill in the spaces opposite item number 2, and so on down the column, working as fast as you can. In the sample at the left, the first three answer spaces have already been filled in. As soon as you have finished one row, go right on to the next. Your score will be the total number of spaces filled.

The answer sheet will be scored by an electric scoring machine, so you must use the special pencils which have been given you. The Scoring machine will not record your answers unless they are made with a heavy black pencil mark between the rows of small dotted lines. Heavy black marks are made by pressing firmly with your pencil and by marking up and down two or three times.

Before the test proper, you will be given a small practice exercise. When the examiner says "go," but not before, start marking the sample above. Begin with the fourth answer space opposite item number 1 and continue working down the page, marking as fast as you can. Remember to start and stop immediately. You will be allowed 10 seconds. Are there any questions?
<table>
<thead>
<tr>
<th>SEX</th>
<th>AGE</th>
<th>INSTRUCTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DATE OF BIRTH</th>
<th>GRADE OR CLASS</th>
<th>NAME OF TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PART</th>
<th>NAME</th>
<th>SCHOOL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BE SURE YOUR MARKS ARE HEAVY AND BLACK.
ERASE COMPLETELY ANY ANSWER YOU WISH TO CHANGE.

Printed by the International Business Machines Corporation, Endicott, N. Y., U. S. A.
IBM FORM T.S. 1000
taking the test emphasize first, that speed on each kind of
test can be determined, and secondly, that the IBM answer
sheets should be marked in a specific and precise manner (2). The dual emphasis on speed and precision of marking answer
sheets is subject to individual interpretation. Inspection of answer sheets clearly indicates that some students empha-
size speed and others emphasize precision.

Rather than devise a series of tasks to isolate each compoenent function in simpler form, the decision was made to have one test which would realistically duplicate the inter-
action of functions required in an actual test situation. Thus, it is assumed that speed, precision, and interpretation of directions will not alone determine how individuals will perform, but rather individual interactions of these variables.

In order to determine an adequate time limit, the probable range and distribution of responses, and the adequacy of instructions, the Test Officer administered the Marking Test to a group of 49 students enrolled in Psychology 6, Reading Improvement, for a preliminary trial. The age range was fairly comparable to that of the experimental group. At time limits of one, two and three minutes, the students were requested to circle the spaces they were marking. The range and distribution were determined for each time interval with so little difference that the one-minute time limit was deemed satisfactory. Only a minor alteration was made in the printed instructions.
Description of the Population. The Test Officer of the College facilitated the experiment by making available for study the 1955 Summer Session group of credential candidates. Only those individuals whose test data were complete were included in the study. The data for two persons with known and highly limiting physical handicaps were withheld. The final data were derived from 193 candidates. Of this number, 49 were seeking administrative or supervisory credentials. The remainder were candidates for regular daytime teacher credentials, many of whom were experienced teachers with provisional status. The two groups were combined for study purposes when it appeared from an inspection of range and distribution of test scores that they were comparable. Of the total group, 124 were women, 69 were men.

Test Procedure. The test design was carried out in a manner which was similar in every respect to the previous summer session procedures except for the additions to the battery mentioned above. All candidates were tested in two consecutive afternoon sessions approximating four hours each. Teacher candidates were administered the Marking Test, MMPI, MTAI, Vocabulary untimed and Arithmetic untimed on the first
afternoon and in the order named. They also had interviews with faculty panels during this time. The second afternoon they were administered the complete timed ACE Psychological Examination and the Cooperative English Examination (CEE). The procedure was exactly reversed for the administrative and supervisory credential candidates.

This order of testing may be criticized in that the untimed Arithmetic test was administered before the timed test to the larger teacher candidate group. To meet the initial considerations mentioned at the beginning of this chapter, the risk had to be taken that if there are significant variations with age in speed of recall, this part of the experiment might be vitiated by practice on the untimed test the preceding day.

Two considerations suggest that this was not the case. In Figure 18, to be discussed later, a comparison of the preliminary study with the experimental data reveals negligible differences between means and standard deviations for the paired age groups. Performance on timed Arithmetic cannot be directly compared, but a comparison can be made between the mean ACE Quantitative raw scores of which timed Arithmetic is a part. The Quantitative test has a mean raw score of 40.23 and a standard deviation of 11.52 for the preliminary group, and a mean raw score of 40.41 and standard deviation of 11.80 for the experimental group. The preliminary study,
it may be remembered, was a composite of data from three testing sessions conducted in 1954. A comparison similar to the preceding may be made between the experimental group and the Spring and Fall 1955 groups of teacher candidates (33). N is 220 and 242. The mean Quantitative raw scores for these groups are 39.6 and 39.3 respectively. This is reasonable evidence that the order of testing had little effect on overall results.

II. RESULTS OF THE EXPERIMENT

The experimental data are discussed in the following order: the Marking Test is compared with age, verbal tests and quantitative tests; in the same order, verbal and quantitative tests are compared with age. In addition, comparisons are made between Vocabulary timed and untimed, and Vocabulary and the Linguistic scale. Arithmetic timed is compared with both Arithmetic untimed and the Quantitative test of which timed Arithmetic is a part.

Following this, and in view of the observations which initiated the experiment, an out-of-pattern high scoring Vocabulary group is compared with the residual group, both groups together constituting the experimental population. Finally, the results of the preliminary study are compared with those of the experiment, including a comparison of test performance by age groups on the percentile and decile
norms in present use for the credential candidate population.

As in the preliminary study, all mathematical formulae and methods used in the statistical treatment of the data are according to McNemar (44). Means and standard deviations were computed for the total test scores and for age group distributions. Pearsonian product-moment correlation coefficients and correlation ratios were computed for each comparison. The data were subjected to analyses of variance to test the null hypothesis that means of test scores do not differ from zero except by chance.

To maintain comparability with the preliminary study, the age range was again divided into intervals of five years. The intervals represent the year of age at the time of testing regardless of accumulated months. The age range is 18 through 64. Because of the small number of cases at either end of the distribution, the one student below age 20 was included in the 20-24 age group and the 8 students who were 55 years of age and more were included in the 50-54 age group. The age range is thus represented in seven intervals. The number of cases in each group from youngest to oldest is 34, 39, 53, 30, 23, 18 and 16 respectively. The mean age for the total population is 34.26 with an S.D. of 9.36.

**Marking Test Comparisons.**

In the following discussion, the results of the Marking Test will be compared with age, the verbal tests and the
quantitative tests. Except for the first comparison, the regression of Marking Test scores on age intervals, all comparisons will be in terms of the regression of verbal and quantitative test scores on the Marking Test. The number of cases at either extreme of the distribution is small and subject to sampling errors.

**Marking Test and Age.** Table VII presents the product-moment correlation, the correlation ratio and the analysis of variance for the regression of Marking Test raw scores on age.

The correlation coefficient between the Marking Test and age is -.15. To test the significance of the linear correlation, the variance estimates for the linear regression and residuals from the line were calculated at 14.88 and 3.38 respectively, resulting in an F-ratio of 4.40. For 1 and 191 degrees of freedom, this relationship would occur by chance less than five times in one hundred (F < .05) and the null hypothesis is therefore rejected.

A correlation ratio (Eta) of .15 was found to be a chance relationship. The mean squares for between and within-array means were 2.77 and 3.47 respectively, yielding an F-ratio of .80. For 6 and 186 df, this relationship would occur by chance more than 50 per cent of the time, and the null hypothesis of no curvilinear relationship is retained. The conclusion may be made that speed, precision, and interpretation of preliminary instructions for using an IBM answer sheet as
TABLE VII

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of Marking Test Raw Scores on Age for the Experimental Group (N = 193)

<table>
<thead>
<tr>
<th>Marking Test and Age</th>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.15</td>
<td>.15</td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>14.88</td>
<td>1</td>
<td>14.88</td>
<td>4.40</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>1.72</td>
<td>5</td>
<td>.34</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>Between-array means</td>
<td>16.60</td>
<td>6</td>
<td>2.77</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>Within arrays</td>
<td>544.89</td>
<td>186</td>
<td>3.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>546.61</td>
<td>191</td>
<td>3.38</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total 661.49 192
measured by the Marking Test have a significant negative linear relationship with age for the group studied.

A graphic comparison of the means and standard deviations of the Marking Test scores on age groups is given in Figure 2. The graph indicates the tendency for mean performance on the Marking Test to be inversely related to age. This represents a difference in score from 185.5 for the 20-24 age group to 169.1 for the 40-44 group. The two groups of older subjects, 45-49 and 50-64, with mean scores of 171.7 and 169.5 on the Marking Test, differ little from the 40-44 age group. The total group mean is 178.22.

Although the downward trend in mean score performance is apparent, sub-group variabilities tend to be of such magnitude as to lead to the conclusion that individual differences in performance within age groups are greater than differences between age groups. The difference between the highest and lowest mean scores on the Marking Test (185.5 and 169.1) is 16.4 raw score points. The range of mean scores for the total population is only 44 per cent of the population S.D. of 57.03.

There is no trend for variability to increase with age. The 40-44 age group has the largest S.D. of 40.38, but this differs little from that of the 30-34 subjects who have an S.D. of 38.94 and the 50-54 subjects who have an S.D. of 39.35. The 30-34 and 45-49 age groups are the most stable
FIGURE 2
Mean Scores and Standard Deviations for Marking Test Raw Scores on Age for the Experimental Group (N = 193)

Marking Test Mean Scores
M = 178.22, S.D. = 37.03

Standard Deviations

Midpoints of Age Intervals
N = 34 39 33 30 23 18 16
with S.D.s of 30.74 and 29.95 respectively.

Since the comparisons which follow are based on the regression of verbal and numerical tests on the Marking Test, Table VIII presents the mean age of each Marking Test interval for purposes of comparison. The tendency for mean age to decrease with increments in Marking Test scores is apparent. The slowest group on the Marking Test has a mean age of 43.35. The fastest group has a mean age of 28.65. The total group mean age is 34.25 with an S.D. of 9.36.

Marking Test and Timed Vocabulary. Correlations and an analysis of variance for the regression of the timed Vocabulary test on the Marking Test are given in Table IX.

A correlation coefficient of -.07 was obtained between these two variables. Testing for significance of the relationship, variance estimates of 4.79 and 5.20 for the linear regression and residuals from the line yielded an F-ratio of .92. For 1 and 191 degrees of freedom (df), a ratio of this magnitude would occur by chance more than fifty per cent of the time.

The Eta coefficient is .18 and when this was subjected to analysis, an F-ratio of .65 was calculated from mean squares of 3.45 and 5.38 for between and within-arrays. For 9 and 183 df, a ratio of this size would occur by chance fifty per cent of the time. The null hypothesis of no linear or curvilinear relationship between Marking Test scores and timed Vocabulary is retained.
TABLE VIII

Means for the Regression of Age on the Marking Test for the Experimental Group (N = 195)

<table>
<thead>
<tr>
<th>Marking Test Intervals</th>
<th>Mean Age</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 271</td>
<td>28.65</td>
<td>3</td>
</tr>
<tr>
<td>251 - 270</td>
<td>32.00</td>
<td>7</td>
</tr>
<tr>
<td>231 - 250</td>
<td>28.65</td>
<td>3</td>
</tr>
<tr>
<td>211 - 230</td>
<td>35.95</td>
<td>19</td>
</tr>
<tr>
<td>191 - 210</td>
<td>32.55</td>
<td>37</td>
</tr>
<tr>
<td>171 - 190</td>
<td>32.35</td>
<td>43</td>
</tr>
<tr>
<td>151 - 170</td>
<td>34.55</td>
<td>35</td>
</tr>
<tr>
<td>131 - 150</td>
<td>36.40</td>
<td>28</td>
</tr>
<tr>
<td>111 - 150</td>
<td>32.35</td>
<td>14</td>
</tr>
<tr>
<td>&lt; 110</td>
<td>43.25</td>
<td>4</td>
</tr>
</tbody>
</table>

Group Mean: 34.25
S.D.: 9.36
TABLE IX

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of TIMED Vocabulary Scaled Scores, Form E of the OHE, on the Marking Test for the Experimental Group (N = 193)

<table>
<thead>
<tr>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marking Test and Timed Vocabulary</td>
<td>-.07</td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>4.79</td>
<td>1</td>
<td>4.79</td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>26.24</td>
<td>8</td>
<td>3.28</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>Between-array means</td>
<td>31.04</td>
<td>9</td>
<td>3.45</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>Within arrays</td>
<td>966.89</td>
<td>188</td>
<td>5.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>993.14</td>
<td>191</td>
<td>5.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>997.95</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Marking Test and Untimed Vocabulary. Correlations and an analysis of variance are given in Table X.

A correlation coefficient of -.09 between Marking Test scores and untimed Vocabulary scores was tested for significance. Variance estimates of 6.68 and 3.89 for the linear regression and residuals from the line respectively, yielded an F-ratio of 1.72 which for 1 and 191 df would occur more than five times in one hundred by chance. The null hypothesis is retained.

An Eta coefficient of .24 was tested for curvilinearity of regression. Variance estimates of 4.64 and 3.67 for between and within-array means were computed and resulted in an F-ratio of 1.80. For 9 and 183 df, this ratio would occur more than 5 times in one hundred by chance, and the null hypothesis is again retained.

Means and standard deviations for the regression of timed and untimed Vocabulary scores on the Marking Test are presented in Figure 5. No definite trends are noted in mean score performance on timed and untimed Vocabulary in relation to the Marking Test. Deviations from the population means of 87.52 and 89.60 respectively, occur only in groups with a very small number of cases. Group variability decreases from S.D. 11.35 on timed Vocabulary to 9.85 on untimed Vocabulary. Gains on the untimed test are even throughout the age range except for the three subjects who scored
### TABLE X

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of UNTIMED Vocabulary Scaled Scores, Form B of the CEE, on the Marking Test for the Experimental Group (N = 193)

<table>
<thead>
<tr>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marking Test and Untimed Vocabulary</td>
<td>-.09</td>
</tr>
</tbody>
</table>

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>6.68</td>
<td>1</td>
<td>6.68</td>
<td>1.72</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>35.11</td>
<td>8</td>
<td>4.39</td>
<td>1.13</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Between-array means</td>
<td>41.78</td>
<td>9</td>
<td>4.64</td>
<td>1.20</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Within arrays</td>
<td>708.36</td>
<td>183</td>
<td>3.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>743.46</td>
<td>191</td>
<td>3.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>750.14</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 3

Mean Scores and Standard Deviations for Timed and Untimed CEE Vocabulary Subtests (Forms Z and S) on the Marking Test for the Experimental Group (N = 193)

\[ \text{Mean Scores} \]

\[ \text{Standard Deviations} \]

\[ N = 4 \quad 14 \quad 28 \quad 35 \quad 43 \quad 37 \quad 19 \quad 3 \quad 7 \quad 3 \]
third from the highest on the Marking Test. Whatever differences occur between timed and untimed Vocabulary appear to be unrelated to the Marking Test.

**Marking Test and Linguistic Scale.** Correlations and an analysis of variance for the regression of ACE Linguistic raw scores on Marking Test scores are presented in Table XI.

A linear coefficient of .08 between the two variables was tested for significance. Mean squares of 12.72 and 10.95 were calculated for the linear regression and residuals from the line respectively. This yielded an F-ratio of 1.17 which, for 1 and 191 degrees of freedom, would occur by chance more than five times in one hundred (P > .05).

The correlation ratio of .14 was subjected to analysis. An F-ratio of .40 was obtained from variance estimates of 4.52 and 11.17 for between and within-array means respectively. For 9 and 165 degrees of freedom, this relationship would occur by chance more than 50 per cent of the time. The null hypothesis of no other than chance relationship between the Marking Test and the Linguistic scale is retained.

Figure 4 presents the means and standard deviations for the Linguistic scores on the Marking Test. Excepting the means of the slowest and the fastest subjects on the Marking Test, the Linguistic mean scores cluster around the group mean of 75.50 (S.D.=16.42). The tendency of the slowest subjects on the Marking Test (N = 4) to score low on the Linguistic
TABLE XI


<table>
<thead>
<tr>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marking Test and Linguistic Score</td>
<td>.06</td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>12.72</td>
<td>1</td>
<td>12.72</td>
<td>1.17</td>
<td>&gt; .05</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>27.95</td>
<td>8</td>
<td>3.49</td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>Between-array means</td>
<td>40.65</td>
<td>9</td>
<td>4.52</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>Within arrays</td>
<td>2043.92</td>
<td>193</td>
<td>11.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>2071.85</td>
<td>191</td>
<td>10.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2084.57</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 4

Mean Scores and Standard Deviations for ACE Linguistic Test Scores (1952 Edition) on the Marking Test for the Experimental Group (N = 193)

Linguistic Mean Scores
M = 75.50, S.D. = 16.42

Standard Deviations

N = 4 14 28 35 43 37 19 3 7 3
Scale, and the fastest subjects ($N = 5$) to score high, is an interesting reminder of studies which have found positive correlations between speed and intelligence. However, if Table VIII, page 96, which lists the mean ages for Marking Test intervals is referred to, it may be noted that the slowest group on the Linguistic and Marking Test has a mean age of 43.25 and that the highest group on both tests has a mean age of 28.65. A logical conclusion is that for these two small groups at the extremes of the distribution, age is perhaps uniquely involved in the relationship between the Marking and Linguistic tests.

**Marking Test and Timed Arithmetic.** Correlations and an analysis of variance for the regression of timed Arithmetic scores on the Marking Test are given in Table XII.

A correlation coefficient of -.11 was calculated for the two variables. The F test of significance was computed from mean squares of 27.78 and 12.80 for the linear regression and residuals from the line respectively, yielding a ratio of 2.22. For 1 and 191 df, a ratio of this magnitude would occur more than five times in one hundred by chance ($F > .05$). The null hypothesis of no relationship is retained.

When the Eta coefficient of .20 between the Marking Test and timed Arithmetic was tested for significance variance estimates of 11.09 and 12.66 for between and within-array
TABLE XII

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of TIMED Arithmetic Raw Scores, 1966 Edition of the ACE, on the Marking Test for the Experimental Group (N = 163)

<table>
<thead>
<tr>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marking Test and Timed Arithmetic</td>
<td>-.11</td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>27.78</td>
<td>1</td>
<td>27.78</td>
<td>2.22</td>
<td>&gt; .05</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>72.00</td>
<td>8</td>
<td>9.00</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>Between-array means</td>
<td>99.78</td>
<td>9</td>
<td>11.09</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>Within arrays</td>
<td>2316.34</td>
<td>183</td>
<td>12.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>2388.24</td>
<td>191</td>
<td>12.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2416.02</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


means yielded an F-ratio of .89. For 9 and 183 degrees of freedom, this ratio would occur by chance more than 50 per cent of the time.

**Marking Test and Untimed Arithmetic.** Table XIII presents the correlations and analysis of variance for the two tests.

The linear coefficient of -.007 was obtained between the Marking Test and Untimed Arithmetic. An F-ratio of .009 between mean squares of .10 and 10.75 for the linear regression and residuals from the line respectively, indicated that for 1 and 191 df the relationship would occur by chance more than fifty per cent of the time. The null hypothesis of no relationship is accepted.

The R2 coefficient of .24 between the two tests was analyzed for significance of curvilinearity. Between and within-arrays were found to have mean squares of 13.64 and 10.55 respectively, yielding an F-ratio of 1.29. For 9 and 183 df, a ratio of this magnitude would occur more than five times in one hundred by chance. Hence, the null hypothesis is accepted for both linearity and curvilinearity of regression.

Graphic comparisons of means and variabilities for timed and untimed Arithmetic on the Marking Test are given in Figure 5. Gains in Arithmetic scores when given on an
TABLE XIII

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of UNTIMED Arithmetic Raw Scores, 1948 Edition of the ACE, on the Marking Test for the Experimental Group, (N = 193)

<table>
<thead>
<tr>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marking Test and Untimed Arithmetic</td>
<td>-.007</td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>.10</td>
<td>1</td>
<td>.10</td>
<td>.009</td>
<td></td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>122.64</td>
<td>8</td>
<td>15.33</td>
<td>1.45</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Between-array means</td>
<td>122.74</td>
<td>9</td>
<td>13.64</td>
<td>1.29</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Within arrays</td>
<td>1989.78</td>
<td>185</td>
<td>10.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>2052.42</td>
<td>192</td>
<td>10.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2052.52</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 5

Mean Scores and Standard Deviations for ACE Arithmetic Raw Scores (Untimed and Timed), 1948 and 1952 Editions, on the Marking Test for the Experimental Group (N=193)

Arithmetic Untimed
M = 15.80, S.D. = 3.26

Arithmetic Timed
M = 8.68, S.D. = 5.54

Standard Deviations

N = 4 14 28 35 43 37 19 3 7 3
untimed basis are remarkably even when compared to the Marking Test. Such deviations as occur are in intervals with a small number of subjects.

**Marking Test and Quantitative Scale.** Data for this comparison are given in Table XIV.

A linear coefficient of .02 was obtained between the Marking Test and Quantitative scores. Testing for significance, variance estimates of .50 for the linear regression and 8.78 for residuals from the line yielded an F-ratio of .06. For 1 and 191 df, a ratio of this magnitude would occur by chance more than 50 per cent of the time. A similar chance relationship was found for the Eta coefficient of .18 when the F-ratio was calculated at .72 for mean squares of 6.58 and 8.86 for between and within-arrays respectively. For 9 and 183 df, a ratio of this size is not significant. The null hypothesis of no significant linear or curvilinear relationship between the Marking Test and Quantitative scores is retained.

Means and standard deviations for the Quantitative scores on the Marking Test are pictured in Figure 6. Except for two groups, the Quantitative mean scores cluster around the total group mean of 40.41.

It may be noted that the slowest group on the Marking Test (N = 4) also scored lowest on the Quantitative scale, a difference of approximately 1 S.D. below the group mean.
### TABLE XIV


<table>
<thead>
<tr>
<th></th>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marking Test and Quantitative Scale</td>
<td>.08</td>
<td>.19</td>
</tr>
</tbody>
</table>

### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>.50</td>
<td>1</td>
<td>.50</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>64.88</td>
<td>8</td>
<td>7.11</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>Between-array means</td>
<td>57.39</td>
<td>9</td>
<td>6.38</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>Within arrays</td>
<td>1620.58</td>
<td>183</td>
<td>8.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>1677.47</td>
<td>191</td>
<td>8.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1677.97</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Quantitative Mean Scores
Mean = 40.41, S.D. = 11.80

Standard Deviations

Midpoints of Marking Test Intervals

N = 4 14 28 35 43 37 19 3 7 3
These are the same subjects who made more than average gains on untimed Arithmetic, who made the lowest mean score on the Linguistic scale, and who scored highest on both timed and untimed Vocabulary. It appears that, at least for this small group of four subjects, the functions measured by the Marking Test may be negatively related to performance on timed tests other than Vocabulary. Since this same group had the highest mean age (45.25) of any Marking Test interval, it appears that the low Marking Test scores may be a function of age.

One may conclude that there is no relationship of statistical significance for the group as a whole between the Marking Test and any of the numerical tests. In all instances, the obtained linear and curvilinear correlations would occur more than five times in one hundred by chance.

Age Comparisons.

In the following discussion, the regression of verbal and numerical tests will be considered. For these comparisons, the number of cases in each interval is more evenly distributed than for the comparisons with the Marking Test. However, there are only seven age groups and, as a result, the obtained correlations may be expected to be somewhat lower than if more intervals had been used.

Since the range is also restricted as compared to the preliminary study where ten intervals were possible, it is probable that the obtained relationships in the experiment
may differ from those of the preliminary study. There is a
difference of over four years in the mean age of the prelim-
inary and experimental groups with fewer subjects in the
lower age range for the latter. This is a characteristic
difference between the summer and regular college semester
populations.

**Age and Timed Vocabulary.** Correlations and an analysis
of variance are presented in Table XV.

A linear coefficient of .19 was obtained between timed
Vocabulary and age. When the $F$ test was applied, mean squares
for the linear regression and residuals from the line were com-
puted at 35.55 and 5.04 respectively, resulting in a ratio of
7.08. For 1 and 191 degrees of freedom, the linear coeffici-
ent is significant beyond the .01 per cent level of confi-
dence. That is, a relationship of this magnitude would occur
by chance less than one time in one hundred. The null hypoth-
thesis was rejected and a linear relationship assumed.

Testing an obtained correlation ratio of .23 between
the two variables for significance of curvilinearity, the
variance estimates were found to be 8.56 and 5.09 for between
and within-array means. This results in an $F$-ratio of 1.68.
For 6 and 186 df, a ratio of this size would occur by chance
more than five times in one hundred ($F > .05$) and the hypoth-
thesis of linearity is retained.
TABLE XV

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of TIMED Vocabulary Scaled Scores, Form Z of the CEE, on Age for the Experimental Group (N = 193)

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>35.53</td>
<td>1</td>
<td>35.53</td>
<td>7.05</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>15.87</td>
<td>5</td>
<td>3.17</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>Between-array means</td>
<td>51.39</td>
<td>6</td>
<td>8.56</td>
<td>1.68</td>
<td>&gt; .05</td>
</tr>
<tr>
<td>Within arrays</td>
<td>948.54</td>
<td>196</td>
<td>5.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>968.40</td>
<td>191</td>
<td>5.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>997.93</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Age and Untimed Vocabulary. Data for this comparison are given in Table XVI.

For untimed Vocabulary and age, a linear coefficient of .15 was calculated. The analysis of variance yielded mean squares of 16.80 for the linear regression and 5.84 for residuals from the line. The resulting F-ratio of 4.38, for 1 and 191 df, would be of this magnitude less than five times in one hundred ($P < .05$). The null hypothesis is rejected.

An Eta coefficient of .24 between the two variables was tested for curvilinearity. Variance estimates of 7.19 and 3.80 for the between and within-arrays yielded an F-ratio of 1.89. For 6 and 186 df, the obtained degree of relationship would occur more than five times in one hundred by chance ($P > .05$), and the null hypothesis that array means do not deviate from a straight line except by chance is retained. The relationship between age and untimed Vocabulary is described by a line.

A graphic comparison of the means and variabilities for timed and untimed Vocabulary on age are given in Figure 7. When compared with the timed Vocabulary test, it may be noted that there are minor increases in mean scores on untimed Vocabulary for every age group except the 45-49 year age group whose mean score drops .55 of a raw score point below their own mean on timed Vocabulary. The total group increase is from a mean of 67.52 to 69.60. The two younger groups, 20-24
### TABLE XVI

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of UNTIMED Vocabulary Scaled Scores, Form B of the CER, on Age for the Experimental Group ($N = 183$)

<table>
<thead>
<tr>
<th>Age and Untimed Vocabulary</th>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.15</td>
<td>.24</td>
</tr>
</tbody>
</table>

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>16.80</td>
<td>1</td>
<td>16.80</td>
<td>4.58</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>26.33</td>
<td>5</td>
<td>5.27</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>Between-array means</td>
<td>43.13</td>
<td>6</td>
<td>7.19</td>
<td>1.69</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Within arrays</td>
<td>707.01</td>
<td>186</td>
<td>3.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>753.54</td>
<td>191</td>
<td>3.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>750.15</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 7

Mean Scores and Standard Deviations for CEE Vocabulary Scaled Scores, (Timed and Untimed) Forms Z and S, on Age for the Experimental Group (N=193)
and 25-29, have lower mean scores on both tests than any of 
the older groups.

Variability is very consistent between age groups on 
both Vocabulary tests with a decrease in standard deviation 
from 11.35 on the timed test to 9.46 on the untimed test. 
These relationships support the hypothesis that the Vocabulary 
test as given on a timed basis in the regular battery is equival-
ent to a power-type task.

**Vocabulary Timed and Untimed.** Further confirmation of 
the above hypothesis was established by computing the correla-
tions and an analysis of variance for the regression of timed 
Vocabulary on untimed Vocabulary. These data are presented 
in Table XVII.

Both the linear correlation and Eta were calculated at 
.84. Testing for significance of linearity, variance estimates 
of 703.64 and 1.54 were computed for the linear regression and 
residuals from the line respectively. This resulted in an F-
ratio of 456.91 which, for 1 and 191 degrees of freedom, would 
occur by chance less than one time in one thousand (F < .001).

The correlation ratio was also found to be highly sig-
nificant when the obtained mean squares of 78.76 and 1.88 for 
between and within-arrays respectively, resulted in an F-ratio 
of 49.85 (F < .001).

To ascertain whether a line or a curve would describe 
the data better, the variance estimates for the deviation of
TABLE XVII

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of TIMED Vocabulary Scaled Scores, Form 2 of the CEE, on UNTIMED Vocabulary Scaled Scores, Form 8 of the CEE, for the Experimental Group (N = 193)

<table>
<thead>
<tr>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary Timed and Vocabulary Untimed</td>
<td>.84</td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>705.64</td>
<td>1</td>
<td>705.64</td>
<td>456.91</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>5.19</td>
<td>8</td>
<td>.65</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td>Between-array means</td>
<td>706.83</td>
<td>9</td>
<td>78.76</td>
<td>49.85</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Within arrays</td>
<td>289.10</td>
<td>163</td>
<td>1.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>294.29</td>
<td>191</td>
<td>1.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>997.93</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
means from a line and for within-arrays were calculated at .68 and 1.88, yielding an $F$-ratio of .41 which for 8 and 183 df is not significant. Since the array means do not deviate from a straight line except by chance, an hypothesis of a significant linear relationship between timed and untimed Vocabulary is retained.

Inspection of Figure 8 further demonstrates the constant relationship between timed and untimed Vocabulary with the exception of minor deviations in the low score range where the number of cases is extremely small. Consistency of scatter is also evident in the small standard deviations for the intervals.

The highly significant correlation coefficient of .84 between the two forms of Vocabulary indicates a high degree of reliability between the two forms of the Vocabulary test and confirms the hypothesis that the timed Vocabulary test of the CER is equivalent to the untimed test.

The answer sheets of the experimental group were checked to see how many students did not complete the timed test. It was found that 19 per cent of the population failed to answer the last few items. Whether this reflects effects of timing or individual limits of knowledge was not determined.

**Age and Linguistic Scale.** The correlation coefficient, correlation ratio and an analysis of variance for the regression
FIGURE 8

Mean Scores and Standard Deviations for TIMED Vocabulary Scaled Scores, Form Z of the CEE, on UNTIMED Vocabulary Scaled Scores, Form S of the CEE, for the Experimental Group (N=193)

TIMED Vocabulary Mean Scores
M = 67.52, S.D. = 11.37

Standard Deviations

Midpoints of UNTIMED Vocabulary Scaled Score Intervals

N = 1 2 5 23 36 24 45 25 12 20
of ACE Linguistic raw scores on age are given in Table XVIII.

A linear coefficient of \( -0.22 \) was calculated between the two variables. Analysing for the significance of the linear relationship variance estimates of 98.39 and 10.40 were computed for the linear regression and residuals from the line respectively. This resulted in an F-ratio of 9.46 which, for 1 and 191 df, would be of this magnitude less than one time in one hundred by chance \( (P < 0.01) \). The null hypothesis is thus rejected and an hypothesis of negative linear regression assumed.

An Eta coefficient of \( 0.25 \) between age and Linguistic scores was subjected to analysis for significance of curvilinearity. An F-ratio of 2.03 was obtained from mean squares of 21.33 and 10.52 for between and within-arrays respectively. For 6 and 186 df, a ratio of this magnitude would occur about five times in one hundred by chance \( (P = 0.05) \).

Testing for significance of linearity of regression, variance estimates of 5.92 and 10.52 for the deviation of means from the line and within-arrays respectively, yielded an F-ratio of \( 0.56 \). For 5 and 186 df, a ratio of this magnitude would occur by chance more than fifty per cent of the time. The null hypothesis is retained for curvilinearity, and it is determined that there is a significant negative linear relationship between age and Linguistic scores for the population studied.
TABLE XVIII


<table>
<thead>
<tr>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age and Linguistic Scores</td>
<td>-.22</td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>98.39</td>
<td>1</td>
<td>98.39</td>
<td>9.46</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>29.60</td>
<td>5</td>
<td>5.92</td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td>Between-array means</td>
<td>127.99</td>
<td>6</td>
<td>21.33</td>
<td>2.05</td>
<td>.05</td>
</tr>
<tr>
<td>Within arrays</td>
<td>1966.59</td>
<td>190</td>
<td>10.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>1984.18</td>
<td>191</td>
<td>10.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2084.57</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Means and variabilities for the regression of Linguistic scores on age are graphically presented in Figure 9. Mean Linguistic scores are on a plateau from the youngest group of 20-24 years through the 35-39 year old group, and are above the group mean of 75.80. The 40-44 year group drops more than one-third of a standard deviation (S.D. = 16.43) below the plateau group, and the two oldest groups, 45-49 and 50-54, have mean scores which are more than one-half the standard deviation below the high-scoring younger groups. There seems to be a trend, but with exceptions, for variability to increase with age.

The progress of Linguistic mean scores with increasing age is a reversal of the trend found between Vocabulary and age where the two younger age groups have mean scores below all other groups. Only the 30-34 and 35-39 age groups maintain their relative standing above the total mean score on both the Vocabulary and the Linguistic tests.

**Linguistic Scale and Timed Vocabulary.** The review of the literature (Chapter II) has established that vocabulary tests are a relatively stable measure of intelligence throughout the span of years considered in this study. Since the Linguistic scale is designed to be a measure of scholastic ability and, as such, is closely related to verbal tests of intelligence, it is to be expected that there will be a high
FIGURE 9


Linguistic Mean Scores
M = 75.50, S.D. = 16.43

Standard Deviations

Midpoints of Age Intervals

N = 34 39 33 30 23 18 16
correlation between the timed CKE Vocabulary subtest and the Linguistic Scale.

Correlations and an analysis of variance for the two tests are given in Table XIX. A linear correlation of .70 was obtained between the two variables. When analysed for significance, variance estimates of 1031.44 and 5.57 for the linear regression and residuals from the line respectively, yielded an F-ratio of 183.50. For 1 and 191 df, a ratio of this magnitude would occur by chance less than one time in one thousand, and hence is highly significant. The null hypothesis is rejected.

A correlation ratio of .72 for the timed Vocabulary and Linguistic scale was analyzed for significance of survi-linearity. An F-ratio of 17.93 was calculated from variance estimates of 98.81 and 5.51 for between and within-array means respectively. For 11 and 181 df, this ratio is also highly significant (F < .001).

To discover whether a line or a curve would describe the line of best fit, the variance estimates for the deviation of means from a line and for within-array means were calculated at 6.55 and 5.51 respectively. This yielded an F-ratio of 1.18 which, for 10 and 181 df, would occur by chance more than five times in one hundred. Thus the apparent departure of the array means from linearity in Table XIX is not sufficiently great to lead to a rejection of an hypothesis of linear regression.
### TABLE XIX

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of AGE Linguistic Raw Scores, 1959 Edition, on TIMED Vocabulary Scaled Scores, Form E of the CER, for the Experimental Group (N = 182)

<table>
<thead>
<tr>
<th>Source</th>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistic Scores and Timed Vocabulary</td>
<td>.70</td>
<td>.72</td>
</tr>
</tbody>
</table>

### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>1021.44</td>
<td>1</td>
<td>1021.44</td>
<td>185.58</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>65.48</td>
<td>10</td>
<td>6.55</td>
<td>1.18</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Between-array means</td>
<td>1086.89</td>
<td>11</td>
<td>98.81</td>
<td>17.93</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Within arrays</td>
<td>997.68</td>
<td>191</td>
<td>5.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>1083.13</td>
<td>191</td>
<td>5.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2084.57</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Means and standard deviations for the regression of Linguistic scores on timed Vocabulary are given in Figure 10. The high positive correlation is clearly evident in the upward progress of interval mean scores and in the consistency of standard deviations, except for the smaller sigma in the groups with a low N at either end of the distribution.

Since both the Vocabulary and Linguistic tests have been shown to have significant linear relationships to age (.19 and -.22 respectively), it is probable that the coefficient of .70 between the two verbal tests is influenced by heterogeneity of age. According to McNeerar (44, pp. 139-142), when a heterogeneous variable contributes to the correlation coefficient obtained between two measures, the influence of the variable can be eliminated by use of the formula for partial correlation. This represents an estimate of what the correlation between two measures would be if the heterogeneous variable were held constant.

In order to determine the coefficient between timed Vocabulary and Linguistic scores with the influence of age held constant, substitutions were made in the formula for partial correlation (44, p. 141):

\[ r_{12.3} = \frac{r_{12} - r_{15} r_{23}}{\sqrt{1 - r_{15}^2} \sqrt{1 - r_{23}^2}} \]
FIGURE 10


Linguistic Mean Scores
M = 75.50, S.D. = 16.43

Linguistic Raw Scores

Standard Deviations

Midpoints of TIMED Vocabulary Scaled Score Intervals

N = 4 3 13 33 30 30 29 16 21 9 3 2
When the calculations are made, the resulting partial coefficient between the Vocabulary and Linguistic tests is .77.

The Linguistic Scale: Partial Correlations. Since the functions measured by the Marking Test may constitute an heterogeneous variable affecting the relationship between age and Linguistic scores ($r = -.22, P < .01$), the partial correlation method was used to eliminate the effect of the Marking Test. When the correlation between age and the Marking Test ($r = -.15$) and between Linguistic scores and the Marking Test ($r = .09$) are substituted in the formula on page 128, the resulting coefficient between age and Linguistic scores is -.21. Comparing this with the linear coefficient of -.22, it is obvious that the functions measured by the Marking Test have little or no effect on the size of the relationship between age and Linguistic scores.

Age is an heterogeneous variable which may affect the size of the linear coefficient of .08 between the Marking and Linguistic tests. Given the coefficient of -.15 between age and the Marking Test, and -.22 between age and Linguistic scores, the formula for partial correlation yields a coefficient of .08. The conclusion can be made that age has little or no effect on the magnitude of the obtained linear correlation of .08 between the Marking Test and Linguistic scale.
**Age and Timed Arithmetic.** The correlation coefficient, Eta and an analysis of variance for the regression of timed Arithmetic raw scores on age are given in Table XX.

A linear coefficient of -.03 between the two variables was analyzed for significance. Mean squares of 3.68 for the linear regression and 12.64 for residuals from the line yielded an F-ratio of .21. For 1 and 191 df, this relationship would occur by chance more than 50 per cent of the time. Hence the null hypothesis is accepted.

An Eta coefficient of .00 was calculated for age and timed Arithmetic. When the F test for significance was applied, mean squares of 15.74 and 12.48 for between and within-arrays yielded a ratio of 1.26. For 6 and 186 df, a ratio of this magnitude would occur more than 5 times in one hundred by chance. The null hypothesis is accepted for both linearity and curvilinearity of regression.

**Age and Untimed Arithmetic.** Correlations and data for the analysis of variance are given in Table XXI.

A correlation coefficient of -.04 was obtained for age and untimed Arithmetic. When the F test for significance was applied, mean squares of 3.69 for the linear regression and 10.73 for residuals from the line yielded an F-ratio of 34. For 1 and 191 df, a ratio of this size would occur by chance more than fifty per cent of the time and hence is not significant. The null hypothesis is accepted.
### TABLE XX

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of TIMED Arithmetic Raw Scores, 1958 Edition of the ACE, on Age for the Experimental Group ($N = 193$)

<table>
<thead>
<tr>
<th>Age and TIMED Arithmetic</th>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.03</td>
<td>.20</td>
</tr>
</tbody>
</table>

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>2.66</td>
<td>1</td>
<td>2.66</td>
<td>.21</td>
<td></td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>91.81</td>
<td>5</td>
<td>18.36</td>
<td>1.47</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Between-array means</td>
<td>94.47</td>
<td>6</td>
<td>15.74</td>
<td>1.26</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Within arrays</td>
<td>2391.55</td>
<td>186</td>
<td>12.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>2413.36</td>
<td>191</td>
<td>12.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2416.02</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE XXI

Pearson Product-Moment Correlation Coefficient, Correlation Ratio (Eta) and Analysis of Variance for the Regression of UNTIMED Arithmetic Raw Scores, 1948 Edition of the ACE, on Age for the Experimental Group \((N = 193)\)

<table>
<thead>
<tr>
<th>Age and UNTIMED Arithmetic</th>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.04</td>
<td>.16</td>
</tr>
</tbody>
</table>

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>(F)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>3.69</td>
<td>1</td>
<td>3.69</td>
<td>.34</td>
<td></td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>50.90</td>
<td>5</td>
<td>10.18</td>
<td>.94</td>
<td></td>
</tr>
<tr>
<td>Between-array means</td>
<td>54.60</td>
<td>6</td>
<td>9.10</td>
<td>.94</td>
<td></td>
</tr>
<tr>
<td>Within arrays</td>
<td>1997.92</td>
<td>184</td>
<td>10.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>2048.52</td>
<td>191</td>
<td>10.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2052.82</strong></td>
<td><strong>192</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A correlation ratio of .18 was also found to be not significant when mean squares for between and within-arrays of 9.10 and 10.86 respectively, yielded an F-ratio of .94. A relationship of this degree would occur by chance more than fifty per cent of the time.

Means and standard deviations for the regression of timed and untimed Arithmetic scores on age are presented in Figure 11. No remarkable age differences are apparent on either test. Mean scores of age groups cluster closely around the group means of 8.68 for the timed and 18.80 for the untimed Arithmetic. All age groups demonstrate significant gains on the untimed Arithmetic, and gains are remarkably even between age groups. In terms of lowest and highest raw score gains for separate age groups, the range is from 6.19 raw score points for the oldest group to 7.39 raw score points for the age 40-44 subjects, a difference of only 1.20 raw score points.

There is no marked tendency for variability to increase with age on timed and untimed Arithmetic. The four younger groups are relatively more stable when the age groups are compared on the basis of differences in standard deviations for the two tests. The largest S.D. difference in the younger groups is .45, while S.D. differences in the three older groups range from 1.10 to 1.42. Total group S.D.'s are 3.54 for timed and 3.26 for untimed Arithmetic.

It may be concluded that age groups of the credential
FIGURE 11

Mean Scores and Standard Deviations for ACE Arithmetic Raw Scores (Timed and Untimed) 1952 and 1948 Editions, on Age for the Experimental Group (N=193)

Untimed Arithmetic Mean Scores
Mean = 15.80, S.D. = 3.26

Timed Arithmetic Mean Scores
Mean = 8.68, S.D. = 3.54

Standard Deviations

Midpoints of Age Intervals

N = 34 39 33 30 23 18 16
candidate population do not differ significantly with respect to ability in arithmetical reasoning as measured by the present tests, nor do they differ significantly with respect to the range of individual differences within age groups.

**Arithmetic, Timed and Untimed.** Correlations and an analysis of variance for this comparison are given in Table XXII.

The linear coefficient between timed and untimed Arithmetic is .87. When analyzed for significance, an F-ratio of 91.36 was calculated from mean squares of 664.20 for the linear regression and 7.27 for residuals from the line. For 1 and 191 degrees of freedom, a ratio of this magnitude would occur by chance less than one time in one thousand and hence is highly significant. The null hypothesis may be rejected and a linear relationship assumed.

An Eta coefficient of .63 for the two Arithmetic tests was also found to be highly significant when variance estimates of 48.53 and 7.03 for between and within-array means respectively, yielded an F-ratio of 6.87. For 17 and 173 degrees of freedom, this ratio is significant beyond the .001 level of confidence and the null hypothesis may be rejected.

However, when the data were analyzed to discover if the deviation of array means from a straight line was significant, it was found that mean squares of 5.84 for the
TABLE XXII


<table>
<thead>
<tr>
<th></th>
<th>Correlation Coefficient</th>
<th>Correlation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic Timed and</td>
<td>.57</td>
<td>.63</td>
</tr>
<tr>
<td>Arithmetic Untimed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>664.20</td>
<td>1</td>
<td>664.20</td>
<td>91.36</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Deviation of means from line</td>
<td>157.43</td>
<td>16</td>
<td>9.84</td>
<td>1.40</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Between-array means</td>
<td>821.62</td>
<td>17</td>
<td>48.35</td>
<td>6.67</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Within arrays</td>
<td>1330.90</td>
<td>175</td>
<td>7.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual from line</td>
<td>1368.32</td>
<td>191</td>
<td>7.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2052.52</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
deviations of means from the line and 7.03 for the within-array means yielded an F-ratio of 1.40. For 16 and 175 df, a relationship of this degree would occur by chance more than five times in one hundred. Thus the departure from linearity in Table XXII is assumed to be attributable to chance and an hypothesis of significant linear relationship is accepted.

The course of mean scores and standard deviations for the regression of untimed on timed Arithmetic scores is graphically presented in Figure 18. The mean scores assume a somewhat curved aspect, but when the extremes of the distribution, which are represented by small N, are disregarded, the linear relationship is apparent. Standard deviations for intervals are consistent in the middle range where more cases are represented than at the extremes.

Since the timed and untimed Arithmetic tests are alternate forms, a high degree of reliability would ordinarily be expected. The obtained coefficient of .57 is a low estimate of reliability. This is likely to be the result of giving one form with the time limit on which it was standardized and giving the alternate form on a work-limit basis. Removing the time limit restricted the range of talent and resulted in a negative skew. In other words, the test ceiling was not high enough to differentiate the range of talent at the upper end of the distribution for the untimed Arithmetic test.
FIGURE 12


UNTIMED Arithmetic Mean Scores
M = 15.80, S.D. = 3.26

Standard Deviations

TIMED Arithmetic Raw Scores

N = 3 11 18 22 19 20 9 4 2 1 0
This, of course, leads to a lower relationship between the two Arithmetic tests than had the standard time limits been observed in both instances.

**Age and Quantitative Scale.** Correlations and an analysis of variance for the regression of Quantitative scores on age are given in Table XXIII.

The correlation coefficient between the two variables is -.41. Applying the F test for significance, mean squares of 284.92 for the linear regression and 7.89 for residuals from the line yielded a ratio of 39.08. For 1 and 191 degrees of freedom, a ratio of this size would occur by chance less than one time in one thousand. Hence the null hypothesis may be rejected.

A correlation ratio of .43 was also found to be highly significant when tested for curvilinearity. Variance estimates of 51.99 and 7.34 for between and within-arrays respectively, yielded an F-ratio of 7.08. For 6 and 186 df, a relationship of this magnitude would occur by chance less than one time in one thousand.

Since both the linear and curvilinear correlations are highly significant, the data were further analyzed by computing the variance estimates for the deviation of means from the line and within-array means. These were found to be 5.40 and 7.34 respectively, yielding an F-ratio of .74. For 5 and 186 df, a ratio this size would occur more than fifty per cent of the
<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>d.f.</th>
<th>p</th>
</tr>
</thead>
</table>

**Table XIXI**
time by chance. The null hypothesis is retained as regards
curvilinearity and a negative linear hypothesis between age
and the Quantitative scale is assumed.

Mean scores and standard deviations for the regression
of Quantitative raw scores on age are graphically presented
in Figure 15. Decrements in mean Quantitative score with in-
creasing age are more striking than those found between Lin-
guistic scores and age. The two younger groups, 20-24 and
25-29, are almost one-half sigma (S.D.=11.80) above the group
mean of 40.41. The mean for the 30-34 year old subjects is
about one-fifth of a sigma above the group mean, and the
means for all other age groups drop abruptly to a low mean
score of 30.62 for the 45-49 year old subjects, or nearly 1
S.D. below the group mean. The older 50-54 year old group
show a slight increment in mean score.

Although there is a significant inverse relationship
between age and Quantitative mean scores, there is a definite
and consistent tendency for variability to increase with age.
Standard deviations progress from 9.56 for the youngest group
to 12.21 and 12.18 for the two oldest groups. The extent of
individual differences may be expressed in percentages. Thus,
39 per cent of the subjects who are 40 years or older score
above the group mean, and approximately 10 per cent of this
older group score above the mean of the highest scoring young
group. On the other hand, almost 24 per cent of the subjects
FIGURE 13


Quantitative Mean Scores
M = 40.41, S.D. = 11.80

Standard Deviations

N = 34 39 33 30 23 18 16
who are 40 years or older score below the lowest raw scores made by the two younger groups age 20-24 and 25-29.

Quantitative Scale and Timed Arithmetic. At the outset of this experiment, an hypothesis was made that if timed Arithmetic mean scores were found to have a negative relationship with age, one variable effecting decrements in mean score would be differences in speed of recall, or the effects of lack of practice. However, an insignificant correlation coefficient of \(-.05\) was calculated between age and timed Arithmetic. This suggests that speed of recall is not a variable related to age in the credential population, although it may account for individual differences within age groups.

Since the timed Arithmetic test is a subtest of the Quantitative scale which has a negative coefficient of \(-.41\) with age, the relationship of Arithmetic to the total Quantitative scale was examined.

A correlation coefficient of \(.73\) was calculated for timed Arithmetic and the Quantitative scale. Since a spurious correlation arises when a total score is correlated with a subscore which is a part of the total score, the formula for a part-whole correlation given by McNemar (44, p. 139) was used to discover how highly Arithmetic correlated with the Quantitative scale minus the Arithmetic subtest. A part-whole correlation of \(.85\) was obtained, indicating the extent
of the relationship between timed Arithmetic and the composite of the Number Series Completion and Figure Analogies subtests which compose the remainder of the Quantitative scale.

The Quantitative Scale: Partial Correlations. It has already been observed that there is no significant relationship between age and timed Arithmetic scores \( r = -.03 \). Hence, a logical conclusion is that timed Arithmetic scores, which are a part of the total Quantitative scale, do not contribute significantly to the high negative relationship found between age and Quantitative scores \( r = -.41 \). A second logical conclusion which may be derived from the data is that the remaining two subtests which compose the Quantitative scale (Number Series Completion and Figure Analogies) are the principal source of the negative correlation. Apparently there is a significant tendency for Number Series Completion and/or Figure Analogies to decline with age.

The functions measured by the Marking Test have already been observed to have a significant negative relationship to age \( r = -.15 \). Using the partial correlation method referred to on page 128 of this study, the influence of the Marking Test was eliminated from the obtained coefficient of -.41 between age and Quantitative scores. The linear coefficients between age and the Marking Test and between Quantitative scores and the Marking Test are -.15 and .02 respectively. When these are substituted in the partial correlation formula,
the resulting coefficient is \(-.41\). Since the correlation between age and Quantitative scores is not altered when the Marking Test is held constant, it may be concluded that the functions measured by the Marking Test are not a source of differing test performance on the Number Series and Figure Analogies subtests of the Quantitative scale.

Using the partial correlation procedure to eliminate the effects of age on the correlation of \(0.02\) between the Marking Test and Quantitative scale, where the coefficient between the Marking Test and age is \(-.15\), and between Quantitative scores and age is \(-.41\), the obtained partial correlation between age and Quantitative scores is \(-.05\). This is evidence that age has little or no effect on the chance relationship of \(0.02\) between the Marking and Quantitative tests.

**Vocabulary Groups Within the Experimental Population.**

An observation which initiated this study was that many older students achieve Vocabulary scores on the Cooperative English Examination (CEE) which are markedly high in relation to their own average scores on other tests of the ACE and CEE battery. To discover how students with the above relationship differ with respect to the remaining students of the experimental population, the 195 subjects were divided into two groups.

These students whose Vocabulary scores are nine scaled score points or more above their total CEE scaled scores (to
which Vocabulary contributes a part) were separated into one group which will be referred to as the out-of-pattern Vocabulary group. The cut-off point of 9 scaled score units was arbitrarily selected for no other reason than that it would be an apparent difference between Vocabulary and total CEE score.

The second group of subjects, 110 in number, constitute what will henceforth be referred to as the residual group. Subjects in this group are those whose Vocabulary scores did not differ by more than 8 scaled score points from their own total CEE scaled score.

The means for each vocabulary group and the total group were computed in relation to age, the Marking Test, timed Vocabulary, timed and untimed Arithmetic, and the Quantitative Scale. The results of this comparison are given in Table XXIV.

The out-of-pattern high-scoring Vocabulary group has a mean age of 40.15 and the residual group a mean age of 30.85, a difference of over 9 years. The histograms in Figure 14 show the distributions for the two groups. Of the 73 subjects in the two lower age groups of 20-24 and 25-29, only 19 per cent are in the out-of-pattern group. Of the 34 subjects who are 45 years or older, 82 per cent are in this group. This confirms the observation that more older students than young students have high Vocabulary scores relative to their own average test performance.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Group (N=193)</th>
<th>High Vocabulary Group (N = 83)</th>
<th>Residual Group (N=110)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>34.55</td>
<td>40.15</td>
<td>30.85</td>
</tr>
<tr>
<td>Marking Test</td>
<td>176.22</td>
<td>172.10</td>
<td>182.90</td>
</tr>
<tr>
<td>Vocabulary, Timed</td>
<td>65.95</td>
<td>72.30</td>
<td>63.90</td>
</tr>
<tr>
<td>Linguistic Scale</td>
<td>75.55</td>
<td>74.30</td>
<td>76.50</td>
</tr>
<tr>
<td>Arithmetic, Timed</td>
<td>8.87</td>
<td>7.83</td>
<td>9.65</td>
</tr>
<tr>
<td>Arithmetic, Untimed</td>
<td>15.81</td>
<td>15.40</td>
<td>16.15</td>
</tr>
<tr>
<td>Quantitative Scale</td>
<td>40.46</td>
<td>35.22</td>
<td>44.48</td>
</tr>
</tbody>
</table>
FIGURE 14

Histogram comparing the Out-of-Pattern High Vocabulary Group (N = 83) with the Residual Group (N = 110) on Age for the Experimental Group

Residual Vocabulary Group
M = 30.85

Out-of-Pattern High Vocabulary Group
M = 40.15
On the Marking Test, the out-of-pattern Vocabulary group has a mean score of 172.10 as compared to a mean score of 182.90 for the residual group. Although the out-of-pattern group appears to be slower on the Marking Test, the difference is small, being approximately one-fourth of a total group S.D. of 37.03.

When the two Vocabulary groups are distributed on timed Vocabulary scores, the out-of-pattern subjects have a mean score of 72.30 as compared to 65.90 for the residual group. Since Vocabulary tends to increase with age, and since the out-of-pattern group has a higher mean age than the residual group, this difference in mean scores on timed Vocabulary is to be expected.

Little difference is found between mean scores on the Linguistic scale, the out-of-pattern group having a mean score of 74.30 and the residual group having a mean score of 76.50.

Mean scores for timed Arithmetic are 7.83 for the out-of-pattern group and 9.65 for the residual group. On untimed Arithmetic, the mean scores are 15.40 and 16.15 for the out-of-pattern and residual groups respectively.

Histograms for the distribution of the vocabulary groups on the Quantitative scale are given in Figure 16. The mean Quantitative score for the out-of-pattern group is 35.22 and for the residual group 44.49, a difference of approximately
FIGURE 15

Histogram comparing the Out-of-Pattern High Vocabulary Group (N = 83) with the Residual Group (N = 110) on the Quantitative Scale for the Experimental Group.

Residual Vocabulary Group
M = 44.48

Out-of-Pattern High Vocabulary Group
M = 35.22
four-fifths of the total group S.D. of 11.80. The low score intervals of the Quantitative Scale have 44 subjects of whom 56, or approximately 80 per cent, are in the out-of-pattern Vocabulary group. In the five high score intervals of the Quantitative scale, only 17 of 66 subjects, or approximately 26 per cent, are in the out-of-pattern group. This confirms the observation that more older students than young students have low Quantitative scores relative to their Vocabulary scores.

In sum, approximately 43 per cent of the experimental subjects show an apparent tendency to have high Vocabulary scores relative to their own average performance on the CEE. This group may be clearly differentiated from the remaining subjects of the experiment with respect to having an average age which is over 9 years above that of the remaining subjects where the total group S.D. for age is 9.56, and with respect to having an average Quantitative score which is more than 9 points below that of the remaining subjects where the total group S.D. is 11.80.

Comparison of the Preliminary and Experimental Groups.

The preliminary study of approximately 600 credential candidates may be compared to the experimental study of 193 subjects in respect to the three tests common to both. These are the CEE Vocabulary subtest, and the ACE Linguistic and Quantitative scales. The two groups may also be compared with
respect to their relative standing when mean test scores for the age groups are converted to the percentile and decile norms which are used for the credential population.

The preliminary group is inclusive of an age range from 18 to 64, whereas the experimental group is restricted to a range of 20 to 54 years. However, eight of the subjects in the 50-54 age group are actually 55 or over. In the preliminary study, the younger age groups have a far larger number of cases than in the experiment. In both studies, the older age groups have less representation than the younger groups.

**GEE Timed Vocabulary.** A graphic comparison of the means and standard deviations of the timed Vocabulary scores on age for the preliminary and experimental groups is given in Figure 16.

In the preliminary study, a linear coefficient of .37 and an Eta coefficient of .41 were calculated between age and Vocabulary. When these were analyzed for significance by means of the F test, it was determined that the Eta coefficient was significant at greater than the .001 level of confidence ($P < .001$), and an hypothesis of curvilinearity was assumed.

For the experimental group it was determined that the linear coefficient of .19 adequately described the data and the hypothesis of linearity was retained.
Comparison of Means and Standard Deviations for Timed CEE Vocabulary Subtest Scaled Scores, Form Z, on Age for the Preliminary Group (N=599) and the Experimental Group (N=193)
Inspection of Figure 16 suggested that the linear relationship found for the experimental group may be partly a function of the restricted number of age intervals. The largest difference in mean score is found in the 25-29 age interval where the subjects of the experiment surpassed the preliminary group in mean score by approximately 5 raw score points. However, the number of cases in the high scoring group is only one-fourth the number of cases in the preliminary study. In both studies, the younger subjects, through age 29, have lower mean scores than subjects above 30 years of age.

**ACE Linguistic Scale.** A graphic comparison of the means and standard deviations of the Linguistic scores on age for the preliminary and experimental groups is given in Figure 17.

In the preliminary study, a linear coefficient of -.07 and an Eta coefficient of .81 were calculated for age and Linguistic scores. The F test determined that the Eta coefficient was significant at greater than the .001 level of confidence ($P < .001$) and an hypothesis of curvilinearity was assumed.

For the experimental group it was determined that the linear coefficient of -.22 described the data adequately.

Inspection of Figure 17 suggests that, as in the case of Vocabulary, the linear aspect of the experimental results may be a function of restricted range. The 20-24 year old
FIGURE 17


Linguistic Mean Scores
- Preliminary Group
  M = 74.38, S.D. = 16.91
- Experimental Group
  M = 75.50, S.D. = 16.42

Standard Deviations

Midpoints of Age Intervals

$N_P = 39 \quad 186 \quad 129 \quad 69 \quad 58 \quad 44 \quad 38 \quad 14 \quad 11 \quad 4$

$N_E = 34 \quad 39 \quad 33 \quad 30 \quad 23 \quad 18 \quad 16$
subjects of the experiment again surpass the age 30-24 sub-
jects of the preliminary study, and the slope of the line is
flatter for the experimental subjects from the youngest
through the 35-39 age interval. Nevertheless, a negative
linear relationship with increasing age may be observed for
the subjects of both studies beginning at about age 40.
There is an observable but inconsistent tendency for vari-
ability to increase with age.

ACE Quantitative Scale. A graphic comparison of the
means and standard deviations for the Quantitative scores on
age for the preliminary and experimental studies is given in
Figure 18.

In the preliminary study it was determined that the
Eta coefficient of .44 more adequately described the data
and an hypothesis of curvilinearity was assumed.

For the experimental study, it was found that the hy-
pothesis of linearity could be retained.

Inspection of Figure 18 suggests that subjects of the
experiment are somewhat superior to subjects of the prelimin-
ary study for all age ranges. For both populations there is
a significant and inverse relationship between age and Quan-
titative scores beginning at about age 35.
FIGURE 18

Comparison of Means and Standard Deviations for ACE Quantitative Raw Scores, 1952 Edition on Age
Age for the Preliminary Group (N=586) and the Experimental Group (N=193)

Quantitative Mean Scores
- Preliminary Group
  M = 40.23, S.D. = 11.52
- Experimental Group
  M = 40.41, S.D. = 11.80

Standard Deviations

Midpoints of Age Intervals

$N_p = 39 \quad 189 \quad 120 \quad 67 \quad 58 \quad 46 \quad 38 \quad 14 \quad 11 \quad 4$

$N_e = 34 \quad 39 \quad 33 \quad 30 \quad 23 \quad 18 \quad 16$
Percentile and Decile Comparisons between Age Groups and the Vocabulary, Linguistic and Quantitative Tests. Test evaluation of credential candidates is based on percentile and decile norms which have been established without regard for heterogeneity of age. The number of younger students in the credential population far surpasses the number of older adults, hence the test performance of college age students is given proportionately far more influence in the determination of percentile and decile scores for the total group. Since the norms are in large degree determined by the performance of young adults, any variables related to increases in age which may influence test-taking ability are not considered in the evaluation of older adults. That changes in test-taking ability do take place with changes in age level is apparent from an inspection of Table XXV. The table lists the mean percentile and decile scores for age groups on the Vocabulary, Linguistic and Quantitative tests.

In Part I of Table XXV, the mean scaled scores for age groups on the Vocabulary test have been converted to crude interpolated percentile scores based on CEE national norms for juniors in junior colleges and teacher colleges. It is evident that the mean percentile scores of all age groups, excepting the 19 and less, surpass the 50th percentile on the national norms. Mean Vocabulary percentile scores show a marked increase with age and the oldest group of subjects in
### Table XXV

Mean Percentile and Decile Ranks on the CEE Timed Vocabulary Form Z, and the ACE Linguistic and Quantitative Scales, 1956 Edition, for Age Groups of the Preliminary and Experimental Studies

<table>
<thead>
<tr>
<th>AGE</th>
<th>Ia Preliminary</th>
<th>Ila Preliminary</th>
<th>IIIa Preliminary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CEE Vocabulary</td>
<td>ACE Linguistic</td>
<td>ACE Quantitative</td>
</tr>
<tr>
<td>60 - 64</td>
<td>99</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>55 - 59</td>
<td>86</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>50 - 54</td>
<td>81 85</td>
<td>30 30</td>
<td>10 20</td>
</tr>
<tr>
<td>45 - 49</td>
<td>91 85</td>
<td>40 40</td>
<td>10 20</td>
</tr>
<tr>
<td>40 - 44</td>
<td>92 89</td>
<td>50 40</td>
<td>30 30</td>
</tr>
<tr>
<td>35 - 39</td>
<td>86 88</td>
<td>50 60</td>
<td>30 40</td>
</tr>
<tr>
<td>30 - 34</td>
<td>86 83</td>
<td>60 60</td>
<td>50 50</td>
</tr>
<tr>
<td>25 - 29</td>
<td>75 71</td>
<td>60 60</td>
<td>50 60</td>
</tr>
<tr>
<td>20 - 24</td>
<td>56 75</td>
<td>40 60</td>
<td>50 60</td>
</tr>
<tr>
<td>19 &amp; less</td>
<td>49</td>
<td>40</td>
<td>50</td>
</tr>
</tbody>
</table>

**Note:** Crude interpolated percentile scores based on C.E.E. National Norms for Juniors in junior colleges and teacher colleges.

**Note II and III:** Decile norms based on credential candidates at Sacramento State College in the Fall of 1954.
the preliminary study (age 50-64) are at the 99th percentile. The preliminary and experimental groups are comparable except in the 20-24 age range where there is a difference of 19 percentile points between the groups in favor of the subjects of the experiment.

In part II of Table XXV, the mean raw scores for age groups on the Linguistic test have been converted to decile scores based on credential candidates at Sacramento State College in the Fall of 1954. Linguistic scores for the two younger groups of the preliminary study, 19 and less and 20-24, are in the 4th decile range; the 25-29 and 30-34 age groups are in the 6th decile; the 35-39 and 40-44 groups are in the 5th, the 45-49 group in the 4th, the 50-54 group in the 3rd and the two older groups, age 55-59 and 60-64, are in the 2nd decile range. In the experimental study, the age groups 20-24 through 35-39 maintain a position in the 6th decile. This drops to the 4th decile for the 40-44 and 45-49 subjects and to the 3rd decile for the 50 and over subjects.

In part III of Table XXV, the mean Quantitative scores have been converted to deciles based on the same population as the Linguistic norms. The drop in mean Quantitative decile rank with increments in age is clearly apparent. In the preliminary study, the age intervals from 19 and less through the 30-35 age range are in the 5th decile. This decreases to a mid-point in the 1st decile for the oldest group.
age 60-64. In the experimental study, deciles decrease from the 6th for the 20-24 and 25-29 subjects to the 2nd decile for the age 50 and above group. The subjects of the experiment consistently tend to surpass the subjects of the preliminary study, but the decrements in decile rank follow a similar course.

Differences between age groups on the same test may be compared in terms of the expectancy that average test performance will fall at the 50th percentile. On the Vocabulary test, only the two younger age groups of the preliminary study meet this expectancy with mean percentile scores of 48 and 56 for the age 19 and less and 20-24 groups respectively. In general, all other age groups have increasingly high percentile scores. The total group percentiles of 75 and 81 for the subjects of both studies on Vocabulary are well above the group expectancy for a college population of juniors in junior and teacher colleges.

The departure from expectancy on the Linguistic scale is not marked until age 50 and above. Decile scores vary from 40 to 60 in the groups under 50. Beyond this age the mean deciles depart more radically from expectancy in a negative direction. Total group decile means are 60 and 50 for the preliminary and experimental studies respectively.

On the Quantitative scale, the group expectancy that decile scores will cluster around 50 is maintained for the 19
and less through the 30-34 age interval. Beyond this age there is a consistent and marked tendency for decile scores to decrease with increasing age.

In sum, Vocabulary percentiles show a consistent increase from the youngest through the oldest students. Linguistic mean deciles cluster around the mean group expectancy until about age 50 before there is a marked drop in score. Quantitative mean decile scores begin to drop below expectancy beginning with age 35.

The preceding discussion has pointed out that there are striking differences in percentile and decile ranks between age groups within tests. Table XXV, page 160, affords a comparison between same age groups on different tests that is even more striking. When Linguistic and Quantitative ranks for separate age groups are compared with Vocabulary, it is obvious that the younger the age group being considered, the less difference there is between percentile and decile scores on the different tests. Conversely, the older the age group, the larger the difference between Vocabulary rank and the Linguistic and Quantitative ranks. The differences between tests for separate age groups tend to be consistently less for Vocabulary and Linguistic comparisons than for Vocabulary and Quantitative comparisons. When Linguistic and Quantitative ranks are compared for separate age groups, the differences are not nearly so striking either between tests or between age groups.
The review of the literature has established that vocabulary is a relatively stable criterion of the highest level of intellectual ability achieved by an individual during his life span. If one accepts this conclusion without qualification, it appears that in the credential population, the older age groups are increasingly highly selected with respect to intelligence.

However, if one assumes that the Linguistic and Quantitative scales measure the same variables for all age groups, namely scholastic and numerical ability, then one must conclude that the older subjects of the credential population have suffered a profound loss of efficiency in these respects. Even if allowance is made for increasing vocabulary level with age to the extent that total group homogeneity with respect to intellectual level is assumed, the losses of efficiency in Linguistic and Quantitative skills are still of such degree that disproportionately more mature adults than younger students risk Credential Committee disapproval or referral for re-evaluation on the basis of test results.

The Linguistic and Quantitative scales have content similar to group tests of intelligence which generally have been shown to decline with age. In the review of the literature, it was established that the progress of test scores with age is a function of the nature of the test, of administrative procedures, of physiological variables related to
intake of stimuli and output of response. Little is known about the nature of the mental reaction itself or about changes in reorganization of mental functions with age or of the effects of use and disuse on skills over a span of time. Until more is known about these variables, the investigator may resort to outside criteria of success to evaluate the effectiveness of test measurements of mature adults, meanwhile maintaining a healthy skepticism with respect to the adequacy of using tests designed for homogeneous age groups on heterogeneous age groups.
CHAPTER V

SUMMARY, CONCLUSIONS AND DISCUSSION

The recent legislation pertaining to the renewal of provisional teacher credentials granted after July 1, 1964 resulted in a marked increase in the number of older students who must take screening tests in scholastic ability and English achievement at Sacramento State College.

Over a period of time the observation was made that the performance of many older students differs from that of younger students on the separate subtests of the screening battery. Since the tests are one of several criteria upon which the approval of a student's application for candidacy is based, the decision was made to conduct a study of age differences in test performance. The purpose of the study was to discover the extent and nature of the differences in test performance between age groups, to evaluate the tests and normative standards in present use for their adequacy in measuring skills over a wide age range, and to explore the literature with respect to the findings and conclusions of other investigators in the field.

Summary of the Preliminary Study. That part of the screening battery which is designed to measure scholastic ability and English achievement consists of the nationally known American Council on Education Psychological Examination
for Freshmen (ACE) and the Cooperative English Examination, Higher Level (CEE). In the preliminary study which preceded the experiment, the test data for approximately 600 credential candidates ranging in age from 18 through 64 were examined. Scores on the Vocabulary subtest of the CEE, and the Quantitative and Linguistic sections of the ACE were compared with age by computing Pearson product-moment correlation coefficients and correlation ratios (Ebs). The obtained coefficients were analysed for significance by means of Snedecor's F test, a non-logarithmic version of Fisher's z distribution. Means and standard deviations were computed for the total population and for separate age groups within the population when distributed in intervals of five years.

In each comparison, the relationship of age to test scores was found to be significantly curvilinear, and the age groups, with a mean age of 30.14, were found to differ markedly in their performance on each test.

For the comparison of age with Vocabulary scores, an Eta coefficient of .41 was found to be significant beyond the .001 level of confidence. Vocabulary level increases significantly with increasing age in the credential candidate population. The mean Vocabulary level of the oldest subjects, age 60 through 64, is almost two standard deviations above that of the age 19 and less subjects.

An Eta coefficient of .27 between age and Linguistic
scores was also significant beyond the .001 level of confidence. There are increments in mean score beginning with
the youngest age group of 19 years and less through the age
interval of 30 to 34 years. Following this peak performance,
there is a marked decline in mean Linguistic scores for each
successive age level and a sharp differentiation between the
three older groups of subjects and the rest of the population.
The mean test scores of age intervals from 50 through 64 are
more than one total group standard deviation below the high
scoring 30 through 34 year old subjects.

For age and Quantitative scores, an $r$ coefficient
of .44 was found to be significant beyond the .001 level of
confidence. Mean scores for the younger subjects, age 18
through 34, are on a plateau above the group mean. Following
this, there is a pronounced decrement in mean score for suc-
cessively older age groups. The 60 through 64 year old sub-
jects have a mean Quantitative score which is more than two
standard deviations below the plateau of the younger group of
subjects.

The preliminary study established that age groups of
the credential population differ markedly with respect to
their performance within each test and also that the older the
age group the more marked is the difference in mean scores be-
tween tests. The older age groups scored highest on Vocabulary,
but received the lowest scores on the Linguistic and Quantita-
tive scales.
Summary of the Experiment. Following this study, an experiment was designed to investigate further the relationships between age and tests of the credential candidate screening battery. The Vocabulary, Linguistic and Quantitative tests were again analyzed with respect to age. Three additions were made to the test battery: an alternate form of the CEE Vocabulary subtest, an alternate form of the ACE Arithmetic subtest which is part of the Quantitative scale, and a Marking Test designed by the author.

The Linguistic and Quantitative tests are rigorously timed. Tests for children and young adults are constructed on the assumption that speed and intelligence are highly correlated, but there is no assurance that this parallel exists for mature adults. A number of investigators have established that adults tend to make relatively higher scores on untimed or power-type tasks than on timed tests.

To evaluate the adequacy of the Linguistic and Quantitative scales for measuring the intellectual level of mature adults, it was necessary to establish a simple criterion of intelligence with which to compare the results of these two tests. The review of the literature has established that vocabulary tests are a relatively stable measure of the highest intellectual level attained by an adult throughout his life span. However, several studies present evidence that vocabulary tends to increase with age for persons of above average
intelligence. Unless adjustments are made for age, there is
the risk that the intellectual level of adults will be over-
estimated.

In view of the above findings, the CEE Vocabulary sub-
test may serve as a criterion for evaluating the effectiveness
with which other tests fulfill the purpose for which they are
used. Since the Linguistic and Quantitative scales were de-
signed for and standardized on college age freshmen, they may
not be adequate measures of ability over a wide age range.

The CEE Vocabulary, as it is regularly administered,
has a liberal time limit which apparently permits the majority
of students to reach their level of competency. However, to
verify this hypothesis, the alternate form of the Vocabulary
test was given to the experimental group without time restric-
tion.

The untimed Arithmetic test was included to discover
what differences would occur between timed and untimed ver-
sions with respect to the total population and age groups
within the population. Among variables which may affect per-
formance on the timed subtest are age, speed of recall, ini-
tial test anxiety, and variables related to the use of an IBM
answer sheet.

The Marking Test, given with a one-minute time limit,
was designed to isolate the complex of implicit functions
which are required in using an IBM answer sheet in a timed
test situation. Directions for taking the test were mimeographed on the back of standard IBM answer sheets and paralleled the general instructions for the ACE which emphasize both speed and precision. This dual emphasis is subject to individual differences in interpretation. The assumption was made that speed of response, tendencies to precision, and interpretation of instructions will not singly determine how individuals will perform, but rather individual interaction of these variables.

During the summer session of 1955, the screening tests with the described additions were administered to 195 credential candidates ranging in age from 20 to over 54 and with a mean age of 34.25.

The administrative procedure was similar in every respect to that of previous summer sessions, with the exception that the personality inventory was given in the short form to compensate for additional time required to effect the experiment.

The data were analyzed in the same manner as for the preliminary study. A summary of findings is presented in Table XXVI.

The Marking Test and age have a negative linear relationship of -.18 which is significant beyond the .05 level of confidence. However, the standard deviations for age groups are sufficiently large to lead to the conclusion that individual
### TABLE XXVI

Summary of Results for the Experimental Study: Pearson Product-Moment Correlation Coefficients and Eta Coefficients of Test Comparisons with the Marking Test and with Age for the Experimental Group \((N = 195)\)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>(r)</th>
<th>Eta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marking Test &amp; Age</td>
<td>-.15*</td>
<td>.15</td>
</tr>
<tr>
<td>Marking Test &amp; timed Vocabulary</td>
<td>-.07</td>
<td>.18</td>
</tr>
<tr>
<td>Marking Test &amp; untimed Vocabulary</td>
<td>-.09</td>
<td>.24</td>
</tr>
<tr>
<td>Marking Test &amp; Linguistic Scale</td>
<td>.08</td>
<td>.14</td>
</tr>
<tr>
<td>Marking Test &amp; timed Arithmetic</td>
<td>-.11</td>
<td>.20</td>
</tr>
<tr>
<td>Marking Test &amp; untimed Arithmetic</td>
<td>-.007</td>
<td>.24</td>
</tr>
<tr>
<td>Marking Test &amp; Quantitative Scale</td>
<td>.02</td>
<td>.18</td>
</tr>
<tr>
<td>Age &amp; timed Vocabulary</td>
<td>.19**</td>
<td>.23</td>
</tr>
<tr>
<td>Age &amp; untimed Vocabulary</td>
<td>.15*</td>
<td>.24</td>
</tr>
<tr>
<td>Vocabulary timed &amp; untimed</td>
<td>.84###</td>
<td>.84</td>
</tr>
<tr>
<td>Age &amp; Linguistic Scale</td>
<td>-.22**</td>
<td>.25</td>
</tr>
<tr>
<td>Linguistic Scale &amp; timed Vocabulary</td>
<td>.70###</td>
<td>.72</td>
</tr>
<tr>
<td>Age &amp; timed Arithmetic</td>
<td>-.03</td>
<td>.20</td>
</tr>
<tr>
<td>Age &amp; untimed Arithmetic</td>
<td>-.04</td>
<td>.16</td>
</tr>
<tr>
<td>Arithmetic timed &amp; untimed</td>
<td>.57###</td>
<td>.63</td>
</tr>
<tr>
<td>Age &amp; Quantitative Scale</td>
<td>-.41###</td>
<td>.45</td>
</tr>
</tbody>
</table>

* indicates relationship to be significant at the .05 level of confidence
** indicates relationship to be significant at the .01 level of confidence
### indicates relationship to be significant at the .001 level of confidence
differences within age groups are greater than differences between age groups.

There is no significant relationship between the Marking Test and any of the verbal and numerical tests which were studied. In all instances, the obtained relationships are not significant at the .05 level of confidence. For only one small group of four subjects is there an apparent age and Marking Test relationship which affects test performance. These subjects were the slowest on the Marking Test, had the highest mean age (42.25), were the lowest on the Quantitative and Linguistic scales, but made more than average gains on untimed Arithmetic and scored highest on both timed and untimed Vocabulary.

The Marking Test does not differentiate levels of performance on timed or untimed Arithmetic, either within the tests or between the tests. Gains on untimed Arithmetic, in general, are consistent between the slowest through the fastest subjects on the Marking Test.

Age and timed Vocabulary have a positive linear relationship of .19 which is significant beyond the .01 level of confidence. Age and untimed Vocabulary also have a positive linear relationship of .16 which is significant beyond the .05 level of confidence. Vocabulary level tends to increase with increasing age and the small gains on the untimed test tend to be consistent between age groups. Individual differences (variability) within age groups are relatively consistent
between age groups. A correlation of .64 between the two forms of the Vocabulary test is significant above the .001 level of confidence, indicating a high degree of reliability. The above relationships support the hypothesis that the timed Vocabulary test is equivalent to a power-type task.

Age and Linguistic scores have a negative linear relationship of -.32 which is significant beyond the .01 level of confidence. Mean Linguistic scores are on a plateau for the age 19 and less subjects through the 35-39 interval and are above the group mean. Successively older groups decrease in mean score to a low for the two older groups, age 45-49 and 50-54, which is more than one-half the total group standard deviation below the high-scoring younger groups. There is a trend, with inconsistencies, for variability to increase with age.

The Linguistic scale and timed Vocabulary have a linear relationship of .70 which is significant beyond the .001 level of confidence. When the effect of heterogeneity of age is removed by means of the partial correlation method, this is increased to .77.

By means of the partial correlation method, it was determined that the Marking Test does not act as an heterogeneous variable on the relationship between age and Linguistic scores, nor does age affect the size of the linear correlation between the Marking Test and Linguistic scores.
There is no significant relationship between age and timed or untimed Arithmetic. The correlation coefficients are -.03 and -.04 respectively. All age groups demonstrate significant gains on the untimed Arithmetic and gains are remarkably consistent between age groups.

A linear correlation of .57 between timed and untimed Arithmetic is significant beyond the .001 level of confidence. However, this is a low estimate of reliability which probably results from the difference in timing of the two tests. For the untimed Arithmetic, the test ceiling was not high enough to differentiate the range of talent at the upper end of the distribution.

Age and the Quantitative scale have a negative linear relationship of -.41 which is significant beyond the .001 level of confidence. The two younger groups, age 19 and less and 20-24, are almost one-half a standard deviation above the group mean. The mean for the 30-34 year olds is about one-fifth of a standard deviation above the group mean, and the means for all other age groups show a negative slope to a mean score for the 45-49 year old which is nearly 1 standard deviation below the group mean. There seems to be a definite and consistent trend for variability to increase with increasing age. Standard deviations progress from 5.56 for the youngest group to 12.51 and 12.18 for the two oldest groups. Thirty nine percent of the subjects who are 40 years or older score
above the group mean, and approximately 10 per cent of this older group score above the mean of the highest scoring young group. On the other hand, almost 24 per cent of the subjects who are 40 years or older score below the lowest raw scores made by the two younger groups, age 20-24 and 25-29.

When the relationship of the Arithmetic subtest to the total Quantitative scale was examined, a linear coefficient of .73 was obtained. A spurious correlation arises when a total score is correlated with a subscore which is a part of the total score, hence the formula for a part-whole correlation was used to discover how highly Arithmetic correlates with the Quantitative test minus the Arithmetic subtest. A part-whole coefficient of .53 was calculated between timed Arithmetic and the composite of the Number Series Completion and Figure Analogies subtests which compose the remainder of the Quantitative scale. Since there is a significant linear relationship of -.41 between age and the Quantitative scale and a -.03 correlation between timed Arithmetic and age, a logical conclusion is that there is a significant tendency for Number Series Completion and Figure Analogies to decline with age.

The partial correlation method determined that heterogeneity of the Marking Test does not affect the size of the relationship between age and Quantitative scores, nor does the heterogeneity of age affect the size of the relationship
between the Marking Test and the Quantitative scale.

The experimental group was divided into two groups, one composed of 83 subjects whose Vocabulary scaled scores were 9 scaled score units or more above their own average total CEE score, and a residual group whose Vocabulary scores did not differ by more than 8 scaled score points from their own total CEE scaled score. The purpose of this division was to discover whether subjects who have outstanding Vocabulary scores relative to their performance on other tests differ in other respects. Approximately 43 per cent of the experimental population show an apparent tendency to have high Vocabulary scores relative to their own average performance on the total CEE. This out-of-pattern group may be clearly differentiated from the remaining subjects of the experiment on the basis of an average age of 40.15 which is over 9 years above the mean age of 30.65 for the remaining subjects where the total group S.D. for age is 9.36. The out-of-pattern Vocabulary group is also clearly differentiated from the remainder of the group on the basis of an average Quantitative score of 35.22 which is more than 9 points below the mean score of 44.48 for the residual group where the total group S.D. is 11.80.

The preliminary and experimental groups were compared with respect to their mean performance on the tests common to both. The Vocabulary, Linguistic and Quantitative scores follow essentially the same course in both studies, but have
somewhat less marked slopes in the experimental group. The
youngest age group in the experiment, age 20-24, consistently
surpasses the same age group in the preliminary study on all
three tests. These two factors, plus the more restricted age
range of the experimental population, probably account for the
significance of linearity of the experimental data as compared
to the highly significant curvilinear relationships between
age and test scores in the preliminary study.

Mean scores for the three tests were converted to the
percentile and decile norms used for the evaluation of creden-
tial candidates. Differences between age groups were com-
pared on the basis of the expectancy that average test per-
formance will fall at the 50th percentile.

On the Vocabulary test, only the two younger age groups
of the preliminary study meet this expectancy with mean per-
centile scores of 48 and 56. In general, all other age groups
have increasingly high percentile scores, reaching a maximum
of 99 for the 60-64 year old subjects. The total group per-
centiles of 75 and 81 for the subjects of both studies are
well above the group expectancy for a college population of
juniors in junior and teacher colleges.

The departure from expectancy on the Linguistic scale
is not marked until age 50 and above. Decile scores vary
from 40 to 60 in the age groups under 50, and proceed from 50
to 80 in the over 50 groups. Total mean deciles are 60 and
50 for the preliminary and experimental studies respectively.

For the Quantitative scale, the group expectancy that decile scores will cluster around 50 is maintained relatively well for the 19 and less through the 30-34 interval. Beyond this age there is a consistent and marked tendency for decile scores to decrease with increasing age. The oldest group of subjects, age 60-64 drop to a midway point in the first decile.

Comparisons between the same age groups on different tests are even more striking. When Linguistic and Quantitative ranks for the separate age groups are compared with Vocabulary, it is obvious that the younger the group being considered, the less difference there is between percentile and decile scores on the different tests. Conversely, the older the age group, the larger the difference between Vocabulary rank and Quantitative and Linguistic ranks. Differences between Vocabulary and Linguistic rank tend to be consistently less than differences between Vocabulary and Quantitative rank for the older age groups. When Linguistic and Quantitative ranks are compared for the separate age groups, the differences are not nearly so striking either between tests or between age groups.

Conclusions. The Vocabulary subtest of the Cooperative English Examination used in this study is relatively equivalent to a power-type task and may be used as a fairly stable index of intellectual level for all age groups.
However, the credential candidate population has demonstrated that with increments in age there are increments in vocabulary level. One cannot state, of course, that this is a function of age, since factors may be operating whereby persons at the higher age levels are more highly selected than those at the younger age levels. This is a limitation of the cross-sectional type of age study where different age groups are represented by different individuals and where there is no assurance of initial homogeneity of ability. However, since there is evidence in the review of the literature that vocabulary level tends to increase with age, it is reasonable to assume that the age groups are relatively homogeneous with respect to initial intelligence and to make adjustments for increasing vocabulary level with age. Failure to make an adjustment is to take the risk of over-estimating intelligence of the more mature adults. This is especially relevant if vocabulary is to be accepted as a criterion of the highest intellectual level attained by an individual during his life span, and if vocabulary is to serve as a basis of comparison with other test performance.

There is a significant negative linear relationship between the Marking Test and increments in age. However, correlations of this test with tests of the credential candidate screening battery are of such low magnitude that the conclusion must be made that whatever variables are measured
by the Marking Test do not perceptibly affect test scores. Significant negative relationships were found between age and Linguistic scores and between age and Quantitative scores, but when the effects of the Marking test were cancelled out by means of the partial correlation method, the coefficients of relationship were relatively unchanged. This confirms the conclusion that the Marking Test is apparently unrelated to performance on the timed tests of the screening battery.

It may be argued that differences in motivation between taking the Marking Test and other candidate screening tests were sufficient to vitiate the results of the Marking Test. However, motivation is generally of a high level for this group of students since they have much at stake with respect to their vocational plans.

Age groups of the credential population do not differ significantly with respect to arithmetical reasoning ability as measured by the timed and untimed Arithmetic tests used in the experiment, nor do they differ significantly with respect to the range of individual differences within age groups. Since gains made on the untimed version of Arithmetic were comparable for all age groups, it may be concluded that there are no test-taking variables which differentiate the age groups and no differences in speed of the mental reaction which might be related to lack of practice. Arithmetical
skills and availability of these for immediate use appear to be comparable between age groups. The credential candidate population may be selected in this respect. Since many of the subjects are experienced teachers or have been engaged in a college training program, recency of experience with arithmetical procedures may not be a source of difference between age groups as has been found in the populations of other studies. The present results are comparable to those of other investigators who have found that Arithmetic occupies an intermediary position between increments in verbal skills on the one hand and decrements in non-verbal type tasks on the other hand.

Performance on the Linguistic test of the ACE is relatively stable until about age 50. Nevertheless, within this age range there may be both increments in score from the youngest through about age 35, and decrements beyond this point to age 50 which, however, do not fall below the mean of the youngest group. Beginning at about age 50, the decrements in mean score are of such degree that proportionately far more mature adults than younger students are likely to have their teacher effectiveness challenged on the basis of test results.

Beginning at about the age of 55, the same conclusion may be made about the results of the Quantitative test. Decreasing mean scores with age will seriously limit the total ACE scores.
The source of the highly significant negative relationship between age and Quantitative score is in the Number Series and/or Figures Analogies subtests. This is a logical conclusion based on the evidence that the Arithmetic test, which is one of the three Quantitative subtests, yields comparable mean scores for the various age groups and has a very low correlation with age. This finding is in agreement with those of numerous other investigators reviewed in the literature. Tests with non-verbal content consistently show decreasing scores with increasing age. A second logical conclusion is that the contribution of Arithmetic to the total Quantitative score is proportionately far larger for older adults than for the younger subjects.

Discussion. The purpose of using the ACE tests is to provide an estimate of the scholastic ability and numerical facility of credential applicants in order to evaluate their prospective success in the various areas of concentration open to them in the public school system. It is obvious that when the vocabulary scores of older adults are compared with their test scores on the Linguistic and Quantitative scales, there has apparently been a serious loss of efficiency in the skills which the tests were designed to measure. This loss is extensive enough that proportionately far more older students than younger ones risk having their potential teacher or administrator effectiveness challenged on the basis of test results.
The ACE tests were designed for and standardized on a college age young adult population. For this population, the usefulness of the ACE for predicting academic success has long been recognized, and the tests have been used in this capacity on a nationwide basis for many years.

In the review of the literature, it is evident that many investigators question the practice of measuring adult skills on tests designed for the young. Sorenson (69, p. 436) has pointed out in connection with his investigation of college extension students with a wide age range that increased vocabulary with age may not represent increased mental growth and decreasing test ability may not represent actual deterioration or real or intrinsic capacity. A decrease in test ability among adults is probably caused by the fact that adults as they grow older, exercise their minds less and less with the material found in psychological tests.

He concludes that "the mental abilities of adults are determined within reasonable limits largely by adult intellectual habits."

A previously mentioned finding is that while speed and intelligence appear to be highly correlated in children and young adults, this parallel apparently does not continue to exist throughout life. Speed is doubtlessly an important variable related to efficiency, but there is no assurance that effectiveness in the majority of life situations depends on the speed of response which is required by a closely timed test such as the ACE.
Lorge (39) emphasizes that one of the sources of error in scales of mental ability is that speed and power are measured in undifferentiated amounts. When the speed factor is eliminated from the test situation, the general finding is that the younger adults take less time to complete the task, but that the differences in test scores between young and older adults tend to diminish or vanish, at least up to the age of about 55.

Definitions of physiological speed variables are more precise and measurable in simpler forms than the kinds of speed required in a timed test of mental ability. Investigators such as Ruger and Stoessiger (60) have found that motor characteristics rise rapidly in childhood, reach a maximum at about 24 or 25 years and gradually decline. Further, they find that sensory acuity (vision and limit of auditory pitch) declines earlier than motor characteristics. Miles (50) found decreasing speed of reaction and coordination with increasing age, but emphasizes that wide individual differences exist.

Just as there are physiological potentials which the individual is never called upon to develop or maintain at capacity, there are psychological or mental potentials which the majority of individuals are seldom called upon to use. A reasonable supposition is that tests designed for the young measure adaptability, while the same tests for adults measure
the consequences of the adaptive process. There is a difference between age changes that occur inevitably from the inherent physiological nature of the organism and the decline or preservation of functions (involving both psychological and physiological variables) which have reference to the adaptive efforts -- or lack of effort -- of individuals over a span of time. With increasing age, it is probable that psychological tests measure increasingly undifferentiated amounts of variables which are inherently related to the aging process and those which are related to the adaptive process.

McFarland (43) points out that speed, as commonly used by investigators of psychological measurements, is carelessly defined and suggests that time rather than speed is a more descriptive term. "Time is speed plus the interaction of many variables."

The emergence of this time or speed factor in adults has been used by Cattell (14) as one of the bases for his theory of "fluid" and "crystallized" ability. He hypothesizes that fluid ability is that which has to do with the discrimination and perception of relationships between any fundament, new or old. This ability increases until adolescence, then slowly declines. It is this fluid ability which is responsible for the intercorrelations, or general factor, found among children's tests and among the speeded or
adaptation—requiring tests of adults. Crystallized ability consists of discriminatory habits long established in a particular field, originally through the operation of fluid ability, but no longer requiring insightful perception for their successful operation. Cattell further states that intelligence tests test at all ages the combined resultants of fluid and crystallized ability, but that in childhood the first is predominant, whereas in adult life, owing to a recession of fluid ability, the peaks of performance are determined by the more crystallized abilities.

Balinsky (6) has found that the same abilities are not tapped by intelligence tests at various age levels, and abilities themselves are not constant. Jones and Conrad (51) for one instance, report that in the 6th decade of life, about 40 per cent of the total Alpha score is derived from two tests, while at age 10, these tests contribute only 25 per cent. The separate tests that compose a scale may not necessarily be described in terms of the same factors from age to age. Lorge and Rush (36) also emphasize that in the measurement of adult intelligence, the various tests are functionally impure.

The scores reflect the operation of many factors, e.g., physiological changes in vision and speed, environmental consequences of schooling, remoteness from schooling, occupational use, habitual versus creative processes in responding, and content variability in non-verbal performance and verbal subtests.
This study and many others demonstrate that decline in ability with age is largely a function of the content of the test, of procedural methods, and of numerous human variables.

Perhaps the assessment of adult intelligence should be undertaken through an analysis of the factors which are characteristic of separate age levels rather than make the tacit assumption that the nature of intelligence is independent of age. Such an evaluation of adult ability should be based on indices of successful performance in lifelike situations, accounting for the kinds of mental reorganization and adjustments that are peculiar to different age levels and paralleling the demands which the adult will be called upon to meet.
BIBLIOGRAPHY
BIBLIOGRAPHY


42. ———. The influence of the test upon the nature of mental decline as a function of age. *J. educ. Psychol.*, 1938, 17, 100-110.


46. ——— and Miles, W. R. The correlation of intelligence scores and chronological age from early to late maturity. *Amer. J. Psychol.*, 1932, 44, 44-77.


57. __________. Vocabulary and efficiency levels as functions of age in the Babcock method. J. consult. Psychol., 1947, 11, 207-211.


