Age-Intelligence Relationship between Ages Sixteen and Sixty-Four: A Rising Trend

RUSSEL F. GREEN
University of Rochester

Four stratified, random groups, ages 25–29, 35–39, 45–49, and 55–64, were drawn from the Puerto Rican population during the selection of the standardization sample for the Spanish language Wechsler Adult Intelligence Scale (WAIS). These groups were altered by the substitution of sufficient cases to make their educational distributions nearly identical. All subjects were administered the newly adapted Spanish language version of the WAIS during the period in which its standardization was being conducted. Results indicate that Full Scale scaled score means rise to age 40 and never fall; Verbal total scaled means rise to age 50 and never fall; Performance total shows a small decline after age 40—due primarily to Digit Symbol Substitution. Various analyses of the Spanish WAIS and of the United States WAIS standardization data give strong support to these results.

Much research has been devoted to changes in intellectual functioning as an adult ages. Confusion pervades the issue. Many people believe that intelligence declines between the ages of 25 and 65 (Jones, 1959; Wechsler, 1958). Others, to the contrary, believe that it continues to rise to at least age 50 (Bayley, 1955; Tyler, 1965). The purpose of this and following papers is to present new data in support of the viewpoint that intelligence as measured by the Wechsler Adult Intelligence Scale (WAIS) increases until approximately age 65. A further purpose of this paper is to demonstrate that if Wechsler and other investigators had conducted a different type of data analysis their conclusions would have been considerably different.

It is surprising that evidence such as that presented by Birren (1964) and others (Bayley, 1955; Bayley & Oden, 1955; Birren & Morrison, 1961; Christensen & Patterson, 1936; Lorge, 1936; Owens, 1953; Thorndike & Gallup, 1941; Tuddenham, 1948) has not seriously eroded belief in the decline hypothesis. Yet as late as 1966 Bromley published a book in which he adopted Wechsler’s (1958) lead. It would seem that if belief in the decline hypothesis is to be finally discouraged, Wechsler’s work will have to be directly attacked.

Wechsler (1958) pointed out that studies of the age-intelligence relationship had shown adult performance reaching a peak in the early 20s and then progressively declining. In answer to an objection that test scores may depend primarily on the acquired and stored knowledge of the individual, Wechsler agreed that evidence indicated that level of education is correlated with performance on intelligence tests. He went on to say that the age and education variables are confounded and that to examine the ef-
fect of age alone on adult test performance education must be held relatively constant. Studies of this nature have been lacking. Investigators of the age-intelligence relationship (Birren & Morrison, 1961; Jones, 1928; Jones & Conrad, 1933; Miles & Miles, 1932; Schaie, 1958; Wechsler, 1941, 1958) to date have given only inadequate consideration to what is undoubtedly a most important variable—this variable is formal education.

Wechsler (1958) summarizes his position as follows:

What is definitely established is: (1) that our intelligence tests can and do measure intelligence in older as well as in younger subjects to a substantial, although not necessarily to an equal degree; (2) that the abilities by which intelligence is measured do in fact decline with age; and (3) that this decline is systematic and after age 30 more or less linear [p. 142].

The data reported here were collected in order to make a direct test of the second and third of Wechsler’s conclusions.

Method

As part of a project to translate into the Spanish language, to adapt to Spanish culture, and to standardize the WAIS (Green & Martinez, 1967) in Puerto Rico, a random sample of the population of Puerto Rico was drawn in 1965. The randomness occurred within the two relevant variables, age and education. Briefly, the sampling method used was as follows. The Labor Department of the Island selected a random sample of dwelling units stratified on region and urban-rural residence. After taking a census of everyone 14 years of age or older who lived in the dwelling units, a random sample of those between the ages of 15 and 64 was chosen as the standardization group. The educational distributions of four age groups within the standardization sample (25–29, 35–39, 45–49, 55–64) were then examined and sufficient randomly chosen cases were added to the groups to make it possible to balance their distributions. Balanced-education groups were thus formed by adding the special subjects to the standardization groups and randomly dropping cases in groups which had too many people at some level of education. This resulted in a 67% average overlap between the educationally balanced groups and the corresponding standardization groups.

The name, address, age, occupational status, sex, and amount of schooling of each subject to be tested, whether in the standardization group or in the supplemental set, were given to an examiner who then contacted that person, solicited his cooperation, and tested him, if cooperation was obtained; approximately 96% of the subjects contacted cooperated. However, additional persons were lost during the course of the study due to death, emigration, military service, etc. Such losses were replaced with similar cases chosen from an auxiliary random sample.

The standardization sample was believed to be a more representative sample of the Puerto Rican population than one which could have been chosen to fit the 1960 census specifications. Any deviations from the 1960 census that occurred in the sample actually used were in the direction of known population trends. Those cases which were added to form educationally balanced groups were randomly chosen from the appropriate age-education strata of the population.

Results

Before considering the relationships between the education and scaled score trends two important facts must be established. First, it should be stressed that a response was obtained from every subject to the question of how many years of school he had completed. Further, this information was compared to the responses given to the Labor Department interviewers when they had previously asked the same question. Exact agreement was obtained in over 99% of the cases, indicating that most responses were reliable in terms of repeatability. Second, it must be established that the well-known decline pattern (Jones, 1955, 1959; Jones & Conrad, 1933; Miles & Miles, 1932; Pacaud, 1955; Schaie, 1958; Vincent, 1952; Wechsler, 1941, 1958; Weisenberg, Roe, & McBride, 1936) is observable in the Puerto Rican intelligence test standardization data. This result is demonstrated in Figure 1. The Full Scale, Verbal, and Performance scaled score means for each standardization age group are plotted in Figure 1 along with the mean educational level (number of years completed) for each group. The declining trend of the Full Scale scores is present and resembles a similar tendency reported by Jones and Conrad (1933) and others.

Note the close correspondence between the education and Full Scale plots of means from approximately age 32 on, and between the education and Performance total means from age 20 on. It is apparent in Figure 1

---

8 Further detail on the selection of subjects is available from the author.

4 The mean education plot is presented twice for the reader’s convenience.
that declines which occur in the Full Scale and Performance mean scores might well be due to age-related differences in education.

The plots of the 11 subtests for the standardization age groups that correspond to the four educationally balanced groups are presented in Figure 2. Since the 20–34 age span was used to calculate scaled scores during the standardization procedure, the mean scaled scores of the 24–29 age group are tightly gathered around 10. Note that five of the verbal subtests cluster rather closely together across age groups and show minimal decline while Digit Span, Block Design, and Object Assembly show a more steady decline. Picture Arrangement and Picture Completion decline still more rapidly with Digit Symbol Substitution declining most rapidly of all. This fanning out effect can be attributed to a changing configuration of mental abilities across age groups. As will be seen, however, these changes cannot be interpreted to represent age-related declines.

In summary, it is apparent that outcomes based on the standardization sample are largely consistent with the observations of others. The only notable difference is a rather tight clustering of the five verbal subtests over the investigated age span.

Comparisons of the most salient characteristics of the standardization sample and of the education-balanced groups are presented in Tables 1 and 2. Information in Table 1 includes (a) the number of subjects in each group, (b) the proportion of the balanced groups drawn from the standardization sample, (c) the mean education for each group, and (d) the urban proportion in the samples.

In passing, note that the urban proportions of the education-balanced groups rise with age. This apparently is caused by a tendency for older people with above average education to live in or to migrate into the towns.

The actual distributions of the number of years of education in each group are presented in Table 2. The distributions in the three younger education-balanced groups are slightly more similar to one another than are these three to the oldest group. Note that the oldest group has a slight deficit at two of the higher educational levels.

In order to avoid bias from a possible Aptitude × Education interaction, simple one-way analyses of variance were performed on the data obtained from the four education-balanced groups. The means, $F$...
### TABLE 1
**COMPARISON OF STANDARDIZATION SAMPLE WITH EDUCATION-BALANCED SAMPLE**

<table>
<thead>
<tr>
<th>Age group</th>
<th>N</th>
<th>Balance standardization</th>
<th>Mean education</th>
<th>Urban proportion in sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Standardization</td>
<td>Balanced</td>
<td>Standardization</td>
</tr>
<tr>
<td>25-29</td>
<td>118</td>
<td>135</td>
<td>.75</td>
<td>8.36</td>
</tr>
<tr>
<td>35-39</td>
<td>104</td>
<td>136</td>
<td>.67</td>
<td>6.82</td>
</tr>
<tr>
<td>45-49</td>
<td>94</td>
<td>134</td>
<td>.61</td>
<td>5.64</td>
</tr>
<tr>
<td>55-64</td>
<td>129</td>
<td>134</td>
<td>.66</td>
<td>3.96</td>
</tr>
</tbody>
</table>

*Proportion of balanced group drawn from standardization sample.*

### TABLE 2
**COMPARISON OF EDUCATION DISTRIBUTIONS OF THE FOUR EDUCATION-BALANCED GROUPS WITH THE CORRESPONDING STANDARDIZATION GROUPS**

<table>
<thead>
<tr>
<th>Yr. of education completed</th>
<th>Education-balanced age groups</th>
<th>Standardization age groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>18+</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16-17</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>14-15</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>12-13</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>10-11</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>8-9</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>6-7</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>4-5</td>
<td>31</td>
<td>33</td>
</tr>
<tr>
<td>2-3</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>0-1</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>WAIS</td>
<td>Age groups</td>
<td>F&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td></td>
<td>25–29&lt;sup&gt;a&lt;/sup&gt;</td>
<td>35–39&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Full Scale</td>
<td>102.01</td>
<td>109.13</td>
</tr>
<tr>
<td>Verbal</td>
<td>56.10</td>
<td>62.02</td>
</tr>
<tr>
<td>Performance</td>
<td>45.92</td>
<td>47.11</td>
</tr>
<tr>
<td>Information</td>
<td>9.38</td>
<td>10.35</td>
</tr>
<tr>
<td>Comprehension</td>
<td>9.27</td>
<td>10.18</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>9.49</td>
<td>10.57</td>
</tr>
<tr>
<td>Analogies</td>
<td>9.34</td>
<td>10.32</td>
</tr>
<tr>
<td>Digit Span</td>
<td>9.35</td>
<td>10.15</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>9.27</td>
<td>10.45</td>
</tr>
<tr>
<td>Digit Symbol</td>
<td>9.10</td>
<td>9.10</td>
</tr>
<tr>
<td>Picture Completion</td>
<td>9.24</td>
<td>9.34</td>
</tr>
<tr>
<td>Block Design</td>
<td>9.01</td>
<td>9.68</td>
</tr>
<tr>
<td>Picture Arrangement</td>
<td>9.12</td>
<td>9.32</td>
</tr>
<tr>
<td>Object Assembly</td>
<td>9.45</td>
<td>9.67</td>
</tr>
</tbody>
</table>

* N = 135.
*<sup>b</sup> N = 136.
*<sup>c</sup> N = 134.
* Simple one-way analysis of variance with α = .05.

Ratios may have been inflated relative to the extent to which an Aptitude × Education interaction occurred.

The covariance adjusted means, F ratios, and omega-squares for both the standardization and education-balanced groups are presented in Table 4. Results for the education-balanced group may be described as follows. The F ratios for Full Scale, Verbal, and Performance adjusted means are significant.
cant as are those for all subtests except Block Design and Object Assembly. They range from one and one half to three times the size of the F ratios in Table 3. Although the omega-square values are in every instance higher than those in Table 3 they are, nevertheless, quite modest. The general trend of the Performance total is downward. The Full Scale mean, on the other hand, rises to age 35-39 and then flattens through age 64. While differential rises are the rule, the net effect is for there to be a rise in means to about age 40 with no net fall before age 64.

It is clear that the age-score trends of the educationally balanced groups are very different from those of the standardization group. This is illustrated in Figures 3 and 4 (compare to Figures 1 and 2). In Figure 3 the Full Scale mean rises to the 35-39 age group and then becomes fairly level. The Verbal total rises at least to the 55-64 age group with no decline. The Performance total also rises to age 35-39 but then shows a slight decline. In Figure 4 one notes that this decline is caused largely by Digit Symbol Substitution scores which show a very marked decrease with age. The other Performance subtest scores also decline with the exception of Block Design. On the other hand, scores on five of the six verbal subtests rise all the way to the 55-64 age level.

**Discussion**

The validity of the conclusions drawn from this study depends on the adequate identification of potential biases and compensation for them. It is, therefore, necessary to scrutinize the procedure used to match groups and to consider its implications.

First, it is apparent in Table 1 that the mean education of the age 25-29 standardization group had to be lowered and the means of the other groups increased in order to form education-balanced groups. This was accomplished by the subtraction and addition of high- and low-education cases to the standardization age groups until their mean education and educational distributions were nearly equal. One might hypothesize that a positive bias was introduced into the three older groups by the addition of surplus high-education cases and that this bias was a result of the Education x Aptitude interaction; that is, high aptitude may have been self-selected for education and vice versa. The author considers this hypothesis to be incorrect and will attempt to refute it, first on logical grounds and then on the basis of supporting evidence.

Until the recent emphasis on educational
achievement in Puerto Rico, length of stay in school was most likely independent of aptitude. Owning a pair of shoes was historically a more important determinant of school attendance than aptitude and, it should be noted, large numbers of children were without shoes until the years following World War II. Therefore, older high-education subjects may have been blessed with certain cultural advantages rather than above average aptitude.

If high aptitude does select for education as the hypothesis suggests, then one would expect the older groups to have higher aptitude levels than the younger groups due to the presence of additional high-education cases. (The 55–64 education-balanced group had nearly twice the mean education level of the corresponding standardization group.) However, the average IQ score of the age 25–29 education-balanced group when converted against the 25–29 age norms was between 99 and 100. The average IQ score of the age 25–29 education-balanced group using its own age norms as a base line was 96, producing a discrepancy of only 3 or 4 points. Clearly, then, if the older group held an advantage in terms of high aptitude, it was not developed to a sufficient extent for it to become manifest in the test results.

Another point which should be made is that Puerto Rican schools have been improving across time; teachers are better trained and facilities are more adequate than in the past. Theoretically, the younger subjects should have received greater intellectual stimulation by the completion of a given level of schooling than older subjects and their potential should be more fully realized. This, of course, would tend to bias the test outcome in favor of the younger rather than the older groups.

Several other potential biases are worth noting. First, English was used as the official instructional language for a time and this could have biased results against the older subjects. Second, cases dropped from the older groups tended to be from rural areas and cases added were for the most part urban. This was not true, however, for the two younger groups which suggests a possible bias in favor of the two older groups.

While one can argue about the possible biases outlined in the preceding paragraphs, it is obvious that the use of education-balanced groups does not of itself yield compelling evidence. This is why it was deemed necessary to present alternative analyses of the data obtained from the standardization sample. The standardization groups are not subject to the same objections to the same degree as the education-balanced groups. If the standardization data yield the same results then the validity of the outcome is strongly supported.

The standardization sample suffered from as few of the potential biases just described as any sample that can reasonably be obtained. It was selected by random methods; each age group was, therefore, a representative segment of the population as it existed in 1965. It is difficult to imagine the presence of much differential Education × Aptitude interaction in a sample of this sort. The oldest group, in particular, had a mean education of only 3.96 years. While cultural advantage may have helped to select these people into school it is unlikely that aptitude selected them out by the fourth grade in generally poor schools. The standardization data thus can provide some of the supporting evidence necessary to refute the Education × Aptitude interaction hypothesis.

It is important to understand that discrepancies between mean educational levels of succeeding age groups within the population in no sense imply differences in intellectual potential. These differences are the product of cultural changes which have resulted in a steady increase in both the amount and the quality of education students have been receiving; they result from changing opportunities, attitudes, and needs, not from Education × Aptitude interaction. Therefore, one cannot meaningfully discuss age differences in intellectual potential unless a serious effort is made to control for the educational differences that occurred. It should be noted here that the educational frequencies for the four standardization groups were in keeping with the changes occurring in the population; that is, the younger groups included individuals with progressively higher levels of education.

Assuming that each age level was originally equal to the rest in potential and as-
assuming there is not much of an Aptitude × Longevity interaction beyond age 25 (which in the light of various results appears reasonable) it is appropriate to covary educational differences out of age-level differences on standardization test performance. This procedure equates the age groups in terms of discrepancies due to number of years of formal education completed. Such a covariance correction for group differences across age levels can be considered a partial correction for cultural change.

What are the consequences of this type of analysis? To the extent that correlation between education and schooling within each age group is due to self-selection of the high-aptitude people the within variance might be somewhat overcorrected. There would not, of course, be any significant overcorrection for the differences between means since the correlation between age and test score means is due primarily to cultural change. The net result might be to inflate F ratios although this tendency should be less than in the education-balanced groups since average educational levels are lower and the distributions have less variance.

The urban residence proportions, Ns, and proportions of common subjects of the appropriate four groups from the standardization sample are presented in Table 1. There exists a considerable overlap of subjects between the standardization and education-balanced groups which could slightly prejudice certain statistical results. However, the extent of nonoverlap should be sufficient to permit the occurrence of significantly large differences in outcome regardless of any statistical bias. The extent of possible difference in outcome is illustrated by noting that the mean education of the oldest group is 3.62 years lower in the standardization sample than it is in the education-balanced sample.

The results of covarying the standardization data for differences in average education and variation in urban residence are presented in Table 4. Although the adjusted means and F ratios are slightly lower than those of the education-balanced group, the pattern of outcomes of the two samples is virtually identical. Therefore, except for intercepts the plots for the standardization sample may be considered essentially the same as those of the education-balanced groups (see Figures 3 and 4). It is clear, then, that the results are largely independent of the experimental design.

Do these findings demonstrate that had Wechsler and others controlled for education in a similar manner they would have obtained similar results? The answer at first seems to be no. The groups in the United States standardization sample reached considerably higher levels of education than the corresponding Puerto Rican groups. Because of this an Aptitude × Education interaction might have had more opportunity to manifest itself in the United States population. There are, however, two ways to illustrate that Wechsler's data would show comparable results under a similar analysis.

The first method is illustrated in Figure 5 which includes plots of the Full Scale, Verbal, and Performance means by age level of Wechsler's standardization data for ages 25–64. The mean education for each age level as reported in the 1950 census (adjusted to 1955) is also presented. Upon

![Fig. 5. Plots of Full Scale, Verbal and Performance means by age level. (Intelligence test data obtained from Wechsler, 1958, plotted with median education as reported in the 1950 Census, adjusted to 1955.)](image-url)
comparing Figure 1 with Figure 5 one discovers that although the educational levels of the two populations differ considerably the important characteristics of the plots are highly similar. Correlations between number of years of school and test scores for the two populations were also found to be very much alike. Thus, it is clear that education is almost certainly related to test performance in the same way in both populations and that a covariance analysis of Wechsler's data would yield plots essentially like Figures 3 and 4.

The second way to illustrate that these results generalize to the United States is by referring to the age-by-educational-level plots of Wechsler's data as published by Birren and Morrison (1961). Information and Vocabulary rise with age while Digit Symbol Substitution and Picture Arrangement show the most marked decline, results which are very similar to the Puerto Rican data. Variation in the other subtest plots indicates that there may be some difference in details of the subtest trends between the two populations. However, such a conclusion cannot be considered final until Wechsler's data are reanalyzed with adequate controls for education. The portion of Wechsler's data used by Birren and Morrison clearly was not a representative sample of the United States population.

Implications and Conclusions

It has been demonstrated by two sets of analyses that intelligence as measured by the WAIS does not decline in the Puerto Rican population before about age 65. It has also been demonstrated that the same conclusion is almost certainly true for the United States. The overall trend is clearly upward until middle age at which time it becomes essentially stable until at least retirement age.

One must be careful to distinguish between decline in test performance and decline in potential or aptitude. It is possible that man's potential declines from age 30 on, but it is here contended that if such a change does occur it is not significantly manifested in WAIS performance before about age 65. The gap between potential and actual development may be large enough that loss of potential due to physical decline does not become manifest in performance on tests of mental ability until sometime around age 65 or later.

It has been shown that the apparently lower overall ability of the older group can be explained in terms of the smaller amount of formal education they received. These results along with others on education-test-score correlations to be reported later lend credence to the proposition that the level of formal education received tends to set the level of cognitive functioning of an individual for life. For verbal subtests educational level sets a general pattern of growth from a particular base line and for performance subtests one of minimal fluctuation over time—except for Digit Symbol Substitution which declines.

The results reported here have been shown to be consistent with results of longitudinal studies which show increases in verbal and other functions to at least age 50 (Bayley, 1955; Bayley & Oden, 1955; Owens, 1953) as well as with other research which shows no decline in verbal functions before age 65. They are also consistent, however, with the observation that certain purely speeded functions such as that tested in Digit Symbol Substitution do, in fact, slow down with age. This result will be reported more fully at a later date.

It is clear, then, that if Wechsler and others who used mostly power tests had controlled for education when determining age-intelligence trends by analysis of covariance, their conclusions would have been very different; they actually would have been inverted. Wechsler and others were understandably led astray by the alleged Aptitude $\times$ Education interaction and failure to recognize that such an interaction does not have the same outcome over all age levels.

One can conclude that (a) as Kuhlen (1940) anticipated, most inferences based on research results in this area have been deficient in that they have taken inadequate account of education as a concomitant of intelligence test performance, (b) a great deal of reanalysis is needed to correct the record, and, finally, (c) future research must account for this important variable. Failure to do so may already have lead to
far more errors of inference and prediction than one might wish to contemplate.

REFERENCES

Kuhlen, R. S. Social change: A neglected factor in psychological studies of the life span. School and Society, 1940, 52, 14-16.
Lorge, I. The influence of the test upon the nature of mental decline as a function of age. Journal of Educational Psychology, 1936, 27, 100-110.
Wechsler, D. The measurement of adult intelligence. Baltimore: Williams & Wilkins, 1941.

(Received August 25, 1968.)