THE MENTAL GROWTH CURVE DEFINED ON THE ABSOLUTE SCALE: COMPARISON OF JAPANESE AND FOREIGN DATA*

TAROW INDW
Keio University

INTRODUCTION

Thurstone's absolute scaling method can be applied to a psychological test which consists of a number of items each of which is responded to either correctly or incorrectly (4, 16, 17, 18). If it is admitted that the actual test data conform to the model involved in the absolute scaling, it will provide us with an interval scale for the ability underlying the test (13). This is one of the advantages of the absolute scale and another is that the scale can be defined, in a way, independent of the composition of items, for instance, the number of easy or difficult items of the test and of the scoring procedure of the raw score. The former will be apparent when we wish to consider, in a rational way, the maturity process of the ability underlying a test and the latter will be helpful when we make an attempt to compare results of several tests which are supposed to measure the same ability.

For the purpose of obtaining rational mental growth curves and comparing them with one another, I constructed the absolute scale for the Japanese Binet-style intelligence tests of three kinds; the Suzuki-Binet (14), the Tanaka-Binet (21) and the Takemasa-Binet (22) and for the WISC (10), the Japanese version of the Wechsler intelligence scale for children (8, 9, 24). Since good agreement was found among the results of these tests, it was possible to define a representative growth curve which should depict, so to speak, the objective feature of the maturity of Japanese children in intelligence. Besides, owing to the advantages of the absolute scale, an attempt was made to compare the results of Japanese data with the results of several foreign tests and some satisfactory agreement was discovered provided that the standards of the scale were equated in an appropriate way.

The assumption underlying the absolute scaling procedure may be best stated as follows;
A mental process which is required to respond to each item \((j)\) will be represented by a hypothetical continuum on which is located the measure of degree of that process possessed by an examinee as \(x'_j\).

There will exist an ability which is fundamental for the test consisting of \(n\) items as a whole. Forming a nucleus of the \(n\) processes assumed above, it will also be represented by a hypothetical continuum \(X\).

Let a projection of \(x'_j\) on \(X\) be denoted by \(x_j\), and it is assumed for \(x_j\) the normal distribution function:

\[
p(x_j) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_j-M_i)^2}{2\sigma_i^2}} \tag{1}
\]

provided that examinees are grouped according to their age \((i)\). In other words, if the distribution of \(x'_j\) is considered on \(X\), it is to be of the same form (1) irrespective of item \(j\).

The mean, \(M_i\), as well as the standard deviation, \(\sigma_i\), is a function of the age \(i\) only. Let a correlation between \(x_{ij}\) for item \(j\) and \(x_{ik}\) for item \(k\) in an age level \(j\) be denoted by \(r_{jk(i)}\), then no restriction is imposed on \(r_{jk(i)}\).

There will exist on \(X\) a measure of the ability \(G_j\) which represents a threshold for passing item \(j\), and \(G_j\) is a function of the item \(j\) only. Then the probability \(P_{ij}\) that an examinee of age \(i\) will answer an item \(j\) correctly is given by the equation:

\[
P_{ij} = \frac{1}{\sqrt{2\pi}\sigma_i} \int_{G_j}^{\infty} e^{-\frac{(x_{ij}-M_i)^2}{2\sigma_i^2}} dx_{ij} = \frac{1}{\sqrt{2\pi}} \int_{T_{ij}}^{\infty} \phi(t) dt \tag{2}
\]

where

\[
\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}
\]

\[
T_{ij} = \frac{G_j - M_i}{\sigma_i}
\]

\[
G_j = M_i + \sigma_i T_{ij}
\]

From Equation (3), we obtain a linear relationship over items between \(T_{ij}\) and \(T_{i-s,j}\) when two age levels, \(i\) and \(i-s\), are considered in pairs.

\[
T_{ij} = \left(\frac{\sigma_{i-s}}{\sigma_i}\right) T_{i-s,j} + \left(\frac{M_{i-s} - M_i}{\sigma_i}\right) \tag{4}
\]

Let \(p_{ij}\) represent the empirical value for \(P_{ij}\) and \(t_{ij}\) for \(T_{ij}\), then \(t_{ij}\) can be obtained by putting \(p_{ij}\) into (2).

![Fig. 1. An example of application of Equation (5) to real data.](image)

In this respect, the model here assumed is not entirely identical with that of Lord where \(G_j\) is defined on \(x'_j\) (II).

Hence, if the assumptions underlying the absolute scaling method are acceptable, we should expect that a linear relation will appear between \(t_{ij}\) and \(t_{i-s,j}\), and this was discovered almost always to be the case. An example was shown in Fig. 1. Then, fitting a straight line

\[
t_{ij} = a_i t_{i-s} + b_i \tag{5}
\]

to the scatter diagram of this kind, the values of \(M_i\) and of \(\sigma_i\) can be
determined as follows. Equations (4) and (5) give
\[ a_i = \left( \frac{\sigma_{i-1}}{\sigma_i} \right) \]
\[ b_i = \left( \frac{M_{i-1} - M_i}{\sigma_i} \right) \]
which express \( M_i \) and \( \sigma_i \) in terms of \( M_{i-1} \) and \( \sigma_{i-1} \). Hence, by adopting \( M \) and \( \sigma \) of an arbitrary age level as the origin and the unit of the entire scale, and by comparing successive pairs of age levels \((s=1)\), it is possible to assign rational values, one by one, to all \( M_i \) and \( \sigma_i \) which should theoretically, as far as the model described above fits the data, be independent of the composition of items and of the raw score arbitrarily defined. And, once the absolute scale is developed, the difficulties in sampling of items \((6, 7)\) can be put aside. However, the interval scale should be admitted only for the values of \( M \) and \( \sigma \), but not for individual score unless the matrix, \( R_i \), consisting of \( r_{jk(i)} \) defined in (III), is of rank one in the sense that all its elements are unity. Of course, this is not realized in the actual data and the following discussions will be restricted to the consideration of an age level as a whole. In order that the absolute scaling be applied, the matrix \( R_i \) should be of rank one, but it is not required that all of \( r_{jki(i)} \)'s are to be unity.

### SCALING PROCEDURE

In fitting a straight line (5) to the scatter diagram, not only random errors involved in \( t_{ij} \) but also those in \( t_{i-1,j} \) should be taken into consideration. Hence, the least square method developed by Deming (2) seems appropriate and was employed all through this study. This procedure will give the best fitted line in the sense that the sum defined as
\[ \chi^2 = \sum \left( \frac{e_{ij}^2}{\sigma^2(t_{ij})} + \frac{e_{i-1,j}^2}{\sigma^2(t_{i-1,j})} \right) \]
is minimum, where \( e_{ij} \) represents a residual parallel to the \( t_i \)-axis and \( e_{i-1,j} \), one parallel to the \( t_{i-1} \)-axis and \( \sigma(t_{ij}) \) represents the standard deviation of the sampling distribution of \( t_{ij} \) etc., which is given by
\[ \sigma(t_{ij}) = \frac{\sigma(p_{ij})}{\phi(t_{ij})} = \frac{\sqrt{P_{ij}(1-P_{ij})}}{\phi(t_{ij})} N \]
The standard deviation (9) consists of two parts; the standard deviation of the distribution of \( P_{ij} \) determined on a sample of size \( N, \sigma(p_{ij}) \), and the influence of fluctuation in \( P_{ij} \) upon \( t_{ij} \);
\[ \frac{dt_{ij}}{dp_{ij}} = \frac{1}{\phi(t_{ij})} \]
In calculation, \( P_{ij} \) was replaced by \( p_{ij} \) and the difference in unit for \( \sigma(t_{ij}) \) and \( \sigma(t_{i-1,j}) \), small if any, was ignored, namely, \( \sigma_i \) was assumed to be equal to \( \sigma_{i-1} \).

If one wishes, goodness of fit of (5) to the data can be evaluated by the use of the chi-square test and, in cases of the three Binet-style tests, necessary information is listed in Table 1 where \( X^2 \) means the observed value of Equation (8) and \( n \) the number of items made use of in the fitting. Items in which either \( p_{ij} \) or \( p_{i-1,j} \) is 1 or 0 have to be excluded in calculation. As seen from Table 1, in a third of the cases of the Tanaka-Binet, and of the Takemasa-Binet, and in all cases of the Suzuki-Binet, goodness of fit might be said to be "poor" if the result of the chi-square test is strictly adhered to. This will not mean, however, that the model of the absolute scaling should be discarded in these cases, since this rather disappointing result comes simply from the nature of the test employed. As a
The property of such a test that labels the discrepancy between observation and theory as either "significant" or "insignificant", the larger a sample, small and unimportant departures from the theory are the more likely to be detected. It is not surprising, therefore, that the chi-square test shows especially unfavourable results towards the Suzuki-Binet where the size of the samples is rather large. As pointed out by Gulliksen, the applicability of the model should be evaluated to the extent to which the model is successful in accounting for the total variance in the data (5). If this point of view is adopted, it is clear that even the case is acceptable where the fit is "poorest" in the sense of the chi-square test because the linear trend is so conspicuous in the scatter diagram that if the correlation coefficient is calculated here, it is never below 0.9. This will not mean, on the other hand, that each of the assumptions (I)-(IV) should be accepted on its face value. For instance, even when the pairs, \( p_{i-1,j} \) and \( p_{i,j} \), are directly plotted, that is equivalent to assume the uniform distribution for \( x_i \) instead of (1), an approximately linear trend can be also observed in this scatter diagram as exemplified in Fig. 2 which deals with the same age levels as in Fig. 1. As a matter of course, the linearity of the former is inferior to that of the latter, but, it is still true, the total variance in Fig. 2 is successfully accounted for by a straight line. It is of my opinion, therefore, that the applicability of the model should be discussed by the final results it brings.

The absolute scale can also be constructed by the use of the raw score as landmark. Then the subscript \( j \) should be understood as denoting one of the discrete steps of the score, so \( P_{ij} \) represents the proportion of the examinees of age \( i \) whose raw score is more than a score \( j \), and \( G_i \) represents a threshold of the ability for obtaining the score larger than \( j \). If the raw score is used as landmark, it is to be noted that \( P_{ij} \) for score \( j \) and \( P_{ik} \) for score \( k \) are not independent. Hence, the scaling procedure described above cannot be applied. I constructed the absolute scale of this type for the Suzuki-Binet with the purpose of making an international comparison of the mental growth curves. The procedure employed here will be described later.

**Growth Curve for the Binet Style Test**

The scaled values of \( M_i \) and \( \sigma_i \) of the three Binet-style tests are given in Table 1. As the origin and the unit of the scale for all the tests, the mean and the standard deviation of the 6-year-group of the Suzuki-
Table 1  Scaled Values for the Three Japanese Binet-Style Tests; The Absolute Scale was Constructed with the Use of the Item as Landmark.  \( P = P_{\alpha}(\chi^2 \leq \chi^2) \) under the Degree of Freedom Indicated.

| Age | Suzuki-Binet | | | | Tanaka-Binet | | | | Takemasa-Binet | | |
|-----|-------------|-----|-----|-----|-------------|-----|-----|-------------|-----|-----|-------------|-----|
|     | M | s | N | \( \chi^2 \) | d.f. | P | M | s | N | \( \chi^2 \) | d.f. | P | M | s | N | \( \chi^2 \) | d.f. | P |
| 4   | -1.050 | 0.851 | 45~71 | 32.2 | 22 | >0.05 | -0.849 | 0.867 | 87 | 16.3 | 28 | >0.95 | -0.975 | 0.841 | 119 | 36.1 | 32 | >0.10 |
| 4\frac{1}{2} | -0.454 | 0.928 | 184~214 | 211.0 | 23 | <0.01 | -0.002 | 0.963 | 162 | 37.4 | 30 | >0.10 | -0.218 | 0.963 | 119 | 45.7 | 33 | <0.05 |
| 5   | 0.000 | 1.000 | 612 | 132.0 | 25 | <0.01 | 0.300 | 1.060 | 111 | 70.7 | 24 | <0.01 | 0.300 | 1.060 | 141 | 74.7 | 37 | <0.01 |
| 6   | 0.550 | 1.074 | 616 | 299.0 | 28 | <0.01 | 0.936 | 1.106 | 160 | 39.5 | 31 | >0.10 | 0.963 | 1.263 | 173 | 48.9 | 35 | >0.10 |
| 7   | 1.489 | 1.274 | 377 | 38.9 | 21 | <0.01 | 1.607 | 1.102 | 100 | 17.0 | 27 | >0.90 | 1.924 | 1.271 | 162 | 71.2 | 34 | <0.01 |
| 8   | 2.241 | 1.365 | 404~428 | 94.4 | 24 | <0.01 | 2.050 | 1.099 | 100 | 35.2 | 21 | <0.05 | 2.663 | 1.321 | 172 | 21.3 | 30 | >0.50 |
| 9   | 2.659 | 1.427 | 563 | 48.1 | 25 | <0.01 | 2.765 | 1.150 | 100 | 20.0 | 18 | >0.30 | 3.200 | 1.343 | 154 | 34.8 | 33 | >0.25 |
| 10  | 3.131 | 1.538 | 586~598 | 85.0 | 22 | <0.01 | 3.333 | 0.938 | 100 | 37.0 | 32 | <0.01 | 4.140 | 1.401 | 186 | 64.2 | 35 | <0.01 |
| 11  | 4.244 | 1.726 | 315 | 138.9 | 28 | <0.01 | 4.884 | 1.267 | 181 | 21.1 | 26 | >0.70 | 4.884 | 1.267 | 181 | 21.1 | 26 | >0.70 |
| 12  | 5.37 | 1.484 | 126 | 14.8 | 26 | >0.95 | 5.596 | 1.479 | 61 | 14.8 | 26 | >0.95 |
Binet is adopted because the size of the sample is largest in this test and it is desirable to choose as a standard an age level around the middle in order to reduce the propagation of error.

Fig. 3. Mental growth curves on the absolute scale, constructed with the use of the item as landmark.

In the Fig. 3, $M_i$'s of the three Binet-style tests are plotted against age. The agreement among the tests is so close that a growth curve can be drawn by 'eye' without trouble. It is well to note that the least square method cannot be employed here because $M_{i-1}$ and $M_i$ are not determined independently. The $i$-age-level of the Suzuki-Binet is defined in such a way that the mean lies in $i$ year and 3 months. The mean of the Tanaka-Binet and of the Takemasa-Binet lies in $i$ year and 6 months and this is the reason why $M_6$ is defined as 0.30 and $\sigma_6$ as 1.06 in these cases. It was discovered that the agreement observed in Fig. 3 was destroyed if this difference of 3 months was ignored. That means, on the other hand, the agreement observed should not be regarded as a matter of course. Thus, the absolute scale seems to be worthy of its name.

Encouraged by this discovery, I made an attempt to compare the growth curve with that of foreign data. The data of the Binet test (Cyril Burt standardization) was the only one of the absolute scale available to me which was constructed with the use of the item as landmark (19, 20). The original absolute scale for this test was developed by Thurstone, but the converted values of $M_i$ are plotted in Fig. 3. The conversion was carried out by changing the unit of the original scale and by shifting the origin in accordance with the unit and the origin of the scale in this study, namely $M_6=0.30, \sigma_6=1.06$. It is satisfactory to see that these also run quite closely to the growth curve defined from the Japanese data. The result of the four tests seems to coincide as far as the range of age indicated in Fig. 3 is concerned.

The parameter corresponding to item, $G_j$, also can be located on the absolute scale. If the model of the absolute scale is acceptable, the empirical value of $G_j$ which can be estimated in each age level should be constant irrespective of the age $i$. As a matter of fact, it was actually

Fig. 4. Examples of the trends observed in $G_j$.
(A) Suzuki-Binet 60: Repeating 7 numbers in reverse order
(B) Tanaka-Binet 44: Counting 13 Go-stones.
the case in the majority of the items, but in some items marked trends of two directions were observed in the value of $G$ as shown in Fig. 4 as examples. Besides, it was discovered that the type A consisted of the items where "memory" plays a critical role, and the type B of the items which have something to do with "information". Thus, strictly speaking, several groups are to be defined* among $x_i'$ postulated in (1), whose rate of maturity is somewhat different from one to another and among them "information" and "memory" are the most remarkable ones. As to the rate, it can be said, from the direction of the trend, the former represents the greatest and the latter the least.

If this interpretation is correct, the growth curve of each of the subtest of the WISC should show a different rate of maturity and the growth curve of the omnibus test such as the Binet-style should run in between them.

**GROWTH CURVES FOR THE WISC**

From among the subtests of the WISC, seven listed in Table 2, were selected as objects of the absolute scaling. The subtests consisting of only a few items and the subtests inappropriate for the scaling were discarded. In this case, age $i$ and age $i+2$ were paired in calculation to make values of adjacent $M$'s independent in the growth curve. The zigzags observed in Fig. 5 should never have appeared if age $i$ and age $i+1$ had been paired as usual. The

* As a result of the factor analysis applied to item-to-item correlation matrix, it became clear in the Takemasa-Binet that the uni-factor model was inappropriate to account for the data (1). The analysis of this kind cannot be carried out in the other tests and it is regrettable that the raw materials of the Suzuki-Binet and of the Tanaka-Binet were lost by fire during the war.
values of $M_i$'s listed in Table 2 and plotted in Fig. 5 are not those that were published before (8). Since the be distorted to the uniform distribution ranging from $-3\sigma_i$ to $3\sigma_i$, the standard deviation of that distribution

![Fig. 5. Mental growth curves for the seven subtests of the WISC and the mental growth curve defined from the four Binet-style tests. The absolute scale was constructed with the use of the item as landmark.](image)

sampling procedure employed in standardization of these tests clearly involved a kind of bias, it seemed appropriate to make a correction as follows.

As examinees, 60 were selected by their teachers from a grade of a school in such a way that 20 should represent the superior ones, 20 the middle ones and 20 the inferior ones of the grade. Two possible models might be thought of to account for the result of such a selection assuming that the original population is normal: $N(M, \sigma^2)$.  

$\alpha$: If the distribution of $t$ age level is $1.73\sigma_i$.  

$\beta$: If teachers select a third of the sample from the superior group $1.17\sigma_i < x_i < \infty$, according to probability density for $x_i$ and the same is assumed for the middle group, $-1.17\sigma_i \leq x_i \leq 1.17\sigma_i$ and for the inferior group, $-\infty < x_i < -1.17\sigma_i$, then the standard deviation of the resulting distribution will be $1.35\sigma_i$. At any rate, the standard deviation must have been enlarged by the sampling procedure. Hence, taking the middle of the estimations in (\alpha) and in (\beta), the correction factor was assumed to be 1.5 and the scale was converted.
The converted values of $M_i$'s were plotted in Fig. 5. They run, as expected, around the growth curve defined by the Binet style tests, besides the curves for the subtests in which "information" plays an important role run above the Binet-style test curve and, as seen from Table 2, the values of $\sigma_i$ in these tests tend to increase with age, while in other tests, $\sigma_i$ remains approximately constant over all ages.

The factor analysis was applied to the pooled data of the five subtests; Comprehension, Similarities, Arithmetic, Picture Completion and Block Design (23). As an element of the correlation matrix, $r'_{jk}$, an average of tetrachoric correlations between items $j$ and $k$ over several age levels was taken. Since 5 or 6 items were sampled from each of the subtests to be pooled and each of $j$ and $k$ varied over all items sampled, the matrix $R$ was $28 \times 28$. Five factors were extracted and clearly separated. There was no difficulty in identifying these factors because only the items belonging to the same subtests clustered around the same factors. Hence, each subtest has a common factor of its own and the result seems to correspond to the fact that each has a growth curve of its own on the absolute scale. In passing, two second-order-factors were discovered to underlie the correlation matrix of these five factors, and the one could doubtlessly be identified as "verbal" and another seemed to be a mixture of "reasoning and space". The details of this study will soon be published.

THE ABSOLUTE ZERO

Thurstone made an interesting discovery. Obtaining a linear relation between $\sigma$ and $M$, he came to the idea of locating the absolute zero of the scale to the point at which $\sigma$ vanishes, since the value of $\sigma$ cannot be negative in its nature. What is most interesting, then, is the fact he succeeded to show in several tests that the mental growth curve, when extrapolated, passed through the absolute zero at or about three months before birth (19). However, it should not be overlooked that the shape of the growth curve is considerably different from test to test. As far as I know, no attempt has yet been made to compare the growth curves of these tests.

Unfortunately this discovery was not confirmed in this study. In the Tanaka-Binet, the Takemasa-Binet and the WISC, the increase of $\sigma$ with $M$ was so slight that the absolute zero could hardly be located. Although the position of the absolute zero could be accurately estimated in the Suzuki-Binet, it was entirely different from that of the Binet (Cyril Burt) and the growth curve did not seem to be asymptotic to this point (24). In accounting for these discrepancies no other reason might be thought of than that something must have been different in the sampling procedure of these tests which show excellent agreement in so far as the growth curves are concerned.

SCALING BY RAW SCORE AS LANDMARK

Since most of the absolute scales developed in the U.S. are constructed by the use of the raw score as landmark, it was felt desirable to construct the scale of the Japanese test in the same way. Necessary information for this purpose was available only
in the Suzuki-Binet among the three Binet-style tests. The result is shown in Table 3.

Table 3  Scaled Values for the Suzuki-Binet Tests. The Absolute Scale Was Constructed with the Use of the Score as Landmark.

<table>
<thead>
<tr>
<th>Age</th>
<th>$M$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/0</td>
<td>-2.60</td>
<td>1.02</td>
</tr>
<tr>
<td>4/6</td>
<td>-1.71</td>
<td>0.90</td>
</tr>
<tr>
<td>5/0</td>
<td>-1.03</td>
<td>1.01</td>
</tr>
<tr>
<td>5/6</td>
<td>-0.65</td>
<td>1.14</td>
</tr>
<tr>
<td>6/0</td>
<td>-0.07</td>
<td>0.94</td>
</tr>
<tr>
<td>6/6</td>
<td>0.30</td>
<td>1.01</td>
</tr>
<tr>
<td>7/0</td>
<td>0.71</td>
<td>1.06</td>
</tr>
<tr>
<td>7/6</td>
<td>1.43</td>
<td>1.11</td>
</tr>
<tr>
<td>8/0</td>
<td>1.77</td>
<td>1.07</td>
</tr>
<tr>
<td>8/6</td>
<td>2.31</td>
<td>1.13</td>
</tr>
<tr>
<td>9/0</td>
<td>2.76</td>
<td>1.13</td>
</tr>
<tr>
<td>9/6</td>
<td>3.06</td>
<td>1.13</td>
</tr>
<tr>
<td>10/0</td>
<td>3.34</td>
<td>1.07</td>
</tr>
<tr>
<td>10/6</td>
<td>3.61</td>
<td>1.14</td>
</tr>
<tr>
<td>11/0</td>
<td>3.88</td>
<td>1.09</td>
</tr>
<tr>
<td>11/6</td>
<td>4.26</td>
<td>1.38</td>
</tr>
<tr>
<td>12/0</td>
<td>5.04</td>
<td>1.14</td>
</tr>
<tr>
<td>12/6</td>
<td>5.25</td>
<td>1.06</td>
</tr>
</tbody>
</table>

It might be well to note that, if the raw score is used as landmark, the scale becomes heterogenous in its content because the same score can be obtained by answering various combinations of items correctly. Thus it is not unlikely that the content of ability supposed to underlie the score somewhat differs in its nature from that under discussion when the scale is constructed by the use of the item as landmark.

As mentioned before, Equation (5) cannot be fitted by the least square method because $t_{ij}$ and $t_{ik}$ or $t_{i-1,j}$ and $t_{i-1,k}$ are not independent here. Hence, the scaling was carried out graphically by the method of successive intervals (3). The weight, the inverse of the square of (9), was taken into consideration and the origin and the unit of the scale were defined as before: $M_6=0.30, \sigma_6=1.06$.

In Fig. 6, $M_i$'s of the Suzuki-Binet thus calculated and those of foreign data converted are plotted against age. The foreign data under consi-

![Fig. 6](image_url)
deration are limited to the case which contains the standard age level, 6 year, in its age. The original absolute scales for these tests were constructed by Thurstone and Ackerson (20) and by Odom (12). As seen from Fig. 6, the agreement among the data is satisfactory indeed. In passing, we may mention that, as to the Suzuki-Binet, the two scaling procedures gave approximately the same growth curve for $M$ but the remarkable difference appeared in the rate of change of $\sigma$ as a function of age (cf. Table 1 and Table 3).

Several more absolute scales of foreign tests are available whose data begin at an age above 6 years, the standard age of the scale under discussion (12). Since, as shown in Fig. 7, no general agreement was discovered with respect to the rate of change of $\sigma$ with age, the result of conversion cannot be independent of the age level at which the scales are equated. Thus, in these tests, the growth curve and the curve describing $\sigma$ as a function of age were extrapolated to estimate the values of $M_6$ and $\sigma_6$ of them, and the scale of these tests were also converted at the 6-year-level as the standard.

* If a slightly skewed to the right and slightly leptokurtic distribution is assumed on $X$ instead of the normal distribution, fit of straight line (5) to the data becomes better in this case, especially when $|t|>2.5$. The details will be published in the near future (15).
It may not be entirely impossible to imagine a growth curve for the converted results of the Dearborn Series II, of the Otis Primary, of the National Intelligence, of the Stanford Binet* (Institute of Juvenile Research data), of the Illinois (Bloomington data) and of the four tests discussed above, although the curve becomes slightly different from that of Fig. 6, and goodness of fit of the data becomes considerably worse. However, the Otis Advanced and the Illinois (Chicago data) have entirely different growth curves of their own (Fig. 8).

Fig. 8. Mental growth curves on the absolute scale, constructed with the use of the score as landmark.

It is well to mention that, if the conversion is carried out at arbitrary age levels, from test to test, to bring the best agreement among the converted results, it is certainly possible to attain an apparently more satisfactory result. But it may not be fair, and, it seems to me, the agreement so far described is great enough to warrant fruitfulness of the absolute scaling in so far as the mental growth curve is concerned.

The diversity observed in the rate of change of $\sigma$ with age will, at least in part, be ascribed to the difficulties in sampling of examinees. Usually one is careful in avoiding the bias which is to affect the mean, but, on the other hand, the circumstances influencing the dispersion are sometimes difficult to control. The difficulties are especially encountered at a higher age level and the distribution is apt to be limited somewhat in its range. No doubt, the extraordinary large dispersions at higher age levels found in the Stanford-Binet of the Institute of Juvenile Research are due to the circumstances that superiors who come to the Institute for scholarship grants are included in its sample as well as idiots and imbecils who come for consultation. In general, the rate is small in Japanese data and $\sigma$ does not much

* Mental age was used as landmark in this case.
increase with age. Presumably the fact has something to do with the wide spread of elementary education in Japan.

SUMMARY

1) The absolute scale was constructed, with the use of the item as landmark, for the Japanese Binet-style intelligence tests of three kinds and for the WISC, the Japanese version of the Wechsler intelligence scale for children.

2) Agreement among the results of these tests was great and a representative growth curve was obtained which would depict the objective feature of the maturity of the Japanese children in intelligence.

3) The growth curve thus defined was also in good agreement with that of the Binet test (Cyril Burt), provided that the standards of the scale were equated in an appropriate way.

4) The model involved in the absolute scaling was discussed in connection with the results of this study and with the results of the factor analysis applied to the Suzuki-Binet test and to the WISC.

5) The absolute scale was constructed for the Suzuki-Binet test by the use of the score as landmark, and the growth curves of eight foreign tests were converted on this scale for the purpose of comparison.

6) As far as the growth curves are concerned, excluding two, an appreciable degree of agreement was discovered among the remaining tests. The agreement was especially close when the comparison was limited within the tests in which no extrapolation was needed for the conversion.

7) Concerning all the tests dealt in this article, no general agreement was found in the standard deviation as a function of age, and possible reasons for this diversity were discussed.

8) Thurstone's discovery concerning the absolute zero was not confirmed in this study.

REFERENCES


17) Thurstone, L. L. The unit of measurement in educational scale. J. educ. Psychol., 1927, 18, 505-524.


(Received January 24, 1958)