PARTICULAR SYSTEM of MEASURING the LEVEL of KNOWLEDGE in HUNGARY, CONFORMING to the POSSIBILITIES of the INTEGRATED SCHOOL-SYSTEM

Sándor Orosz

I.

FUNCTIONS OF THE MEASUREMENT IN THE INTEGRATED SCHOOL-SYSTEM

1. By way of introduction it must be mentioned that schools in Hungary were brought under state management a quarter of a century ago. This led to the conditions for the creation of an integrated school-system. Following this there grandually developed an integrated network of primary and secondary establishments: the compulsory eight-class primary school and the four-class secondary school based on this. The integrated school-system which emerged is one of the important factors in the implementation of the social aims in Hungary.

The integrated school-system has many advantages and of course certain disadvantages too. At this point it is desired merely to indicate briefly those advantages which are of importance as regards the present theme.

a. From the point of view of the schoolchildren the main advantage of the integrated school-system is that it ensures equal rights and possibilities for the completion of primary, secondary and higher studies. In addition to the rights, emphasis must be made regarding the possibilities; these are ensured by the uniform curriculum, the uniform requirements, the same text-books, and the unity in the most important basic principles and methods of the tuition and education. Under such conditions it does not lead to a break in the child's development if, for example, it is necessary for some reason for him to change schools.

Another of its essential positive characteristics is the openness, as a result of which everyone, depending on the subjective factors and the natural endowments, has the same possibility of entering any of the given types of schools after completion of the primary schooling.

All schools of a given type provide identical qualifications, and thus they ensure the same prospects as regards employment opportunities.

b. The integrated school-system also has many advantages with regard to the planning and direction of public education: it is possible to centralize the planning of the contents of the tuition and education, and the preparation and distribution of the curricula, text-books and other means of implementing these contents; the teaching and postgradual education of pedagogues can be integrated; a possibility arises for obtaining information of the results continually, and hence the tuition - education process can be regulable.

The integrated school-system has many economic advantages too, but it is not wished to discuss these here.

One objective means of constant feed-back ensuring control of the process is measurement of the level of knowledge.

2. There are several functions of the measurement of the level of knowledge.

a. As already pointed out above, its main role is continual feed-back. The information obtained on the results of the process are utilized by the teachers, those responsible for preparing the curricula, the authors of text-books, the national directing organs and the pedagogical research workers. The constant feed-back is no less important to the students themselves: the objective evaluation may be an extremely motivating factor in the learning.

The results of the measurements are utilized most directly by the teachers: in all phases of the teaching they can obtain continual information on the effects and deficiences of the work carried out, and hence the methods, means and organizational forms of dealing with the whole of the class and with the individual students may be reshaped at any time and in any detail in an adequate way. The actual evaluation of the achievements of the students is established by the objectivity of the measurements.

For those who prepare the curricula the national measurement results are exact indications of the good qualities and the weak points of the curricula, of their practicabilities and the need for changes and modifications, and of the necessity of decreasing or expanding the subject-matter taught. They can also obtain a picture of whether the principle of the arrangement of the subject-matter was satisfied by the objective requirements, whether there is linear or concentric material-arrangement in concrete cases, where and what concentration would be necessary, and after what time it is justified to carry out curricular modifications and reform, etc.

Similarly, the text-book writers obtain indications from the measurement results as to whether the structure of the text-book and its logical systematic division are suitable, whether the individual thematic units are correctly constructed, whether the presentation and the analysis of the material are appropriate for the developmental level of the given age, whether the students are in possession of the basic knowledge necessary to learn the set material, whether the text-book contains sufficient regulatory and methodological parts, whether the number and quality of the exercises are satisfactory, etc.

The central directing organs of public education can obtain information from the national measurements as to which are the most important difficulties and problems requiring general action, and what are the common difficulties of the schools of the various regional units, which differ from the general problems; these indications provide a basis for the discovery of the mistakes and deficiencies, and establish the steps to be taken to eliminate these shortcomings.

It helps the pedagogical scientists to find the most direct and most topical questions.

b. In the integrated school-system the standardization of the means of measurement is simplified: in general a single national representative measurement is sufficient to establish all those characteristic features which can only be estimated after frequently lengthy and complicated validity and reliability investigations in thae heterogeneous school-system.

c. The results of the national representative measurements provide an objective basis of comparison for the evaluation of the performances of individual students and classes. Indices calculated from the national results can be added to the standardized tests, and form the objective basis of the evaluation.

3. In order for the measurements to be able to perform the above-listed functions completely, it is vital that they satisfy certain fundamental requirements.

a. A most important requirement is totality. This means that the tests must contain all aspects /idea, fact, concept rule, law, theory/ and all types of activity of the material-unit measured. The measurement therefore extends to the curricular maximum. There are three fundamental reasons for this:

Frist: if it is desired to know to what extent the students master the curricular material, then it is not possible to select or to choose in the knowledge to be checked.

Second: the curricula do not display the compulsory minimum. Thus, the selection of the curricular material in the measurement from any point of view at all would be done on the basis of an individual consideration, and this would introduce subjectivism.

Third: in the integrated school-system the tests are prepared for general use, and reach the teachers and students after typographical duplication; they can therefore be regarded as just such educational accessories as, for example, note-books or programmed material. Under such conditions, the selection of the curricular material from any point of view might have the result that the teachers would not strive to achieve the curricular requirements in the course of their tuition, but would concentrate on the knowledge contained in the tests, since the results are determined with the help of the tests.

b. Another basic requirement is representativity.

In the integrated school-system the possibility exists for a given mark to indicate the same level of knowledge everywhere. Consequently a n objective basis of comparison is necessary for the evaluation of the achievements. For this purpose, in place of the speculative norms the national average results automatically suggest themselves. Because of the many hundreds of thousands of students, a full-scale measurement can not be carried out to determine these. But it is not necessary: the representative sample /on the basis of the law of large numbers/ is just like a scaled--down version of the overall population, and so the results of the methodically made representative measurements reliably reflect the national situation. These representative measurements must be repeated at certain intervals, in order that the changes /development/ undergone throughout the years can be taken into consideration so that the norms do not grow rigid, because in this way they would become impediments to further development.

II.

METHODOLOGICAL BASES OF THE MEASUREMENT

1. The reliablitiy and objectivity of the results depend to a large extent on what scale is used in the measurement.

a. The technical literature on the methodology of measurement /e.g. Torgerson, 1962; Sixtl, 1967; Rummel, 1958/ deals in detail with the types of scales. These works come to the uniform conclusion that the metric /absolute and interval/ scales ensuring objectivity to a maximum can be applied in few fields in pedagogy.

As is well known, the main characteristic of these scales is the additivity, which is ensured by the fact that it has unit value, and all other values are definite multiples of the unit value; further, the intervals between the units are identical. With such scales it is possible to determine primarily data regarding the physical development of the students; they are applicable almost exclusively to the measurement of performances in sport and productive work, and are not suitable directly to express study achievements and the level of knowledge.

b. Ordinal scales have extended to the evaluation of the level of knowledge. The ordinal scale is used all over the world in traditional classification, independtly of how many marks /numbers/ are involved in the evaluation, and of whether a better performance is indicated by the larger or the smaller numbers.

In practice, however, ordinal scales are also applied when the evaluation is made by a points system instead of marks. The situation is made more complex here by the fact that /according to the estimated degrees of difficulty of the problems/ a different number of points is established for the evaluation from problem to problem, i.e. ordinal scales of different degrees are used within a given performance.

The determination of the mark is made difficult by a number of factors, which make the objectivity of the marks doubtful.

First: very complex achievements are evaluated overall with a single mark, and so as regards the details it is not possible to give sufficient consideration to the elements of the achievement or to the proportions of correctly and incorrectly solved elements. In other words: the criteria of the establishment of the marks are not clear-cut.

When evaluation is made on a points system, although it differs from problem to problem, in the establishment of the number of points the conditions for decreasing the number of points /subtraction/ in accordance with the incorrect elements can rarely be determined unambiguously.

Second: qualitatively very different things must be measured and compared with the grading, and so it is not possible to satisfy the basic requirement of the methodical scale of order, which can be worded as follows: If some element "a" is greater and better than element "b", and element "b" is greater and better than element "c", then of necessity "a" must be greater and better than "c". /Expressed as a formula: $a > b \wedge b > c \rightarrow a > c$./

For example: the mark for a mathematical problem is 4 /"good"/, while that for a grammatical problem is 3 /"average"/. Does this mean that the mathematical work is objectively more valuable and better than the grammatical work? To some people this question might appear absurd perhaps, but it is certainly a just one, since when the two, qualitatively very different things are evaluated with the units of one and the same measurement system, then - consciously or subconsciously - we compare them quantitatively. And since the qualitative difference between them is very great, they exceed the critical limit of comparability.

In principle the comparison is justified and objective only when the things to be compared are qualitatively homogeneous. Naturally, the homogeneity itself is not absolute, and there are also limits of tolerance to the homogeneity; this gives a critical limit to the quantitative comparability.

The basic relation of the quantitative comparison can be seen from what has been said above, and can be formulated as follows: the greater the qualitative difference between the things to be measured /compared/, the closer is the critical limit of comparability, and vice versa: the smaller the difference between them, the more distant the critical limit of comparability, and the more objective and more accurate the comparison and the measurement.

c. From this relation there arises the possibility of making the measurement of the level of knowledge objective: the performances must be broken down to elements whose qualities do not differ from each other, while one of their qualities can be unambiguously determined. In this case the elements with identical qualities can be counted, and as a result a derived scale is obtained, in which every counted element has unit value.

In effect the well-known nominal scale is applied here. On the application of the nominal scale designated things of the same quality and which can be defined exactly are taken into account.

If the performances of the students are broken down to such small elements of knowledge whose qualities can be unambiguously categorized from the possible alternative solutions good or bad, correct or incorrect, then in effect two designated, accurately defined nominal groups are formed: each of the knowledge elements is taken into account in the "good" or the "bad" nominal group. This reckoning can be carried out by assigning numbers to the two qualities and to the two types of nominal groups: elements belonging to the "good" group receive a value of 1, and those belonging to the "bad" group a value of O.

Since every element is classified into one or other of these groups, the value of the entire performance can be expressed as the quotient of the well and badly solved elements.

2. With the breaking-down to alternative elements the fundamental problem of the scaling is solved: a derived scale has been produced, which satisfies the fundamental requirement of the metrical scales, the additivity.

What constitutes an alternative unit can in general be determined; it is more correct, however, to decide this concretely in every measurement, depending on the subject-matter.

In general an alternative unit is formed by each individual single-term elementary decision. All types of knowledge are composed of these. The individual marks of the concepts are each separate single-term decisions; in the methodical concept-definition the genus proximum too is a single-term decision. The facts are true, single-term, elementary decisions. The rules and laws too are composed of series of elementary decisions.

The operative activities and problems are somewhat more difficult to separate into alternative units. Here the elementary operations constitute in general an alternative unit.

3. However, the subject of our measurement, knowledge, possesses characteristics which give rise to further difficulties.

The problem is that according to their importance the alternative units do not always have the same weight, they are not always homogeneous. It is sufficient to illustrate this by considering a single simple example. If, let us say, the students had the task of writing the date of the outbreak of the Second World War, this problem can be broken down into three, accurately classifiable alternative elements: year, month, day. Of these three facts, however, the knowledge of the year is clearly more important than the knowledge of the month or the exact day: from the point of view of the development of a chronological orientation and a historical attitude, knowledge of the year performs a more important function than that of the month and the day. It is necessary that it receives a greater weight in the evaluation too.

The knowledge of the students represents different values at the different levels. In the multiple choice questions which are fairly widespread in the measurement of knowledge, the students give an account of their knowledge in the main at the level of recognition. A more difficult task than this is the reproduction from memory of learnt knowledge. A still more difficult achievement is the practical application of the knowledge to carry out various operations and solve problems.

And as regards what is in fact difficult to the students themselves, and what is easily studied /in other words what is either important or not important to them/, this depends on a large number of factors. In the evaluation, however, consideration must be given to what experience shows: which elements are more difficult to learn, and which elements present less serious problems for the students to learn.

As can be seen, the alternative achievement-elements are generally not homogeneous, but from several points of view are heterogeneous.

To summarize what has been said, it can be stated that in the measurement of the level of knowledge it is not sufficient to break down the performance to alternative units, and to take these into account unit by unit; it is also necessary to weight these units.

It has already been indicated in the above that three types of weighting are used in our measurements.

a. The functional /importance/ weight must be established. This is done by a statistical procedure. With the help of a table every individual element of the test problems is weighted by an adequate number /80-120/ of theoretical and practical experts, research workers and teachers. Each element is grated separately into the following three categoriest: „very important" /3 points/, „important" /2 points/, and „less important" /1 point/. The functional weight of each individual alternative unit /the value of F_p/ is the average of the obtained values.

The choice of experts is made in accordance with the rules of representative sampling, and their number is adequately large, and so the calculated functional point values are as objective as possible.

b. The level weight /value L_p/ can be determined by any specialist alone.

It should be noted that the determination of the levels is still not sufficiently elaborated. It is certain that the recognition level can be well distinguished from the level of reproduction, but the application of knowledge takes place at several levels: there are simpler, easier and more difficult, more complex applicational problems. We do not yet possess sufficient experimental data for a reliable theoretical differentiation of these. As a provisional solution the value of L_p for accounting problems on the level of recognition is 1, on the level of reproduction is 2, and on every level of application is similarly 3 points.

In our experience the degrees of difficulty within the level of application are expressed by the empirical point values to be discussed below, and so the results are not distorted by the lack of further differentiation.

c. The empirical weight /value E_p/ is determinde by representative measurement.

Its calculation is based on the consideration that if someone can not answer a question then to reply is an insolubly difficult problem for him. From the point of view of the measurement the reason for this is of no importance: it does not change the fact.

If the majority of the students do not solve a certain problem or problem element in a national representative measurement, this means that its solution is objectively difficult. The degree of difficulty and its measure, the value of E_p, can be expressed by the percentage ratio of those giving the incorrect solution.

It should be stressed again that as regards the determination of the value of E_p the cause of the difficulty is insignificant, since the aim is the establishment of the degree of difficulty necessary for the evaluation; the discovery of the reasons is advisable from a didactical point of view. The reasons can be determined by qualitative analysis and in the main by the experimental study accompanying this. Of course, light can also be shed on certain reasons by mathematical procedures if it is possible to collect suitable causal data on the students taking part in the measurement /family conditions, housing conditions, methods and means of teaching, the equipment of the schools, etc./: the correlation between these and the performance levels can give information on the reasons.

d. Separate evaluation using the three types of points would greatly complicate the use of the tests. With a view to the easier manageability and the elaboration of more comprehensive performance indeces, an amalgamated point value is calculated for each problem element. This is the product of the three point values.

Finally, in order to simplify the work of the teachers applying the tests, the total point values for the problem elements are converted to percentage distribution ratios. These values are called percentage points /%p/. Consequently, in evaluating the test it is sufficient to sum up the %p values of the correctly solved problem elements, whereby the value of the performance is obtained expressed as a percentage.

4. The evaluation practice which has emerged demands the conversion of the measured results to marks. There are several methods known for this.

a. One method, which may be named a procedure based on convertion, is based on theoretical and practical considerations: for example, with a performance of 0-40 % the student receives a mark of „fail", with 41-55 % „fair", with 56-70 % „average", with 71 - 85 % „good", and with 86 - 100 % „excellent".

The main deficiency of such general, predetermined, speculative norms is that they are not related with reality. When such norms are used it may happen that the majority of the students receive a mark of „fail" in a given measurement. For instance, it is not a rare occurrence in reality for the national performance average for students for one or other of the knowledge units of the subject matter to be below 40 %p. In such a case, on the basis of the norm given as an example half or more of the students would receive a mark of „fail".

If the national level is so low, then the reasons and the possibilities of overcoming these must be sought. The reason may be the inappropriate construction of a text-book, the use of incorrect teaching methods, etc. With classification according to a prefixed norm, we should in effect be penalizing the students en masse, and mainly for reasons beyond their control. For this reason we do not consider speculative norms as suitable for use.

b. A more realistic and more well-founded procedure than this is conversion to marks by the method of averaging.

The essence of this is as follows:

In the national representative measurement the performances of the students are traditionally assessed by the teachers into the grades „excellent", „good", „average", „fair" and „fail". The tests with these grades are then separated into the various groups, and the averages of the measured %p values are calculated for each grade.

Let us consider the following case for example: the measured average for the students who obtained a grade of „fail" is 26 %p, for the „fair" grade is 34 %p, for the „average" grade 52 %p, for the „good" grade 78 %p, and for the „excellent" grade 88 %p /Ágoston-Nagy-Orosz, 1971, p. 106/. The norms of the conversion to marks is obtained from these averages by calculating the arithmetical mean of adjacent grades. Thus: 26+34:2 = 30%p; 34+52:2 = 43 %p; 52+78:2 = 65 %p, 78+88:2 = 83 %p. Hence the key to the conversion to marks is:

fail	0 - 30 %p
fair	31 - 43 %p
average	44 - 65 %p
good	66 - 83 %p
excellent	84 -100 %p

It is well known that marking is performed in a fairly subjective way. This is shown by comparison of the grades awarded and the measured results. According to the evidence of national measurements, however, the spread in the grading results is so large that it exceeds all conceptions. Thus, the subjectivity of the grading is much stronger than we might expect. Evidence of this is provided by the following example.

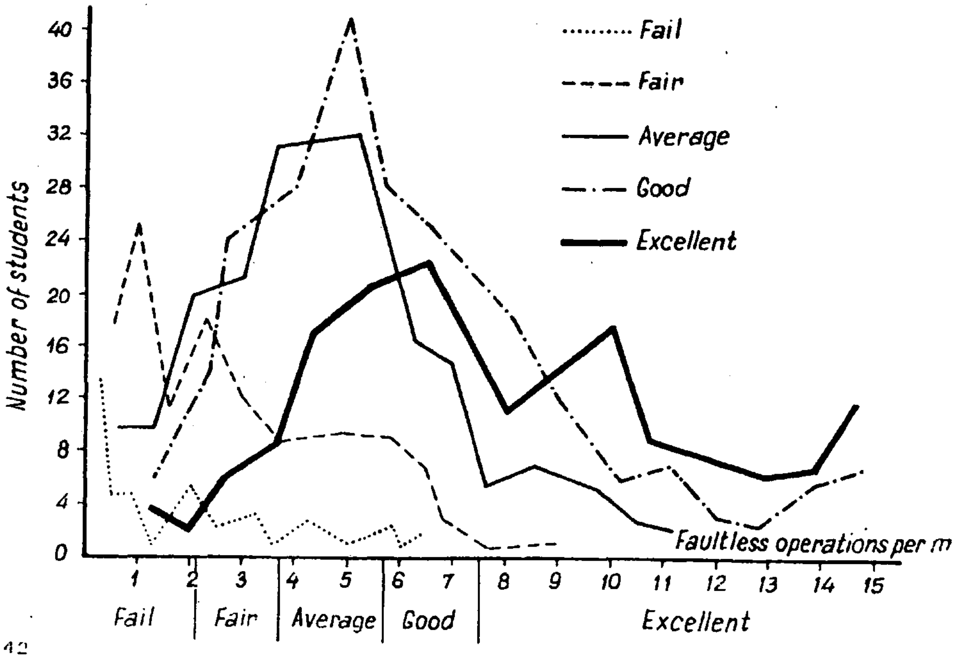
In 1968 we measured /Nagy, 1971/ the national level of development of ability in elementary calculation. One of the indicators of the level is the number of faultlessly performed operations per minute. Based on the averaging of the grades awarded by the teachers, in class 1 of primary schools the average performance of students who received a grade of

„fail" was 1.5 faultless operations per minute; that for a grade of „fair" was 2.7; that for „average" was 4.8; that for „good" was 6.5; and that for „excellent" was 8.5 faultless operations per minute. Accordingly, the norms for the conversion to marks are:

fail	0 - 2.1	faultless operations/minute
fair	2.2 - 3.8	"
average	3.9 - 5.7	"
good	5.8 - 7.5	"
excellent	7.6 -	"

The spread, however, was so great that there were schools where even with 9 faultless operations per minute the students received a grade of „fail", while on the other hand there were also schools where even 1 faultless operation per minute was sufficient to obtain a grade of „excellent"! Thus, there is complete chaos in the awarding of the gradings. This is also shown by the following figure /Ágoston-Nagy-Orosz, 1971, p. 107/.

Figure 1.



In another measurement /Orosz, 1970/ we studied what percentage of the students obtained the individual grades according to the norms produced by averaging. This is shown by the following table:

Table 1

Grade	Percentage of students attaining grade			
	Primary school class 5	Primary school class 8	Secondary school class II	Secondary school class IV
Fail	31.5	22.9	42.9	37.4
Fair	5.6	24.1	7.4	9.2
Average	16.7	9.3	10.0	10.7
Good	22.2	17.0	3.7	14.8
Excellent	24.0	26.7	36.0	27.9

It can be seen from all this that there is no unity in the determination of the grades, even though the basic principle for this is provided in the integrated school-system.

As a final conclusion: the method of averaging is not suitable for making the evaluation objective.

c. In our Department we have developed a procedure for converting to marks by the <u>method of standard deviation.</u>

Our measurements have confirmed that the well-known law of normal distribution is also valid for the distribution of the measured results; according to this law 68 % of the elements can be found within the standard deviation, while 32 % /in proportional distribution/ lie in positive and negative directions outside the standard deviation.

From this regularity there arises that natural method of conversion to marks, according to which performances lying outside the standard deviation in the negative direction are graded as „fail", and those outside it in the positive direction as „excellent"; the region of the standard deviation is divided into three equal parts to give the intervals of the „fair", „average" and „good" grades.

In one of our measurements /Orosz, 1971/ the average of the national performance was $\bar{x}$ = 56.8 %p, and the standard deviation was s = 23.1. The range of the standard deviation was therefore from 56.8 - 23.1 = 33.7 %p to 56.8 + 23.1 = 79.9 %p. A grade of „fail" is thus awarded to those students whose performance is poorer than 33.7 %p, and „excellent" to those exceeding 79.9 %p. The range of the standard deviation = 2s = = 2 x 23.1 = 46.2. This is divided into three equal parts, i.e. ranges of 15.4 %p. The range of the „fair" grade is therefore from 33.7 to 33.7 + 15.4 = 49.1 %p; the „average" grade from 49.2 to 49.1 + 15.4 = 64.5 %p; and the „good" grade from 64.6 to 64.5 + 15.4 = 79.9 %p.

In such a way we have obtained a mathematically based procedure with the help of which the conversion to marks can always be carried out in accordance with the real situation.

From a pedagogical point of view, however, this method can be criticized. In every case about 16 % of the students would receive a „fail", independently of whether the national results are good or bad. Experience of our measurements shows that it may happen that the average of the performances of the students is higher than 70 %p, and the standard deviation is lower than 20 %p. Thus, even those students attaining a performance of 50 %p would receive a „fail", and this is clearly too severe a requirement.

For this reason /and because of other technical considerations/ the limits of the „fail" and „excellent" grades are taken not as the simple extent of the standard deviation but as one and a half times the standard deviation, while the ons and a half times expanded standard deviation-range is divided into three parts to give the „fair", „average" and „good" grades.

Experience has also shown that a really normal, uniform perfomance distribution is obtained if the average is around $\bar{x}$ = 50 %p. In the case of an average significantly lower /higher/ than this, the distribution curve exhibits slant and asymmetry on the left /right/ side, i.e. in the negative /positive/ direction, and this distorts the scale of the conversion to marks. In order to eliminate this, a slant correction coefficient calculated with the help of the median is taken into consideration.

On this basic principle we have constructed a universal table for conversion to marks /Nagy, 1970/ for all those cases where the measured average is between 10 %p and 90 %p. Theoretically and practically, national averages lower or higher than this are not possible.

The proportions of the performances lying in the grade-ranges thus obtained are generally /with slight deviations/ the following:

fail	7 %	of the students
fair	24 %	"
average	38 %	"
good	24 %	"
excellent	7 %	"

We have thus arrived at a uniform method of evaluation which conforms to the real situation, has been mathematically confirmed and is pedagogically acceptable and objective.

As indicated earlier, however, this system is associated in principle by the danger that the level in existence at the time of the national representative measurement will be made the general requirement level, and as a result the evaluation will become rigid and development will be impeded.

Practice shows, however, that the mere possibility of objective comparison to the national level spurs the teachers to better work. Those teachers whose classes do not reach the national average feel ill at ease, and strive to carry out their work so that their class rises to the level of the national average. In this way of course the national level itself also rises in the course of the years. It is sufficient therefore to repeat the national representative measurements every 3 - 5 years, and the new level indicators will ensure constant and continuous development.

5. The measurements elaborated in our Department are divided into two groups. One of these is known as the theme-concluding measurement of the level of knowledge, while the other is the type of measurements determining the development of the activities.

a. The essence of the theme-concluding measurement of the level of knowledge is that the knowledge of the subject-matter for the whole year is checked after breaking it down into thematic units.

Why did we select this type?

As pointed out in the Introduction, in accordance with our aims the basic characteristic of our measurements is totality. It is easy to see that it is not possible to accomplish the total checking of the entire matter taught in a year in any subject by means of a single measurement. Because of the extent of the subject-matter and the large number of facts taught, the time required for such measurements would be so long that this would undoubtedly result in considerable tiredness to the students. Consequently, the level of the performances would reflect not the extent of their knowledge, but rather that of their tiredness.

It is necessary, therefore, that the knowledge be studied by breaking down the subject-matter into parts. This is justified from another aspect by the aim of the measurement given earlier that it should give a continuous report of the course of the teaching, and so would permit the regulation of the process. For this reason it is justified to obtain total information after the teaching of every individual part of the material forming a logically and structurally natural unit, a so-called thematic unit: the further teaching work can thus be directed in accordance with what inadequacies have been observed and what shortcomings must be eliminated before the teaching of the following thematic unit is begun. It is extremely rare for the individual thematic units to be completely isolated and independent from the other parts of the subject-matter: the elements of knowledge acquired in a given unit are normally necessary as a means of mastering the knowledge of the subsequent thematic unit.

The individual themes of the subject-matter are in the main of such an extent that the natural organizational unit of the teaching, one lesson, is sufficient to check the knowledge. It was made a definite principle, therefore, that no single measurement should exceed one teaching lesson.

In this time /45 minutes/, however, it is not possible to question every student on his knowledge of every fact in the thematic unit. Hence, test-variants are prepared; there are at least four of these, and more depending on the extent of the subject-matter of course, but at most eight. The preparation of more variants than this would complicate the measurement, and would also endanger the reality of the results. So few students within one class would attempt each variant that as a result the teacher would not obtain a realistic picture of the knowledge of the entire subject-matter.

The preparation of the test-variants also has the advantage that no two adjacent students are working on the same variant, and any attempt at cooperation is therefore pointless and has no result.

As has been pointed out, the variants between them contain the whole of the thematic unit, and so their analysis gives a comprehensive picture of the entire knowledge of a single class. In representative measurement each variant is attempted by a sufficient number of students for each variant to provide representative samples too. Consequently, a complete picture is obtained of the national level of knowledge.

b. The main characteristis of measurements establishing the development of the activities /in brief: development measurements/ is that they are used to study the development of activity in an independent ability or proficiency over the entire vertical cross-section of the school-system, i.e. in every class, beginning from the year in which it is first taught, right up to the end of the scholastic teaching, the years in which the students take their final school examinations.

The measurements are therefore complete from two points of view: they are quantitative, since they extend to every element and operation of the activity measured, and they completely embrace the dynamics of the development.

In the measurement of some slowly developing activity /such as correct spelling, or composition/ it may be disregarded that measurements are made in all individual consecutive classes: it is sufficient to study the development of the activity in definite cross-sections /e.g. in classes 5 and 8 of the primary schools, and classes II and IV of the secondary schools/.

Studies into development are generally very time-consuming; longitudinal researches require many years. In the case of the measurements, however, we may neglect the usual investigation procedure following the development in its course with time in a given group. In the case of the use of a curriculum which is unchangingly valid for a prolonged period, a representative measurement carried out by the same means at the same time for each year accurately reflects /in accordance with probability theory/ the realization of the curricular requirements. Hence, our measurements to establish the development of activities are carried out at the same time in all classes.

Considering that this type of measurement too is complete as regards content, it may also occur here that we can solve every problem in the time available only if the students are given test-variants. In accordance with the above-mentioned considerations, at least four variants are prepared from these tests too.

As has been stated, with the tests it is desired to determine the level of development of the activities. In addition to the quality, however, an important role is generally played in the activity by the time factor. In the establishment of the levels of most activities, therefore, we use not only a qualitative index but a tempo index too.

It has already been pointed out that the quality is in general expressed by the ratio of the incorrect and the correct elements, or by the ratio of the number of incorrect elements to the total number of elements, if the performance elements are homogeneous. In the case of heterogeneous elements weighting is performed. In essence, this procedure is also applied in the calculation of the qualitative index of the activities, with the difference that here the performance elements are so-called elementary operations.

The tempo index can be given as the number of elementary operations carried out in unit time. Experience indicated that one minute is the most appropriate time unit in the determination of the tempo. The tempo index therefore is: operations/minute.

Since the tempo index represents an average performance, for the calculation of this average it is necessary for the students to work with a sufficient number of problems for a sufficient period of time. The problem, therefore, is the determination of the number of problems of each type to be set and the amount of time to be devoted to their solution. Two possibilities arise.

First: For every student separately we note accurately the time required to solve the fixed number of problems. However, this is extremely difficult, because it cannot be entrusted to the students, while it is not possible for the teacher to measure every child separately. The situation is made even more complicated in this case by the fact that for every type of problem it is necessary to wait until every student solves every problem. Because of the slowly working students, however, this is accompanied by the danger that the measurement can not be completed in one lesson, while the faster workers begin to solve the problems of the following type without waiting for instructions, and hence the measurement of time becomes impossible.

Second: The time for working on each type of problem is determined in advance /e.g. 4 minutes/, and after the elapse of this time the teacher instructs the students that they schould proceed to the solution of the following type of problem. In this way the period of time is the same for every student, and the problems completed can be counted on the tests. /This can also be done by the students themselves at the end of the lesson, and they can note the number of problems solved per type of problem./ In this case of course there must be enough problems of each type for the fastest working student not to be able to solve all of them in the given time.

From the indeces of quality and tempo it is practical to calculate a combined index, which can be expressed as the number of perfectly completed operations per minute.

The tempo is not an equally important determining factor in every activity. In the solution of very involved mathematical problems or in composition, for example, it is by no means as important as, say, in elementary calculation operations or in ability to read or write. The tempo of the activity generally has a determining role in those simple abilities which function as media in the complex, higher-order activities. From another aspect: the tempo of these higher-order activities would be very difficult to determine directly, since it is almost impossible to separate it from the abilities functioning as media. For example: if we should wish to express the tempo of composition or correct spelling by the number of operations performed per minute /e.g. words written/, then not only the speed of the composition and correct spelling would appear in this index, but also the tempo of the ability to write, in that the data referring to the tempo of the composition would in effect arise in many cases exclusively from the insufficient development ot the writing tempo; i.e. the activity the level of which it was desired to determine would not be the one which would be described. In this way misleading data would be obtained.

The situation is similar in the case of higher mathematical problems too. The main role in the fact that a student solves a quadratic equation very slovly, for example, may well be that the tempo of carrying out the elementary operations /multiplication, division, raising to a power, extraction of roots/ has not developed sufficiently.

The tempo as an index of the level of development, therefore, is worthy of calculation only in those more simple activities, in which it can be established in an adequate way. As to which are the activities for which it is possible to calculate this index, this depends decisively on the structure of the activity. More will be said of this in the following Chapter.

III
PRACTICAL PROBLEMS OF THE MEASUREMENTS

1. Preparation of the theme-concluding tests

This work is divided into four stages. These are: analysis of the material to be learnt in the thematic unit, construction of the tests, the preliminary checking, and finally the standardization.

a. Analysis of the material of the thematic unit. First step: to take into account the concepts /target ideas/ to be taught. This is done simply by writing out the concepts from the teaching program or from the text-book.

It is necessary to distinguish two types of concepts. The first is the group of so-called individual concepts, and the other the group of sets. It is well known that the logical structures of these two types differ fundamentally from each other.

The second step of the analysis of the thematic unit is the discovery of the structures of the concepts.

The extent of the individual concept is formed by the facts reflecting the properties of this concept. The structure of the concept is displayed in the logical relations of these facts. Examples of individual concepts in geography for instance: a continent, a country, a concrete region; and in history: every individual historical event. Thus, the individual concept might be Hungary; the facts reflecting its properties are: its geographical situation, its climate, its relief, and the data determining these, etc. The individual concept might be the First World War; and the facts reflecting its properties: its beginning, its end, its duration, the opposing sides, the individual battles, etc.

The number of facts relating to the individual concepts is very high, and in principle is infinite in most cases. It is not possible, therefore, for every fact relating to an individual concept to be taught. In the case of concentric arrangement of the subject-matter, an ever increasing number of facts are taught by re-teaching the concepts in higher classes.

As regards the measurement of the level of knowledge, the analysis of the individual concepts finishes with the taking into account of the facts to be taught in the given year.

The sets /multiple concepts/ include within their scope the partial sets and the facts reflecting the properties of the set. The number of facts reflecting the properties of sets is generally not large, since these are in effect the common properties of the individuals belonging to the set, or the partial sets; ab ovo the number of common properties can only be less than the whole of the specific ones of the individuals. Naturally, the data to be taught are also the facts reflecting the properties of the partial sets.

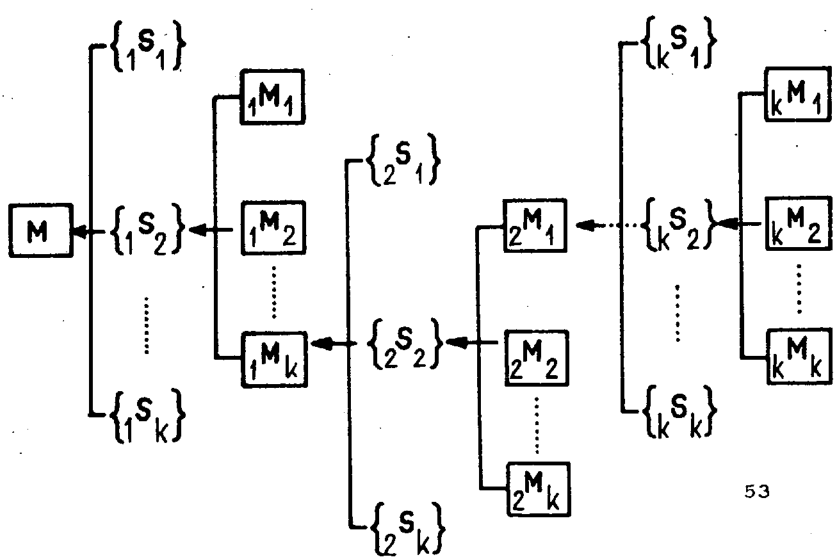
The number of partial sets is also limited.

In lower classes in general we do not teach every property and every partial set of the sets. With concentric arrangement of the subject-matter, on re-teaching in the higher classes the instruction in the properties of the sets and the partial sets will become increasingly complete.

It follows from what has been said that the structures of the sets are more complex than those of the individual concepts. The structure is complicated by the fact that it also involves the divisional bases: the so-called system-forming concepts. For just this reason the discovery of the structures of the sets is carried out in two stages: we first prepare the division of the concepts, and next take into account separately the facts belonging to each concepts. For the sake of clarity it is fitting to plot the division of the concepts. The plotting throws light on whether there is an excess concept or a break in the system.

The Figure will be more simple or more complex in accordance with whether complete or incomplete division is performed, according to one or more divisional bases, and whether one or more layers /levels/ of the partial sets of the set are discovered. The general scheme of the division can be seen below /Nagy, 1970/:

Figure 2.



Symbols used:

M = set

S = divisional basis, system-concept

The numbers before the large-letter symbols represent the level of the division on which the concepts are, while the number after the symbol is the serial number of the concept; the final member of each series is denoted by „k", in an attempt to express that all the series involved are finite.

After the structure of the multiple concept has been discovered, there follows the discovery of the extent of every concept in that structure. This means taking into account the facts relating to the concepts. Particular attention must be paid to whether the book gives a regular definition of the concepts. If this is the case, then it is simple to take the facts into account, since the definition contains concisely the most necessary data, and at such time all of the facts to be taught.

The third step in the analysis of the thematic unit is to take into account the laws, principles and operations.

According to their structures these pieces of knowledge all consist of logically related series of descriptions, i.e. judgements. It is very important to discover exactly the logical connection between the judgements. Text-books today do not yet devote sufficient attention to the precise linguistic formulation of these. It happens fairly frequently that a connecting conjunction is used to link judgements between which there is not a conjunctive, but a disjunctive logical connection. In other cases, because of the omission of the conjunction the logical connection is not clear.

It is true that the writers of text-books should think more about these problems, but we cannot neglect them either in the preparation for the measurement: the deficiencies of the text--books often become known from the results of the measurement themselves. The qualitative analysis of the measured results shows that the reasons for the weak knowledge of the students are frequently such text-book deficiencies.

As the fourth step the activities are taken into consideration. This is a special problem of the theme analysis, since our teaching programs in general do not list the activities to be taught in an organized way. The starting point here too is the teaching program /and primarily its instructional part/, but we must rely more strongly on the text-book, and on that part of this containing the problems. In addition to these we must also study the teacher's auxiliary books.

Even today very little can be said about this problem in general. The characteristic features of the activities are very different from subject to subject. The problems must be studied very thoroughly.

Merely as an example, it will be mentioned that four types of activities have to be taught in the thematic unit „Fundamental ideas and laws of chemistry" in the chemistry teaching to class I of the secondary schools: 1. calculation of molecular weight, 2. writing of structural formulae, 3. writing of equations, 4. „preparation of compounds" by solution of equations. Again: the thematic unit „Introduction to syntax" in the teaching of Hungarian grammar to class 5 of the primary schools contains a total of 17 activity forms in seven activity types. These are: 1. systematic division /breaking down compound sentences to simple sentences, and unpunctuated text to sentences/, 2. supplementation /insertion of punctuation and interpunctuation, addition of omitted elements of proverbs, addition of missing case-endings on the insertion of given words into sentences/, 3. analysis /simple sentences according to content, simple sentences according to structure, simple sentences according to sentence-parts/, 4. sentence expansion /by given sentence--parts/, 5. sentence construction /sentences of given content, simple sentences on a given theme, sentences with given sentence-parts, sentences with adjectival constructions according to choice/, 6. formation of structures /addition of various attributes to a given word, arrangement of adjectival constructions from given words/, 7. explanation of spelling phenomena.

The examples confirm that the activities differ considerably from subject to subject; their typization according to subject, however, may give a basis for didactic generalizations of a more far-reaching nature. Our experience to date shows that in general two fundamental types of problems are to be taught: those of operative and cognitive problems. Problems are termed operative, in which a change must be brought about in the object of the activity. /Of course, the objects may be symbols too, e.g. in mathematics, chemistry and physics./ In contrast, problems are termed cognitive where it is not necessary to produce a change in the object of the activity. These problems can be solved by the logical operations of arranging, systemizing, the construction of systems, etc.

After the activity has been taken into account the requirement levels must be determined. For example, molecular weight can be calculated with or without atomic weight tables. We must decide, therefore, on the way in which we wish the problems to be solved.

b. Determination of the types of problems on the tests

It was mentioned earlier that our aim with the measurements is to ensure objective evaluation, and to be able to analyze the mastery of the thematic unit as regards content.

Since the nature of the knowledge to be measured is very varied, it is fairly difficult to find the appropriate question and problem forms. The alternative and multiple-choice problems widespread in teacher-made tests are not suitable for the measurement of all types of knowledge at every level. There is a need for those problems which conform to the structure and levels of the knowledge. We have developed the so-called combined question technique, which adapts itself to the type and nature of the knowledge.

Several problems had to be solved.

First: an adequate compulsion of answer had to be ensured. Thus, the question or instruction has to be composed so that the student answers only what is asked of him. All types of problems which contain precomposed optional answers satisfy this condition. But, as we have indicated, for reasons of principle such problems are used only rarely, and therefore the informative - instructional parts of the problems must be composed with particular care. Trial measurements show only whether we have succeeded in ensuring the adequate compulsion in the answer.

Second: the objectivity of the evaluation can be assured only if the expected correct answers are first broken down to alternative elements, which can always be described unambiguously. We must also ensure that these elements can be uniformly separated in the freely composed answers. This aim is served by coding, two fundamental means of which are used: coding on the test-paper and with the correcting key. We code on test-paper when the test provides the possibility of identifying the lettering and the percentage point with the corresponding alternative elements. We now give an example of coding by test:

Name one animal belonging to each group:

a. mammal: . . . .

b. bird: . . . .

c. insect: . . . .

a	b	c

This coding has the advantage that the teacher can carry out the marking in a meaningful way, and hence his work is rapid and easy.

Coding with the correcting key is used for those problems where the identity of the code symbol and the corresponding alternative element is not possible on the test, either for the reason that the answer is freely composed, or because the students can give the elements of the answer in the order of their own choice. For example:

Diffusion is defined as . . . .

. . . .

a	b

In the correcting key: a. molecules

b. their mixing

It is not important in this case how the student composes the answer; the vital thing is that it should contain the two alternative elements listed.

An example for the case when the order of the elements of the answer can be selected at will:

List the main parts of a tree.

. . . .

a	b	c

In the correcting key: a. roots

b. trunk

c. crown

In the commonly known technical literature closed and open questions are distinguished according to coding and codability. The closed questions are coded and the open ones not, since the answers are not known. It can be seen that by breaking down to alternative elements all types of problems can be coded, and so the evaluation may be made objective.

Third: The objectivity and content analyzability of the measurement are affected decisively by the types of the questions and problems. As already pointed out, a combined question technique is used. We have elaborated the possible question types and question forms. /Question and problem are of identical meaning in such a respect./

Corresponding to the content of the problems, four question types have been established. These are:

Factual question. With these we check the knowledge of the individual facts.

Fact-connecting question. To check not only the knowledge of the individual facts, but also that of the connections between them.

Cognitive problems. These check the cognitive use of the knowledge explained above.

Operative problems. With these we evaluate the operative use of the knowledge.

Ten question forms are generally employed. These are:

Ouestioning.

Instruction.

Table, table-type problem.

Supplement: verbally,
naming a given picture /figure/,
depiction of a given object,
supplementing a given figure by drawing,
supplementing a figure by drawing and with text.

Choice of answer.

Solution of problem.

Even if only the four question types and the ten question forms are taken into consideration, it is obvious that the questions on the tests can be assembled in a large variety of ways. And if the contents of the individual problem types are given in detail too, then the number of possibilities of variation increases still more. As an example, let us try to take into account the types of facts, and hence list the variations in the forms of factual questions.

The facts connected with the individual concepts are judgements which involve each of the properties of the individuum and which designate the individuum. Our factual questions can aim at the designation of the individual properties and the concept.

Let us consider the most important of the possible variations.

Designation of one or more properties of a given individuum

By questioning:

What is the name of the capital of Canada?

. . . .

a

By instruction:

Name the highest peak in Europe.

. . . .

a

By verbal supplementation:

The constitutional form in Austria is . . . .

. . . .

a

By supplementing a drawing with a name:

/e.g. a photograph of the Colosseum/

What is the name of this building?

. . . .

a

By supplementing a given figure by drawing:

Put in the capital on an outline map of Austria.

a

By supplementing a given figure by drawing and with text:

Draw the right-hand tributaries of the river Tisza on an outline map of Hungary. Write the names of the rivers on the map.

a	b	c	d	e	f	g

By choice of answer:

Which is the capital of the USA? New York
Chicago
Washington
San Francisco

a

By table:

Hungary	
Its area	a. . . . . km^2
Its population	b. . . . . million
Neighbouring countries	
To the north	c. . . . .
To the north-east	d. . . . .
To the east	e. . . . .
To the south	f. . . . .
To the west	g. . . . .

a	b	c	d	e	f	g

Designation of the individuum on the basis of given properties

By questioning:

Who was the Russian general who joined battle with Napoleon at Borodino? . . . .

a

By instruction:

Name the town on the river Tisza which has a university and important industry. . . . .

a

The facts relating to the sets form a more complex system, and so there is much more variation in the questioning to check them. Let us disregard the illustration of variations and examples which depend on the question forms. We shall attempt merely to list those question combinations which come about from a consideration of the contents and structures of the sets. These are:

designation of one or more properties of a set;
designation of a set on the basis of given properties;
designation of the divisional basis on the basis of given properties;
listing of the elements of the set: complete list,
incomplete list;
designation of the set on the basis of listed elements;
designation of the divisional basis on the basis of listed elements;
listing of partial sets: complete list,
incomplete list;
designation of the complete set on the basis of the designation of the partial sets;
designation of the divisional basis on the basis of the listing of the partial sets;
listing of the divisional bases.

The fact-connection questions have many variants, since there are several types of connection between the facts: logical, spatial, time. Depending on the nature of the connection, the order of the facts may be of decisive importance as regards the knowledge. The knowledge containing fact-connections includes definitions, laws and rules. Fact-connection questions are mainly coded by correcting key. We shall consider only a few examples.

What is oxidation? . . . .
. . . .

a	b	c

The drafting of the answer may be optional. In the correcting key, however, we give the facts essentially required: a. substance b. combine c. with oxygen

Give the law of constant weight proportions. . . .

. . . .

. . . .

a	b	c	d

In the correcting key: a. in compounds
b. the constituent parts
c. weight ratio
d. definite

Let us look at a typical question in which the order of the alternative elements is of vital importance:

Give the correct sequence of cultivation of the grape.

a. . . . . b. . . . .

c. . . . . d. . . . .

e. . . . .

a	b	c	d	e

This type must be dwelt on separately because its evaluation is special. It is in vain that the student well describes the individual alternative elements: if their sequence is incorrect, then the entire answer is wrong, and so every element must be regarded as incorrect.

With factual questions and fact-connecting questions we can only make the students reproduce their knowledge. With the cognitive and operative problems, on the other hand, we check their ability to use the knowledge.

As regards the <u>cognitive problems</u> it should be noted that their most frequent form is answer-selecting, in which the student must always indicate several answers. These answers are mainly concepts, and their selection requires a particular logical operation, in general their classification into some system. Of course, other forms too are employed. The following is a typical example of answer-selection:

Of the following changes, underline those which are chemical transformations.

a. the rusting of iron
b. the evaporation of water
c. the breaking of a glass beaker
d. the burning of sulphur
e. the turning sour of milk
f. the fermentation of grape-juice

a	b	c	d	e	f

Next follows a tabular problem form where the solution is achieved by other than answer-selection:

Fill in each of the columns of the table for each of the verbs, showing in what form of conjugation it is.

Verb form	What mood	What tense	What number	Which person
nézem	a.	b.	c.	d.
vinnének	e.	f.	g.	h.

a	b	c	d	e	f	g	h

The <u>operative problems</u> too can often be solved in several ways. Because of this their breaking down to alternative elements and coding gives rise to a fair number of difficulties. This problem is still not solved: further research is required. It is a general principle that if a problem has not more than two or three solution possibilities, then every possibility is broken down to the same number of elements, and in addition to the individual code symbols every correct solution possibility is given in the correcting key. In the event of more solution possibilities, the alternative elements corresponding to the most frequent variants are given.

Even in the knowledge of this general principle the breaking down to alternative elements can only be carried out on the basis of the nature and construction of each individual concrete problem.

Let us consider an example.

Solve the following equation on the basis of the relations of the operations, and check the answer by substitution.

$$\frac{x}{1.1} = -6$$

a	b	c

In the correcting key: a. $x = -6 \cdot 1.1$

b. $x = -6.6$

c. $\frac{-6.6}{1.1} = -6$

Alternative element a. can be accepted even if the multiplication is not given in the written-out form, as long as alternative element b. is correct. In this case the student receives the %p value for both elements.

Another example:

Write the structural formula of H_2CO_3.

a	b	c

In the correcting key: a. O with a double bond to C

b. two O's with single bonds to C

c. H's with single bonds, each to a different O

Operative problems in grammar are all types of activity such as declension, transformation, sentence construction, etc.

It should finally be noted that we frequently employ more complex problems, in which the various types of questions and problems appear in different combinations. These still further extend the variational possibilities. They also give the advantage to the compiler of the test, that the information given in the initial question simplifies and shortens the drafting of the related problems.

From a practical point of view, however, it is advisable to compose the more complex problems in such a way that they do not consist of too many alternative elements.

The types of problem which can be used in the measurement are also affected by the nature of the subject concerned; the types reported, however, can be employed in almost all subjects, even though not all on a single test--paper.

c. Compilation of the tests

This work is comprised of several stages.

First: it must be decided how many variants the test will consist of. In this, consideration must be given to the number and difficulty of the problems.

A simple practical procedure is used in the estimation of the number of variants: one or two subject teachers are asked to solve all the problems, and in such a way that they too write down all that the students must later. If this work does not last more than 35–40 minutes, then it is sufficient to separate the problems into four variants, because the average student too can solve every problem in each variant within one 45-minute lesson. If the teacher works for 40-60 minutes, then 5-6 test variants are necessary. If the work requires even more time, then it should be considered whether to compose a larger number of test variants, or to divide the thematic unit into two parts. One factor against more variants is that in this case very few students would solve each variant within one class; the separation of the thematic unit into two can be criticized on the grounds that this may interrupt the logical unity of the theme.

Second: the problems are arranged into questions of fact, fact-connecting questions, and cognitive and operative problems. All four types of problem are then divided into as many parts as the variants planned. By this it is hoped to ensure that equal numbers of problems of each individual problem type will appear on each test variant. In addition, this also serves to ensure that the extents and difficulties of the variants correspond.

With a little practice the number of problems can already be planned in advance so that only the factual questions and fact-connecting questions require division, the cognitive and operative problems being prepared in such a way that as many of each type are available as the number of test variants expected to be necessary. The type, number of elements, and structure of these must be in perfect agreement, and they can differ from each other only in their concrete contents and data.

Third: provision is made for voluntary problems, so that the faster working, better students too can work throughout the entire lesson.

Fourth: control of the tests is necessary. Such control is repeated. The compilers must first carry out a thorough technical and logical control, and the test must then be looked through by specialists who are familiar with our measurement methodology, but who did not take part in the compilation.

The errors detected must be carefully corrected.

The tests are next duplicated for the trial measurement. It is this trial measurement which constitutes the most important control.

The trial is organized in such a way that the problems are given for solution to the students in one class of each of 4-6 schools; we thus see the results of the work of 40-50 students per variant. It can be accepted as a rule that at least as many classes must take part in the trial measurement as the number of test variants prepared.

The classes participating in the trial are selected so that there is one good and one weak class among them; it is in this latter that possible schortcomings of the test come to light, among others whether we have succeeded in ensuring an adequate compulsion to answer in every problem.

Care must be taken regarding a suitable measurement guide for the trial, so that the performance is uniform in all respects. The teachers are additionally asked to write or have written on the test of each student the time required to required to answer the compulsory problems. In the case of those who also answered the voluntary problems correctly, the time required for this is written separately on the tests. It is thus established whether the compulsory questions can be answered within one lesson, and whether the tests contain sufficient voluntary questions.

Experience shows that it is necessary to divide the problems on the test into several variants if the average working time devoted to the solution of the compulsory questions is more than 35-36 minutes. Extra voluntary questions must be provided if there are students who had no further questions to answer during the final 4-5 minutes of the period.

The trial measurement shows whether the breakdown to alternative elements was correct, whether the coding was satisfactory, whether the correcting key can be employed, and whether the variants are of equal difficulty.

Much care must be taken about this latter problem. The break-down to test variants is justified in the establishment of the national level only in the case of the variants being of identical difficulty. This identity of course is to be understood within definite limits of tolerance.

On the basis of theoretical considerations and practical experience it has been established that the test variants are to be considered as of identical weights if the measured average result of the individual variants differs by at most 6 % from the average of the variants.

2. Preparation of tests measuring development

The knowledge taught at school is employed in various activities. These activities too must be taught. The means of their development is practice. There are extremely important didactical questions, such as how long they may be developed, when the practice is sufficient, when there is a need for further practice, etc.

As has already been mentioned, the activities to be taught can be separated into two general types: cognitive and operative activities. The tests serving this aim are compiled as follows:

consideration of the operative activities and break-down into operations;
determination of the activity levels;
compilation and testing of the measuring device; elaboration of the evaluating system.

In the following these stages will be considered in turn.

a. Consideration of the operative activities and break--down into operations

Above all we must establish the nature of the operative activities in which the material prescribed in the syllabus is to be applied, and the operative activities to be taught and practised must be listed formulawise and taken into account. These may be very different within an individual subject. In what follows we shall illustrate the essence of the consideration by means of examples. These will be taken from three different areas: one quite simple one from arithmetic /elementary calculating ability in primary forms/; one more complicated one from the operative activity system of correct spelling; and one quite complex one from the area of composing activity.

aa. Operative activities in elementary arithmetic /elementary calculating ability in the numerical range of 100/

The following types of operative activites must be taught:

addition operations in the numerical range of 10,
including the exceeding of 10 /of the type $a+b=x$/;
subtraction operations /of the type $a-b=x$/;
four types of supplementing operation /of the types
$a+x=b$, $x+a=b$, $a-x=b$, $x-a=b$/;
operations of the multiplication tables;
operations of the division table.

Altogether therefore there are 5 types of operative activity, and within these 8 types of operation in this elementary ability system. Since 10x10 numerical combinations are possible in each operational type, we must measure the level of development in a total of 800 operational performances.

ab. Operative activities of correct spelling

Correct spelling in Hungarian is based on three principles. These are:

the pronunciation /phonetic principle/;
the etymology /etymological principle/;
tradition /the principle of tradition/.

Accordingly, three types of activities are to be distinguished. The 10th edition of „Rules of correct spelling in Hungarian" describes 438 rules within these three types. Each individual rule is an operational type. A problem arises in that there are exceptions to many of these rules; these can in part be included in further rules, but in part are quite individual and hence the complete consideration of the operations is impossible.

There is still another reason for us to give up the complete consideration of the operations. This is that it is not possible to prepare a measuring device which contains every operation.

As regards correct spelling, therefore, it is necessary to carry out the analysis from other points of view. It appeared suitable at first to distinguish between the activities related to unambiguous, readily teachable rules, and those which can not be described by such unambiguous rules. The rules based on the principle of etymology by and large describe the activities unambiguously, whereas the rules based on the phonetic principle and tradition in general do not.

As a result of practical considerations, the operations of correct spelling linked to unambiguous rules are divided into 8 groups, containing a total of 28 operations.

There is a total of 6 operations in correct spelling which can not be related to unambiguous rules. The operations which can not be classified into either of these categories are placed under the heading of „Others".

It can be seen that we are immediately forced into a compromise, as complete consideration must be abandoned. Unfortunately, every compromise increases the possibility of measurement error.

ac. <u>Operative activities in composition</u>

The situation here is much more complex than in correct spelling. This is in part due to terminological reasons.

In a broad sense composition is the activity in the course of which the individual doing the composing uses the language /either verbally or orally/ to express the content of his personality, or an actual part of this depending on the theme of the composition. In general use the word „composition" means alike the process, the activity and the product. In the interest of approaching the object of the measurement, however, we must distinguish between these.

Let us name the „ready product", that is the personality content objectivated by means of the language, as the composition. The functioning of the expressing means-system is termed the composing activity. The operational system of the composing activity is collectively named the composition technique. Thus, the personality content in the composing activity is mobilized and objectivated by the functioning of the composition technique.

The composing activity is a dialectical process. Within it are two factors which act on each other mutually and inseparably from each other: these are the personality content to be expressed and the expressing technique. The functioning of the technique is initiated by the range of will of the personality /on the action of external or internal motives/, and the technique mobilizes the actual elements of the personality in accordance with the theme. The personality content is formed in the composing activity, and the technique too is developed by the functioning.

The unity of content and form is a well-known fact, and within this the determining function of the content in general. Linguistics and psychology research workers have revealed the unity of the consciousness content, thought and linguistic expression. The determining function of the content does not mean a function-like determination, nor therefore that a certain increase in the consciousness content /information uptake/ entails an equivalent development in the expression technique. For this reason it is necessary to teach the composition technique separately. From another aspect: the development of the means of expression assists the interiorization and integration of the information. Hence a very important part is played by the expression, and by the technique of composing the thoughts, in the acquisition of knowledge, and in the whole process of studying.

All this justifies why we deal with the measurement of the composition technique, but at the same time it also indicates that the consideration of the operations of this activity differs from the consideration of the operations of the activities so far reported. Let us neglect here an account of the complete operational system of the composition technique. We shall make do merely with a list of the operations. These are as follows:

1. Collection of the factual material belonging to the theme.
2. Selection of appropriate concepts.
3. Formation of correct judgements.
4. Connection of decisions, conclusions, argumentation.

5. Selection of the principle of compilation, and corresponding arrangement of the material.
6. Division into sections and paragraphs.
7. Construction of the complete sentences.
8. Formulation of the structures /syntagmae/ within the sentences.
9. Harmonizing within the sentence and the section.
10. Assuring of logical and grammatical consistency.
11. Selection of the correct comparative elements.
12. Selection of the appropriate words.
13. Construction of the correct word-oder.
14. Correct use of the suffices.
15. Use of suitable pictorial expressions.
16. Selection of words.
17. Selection of sentences.
18. Elimination of word-repetition.

These operations can be further broken down into elementary operations. Complete consideration is not possible here either. Instead we have strived to give an unambiguous formulation of the criteria for listing in operations.

b. <u>Determination of the levels of activity</u>

It is expedient to distinguish three levels in the mastery of the operative activity.

ba. <u>Level of the external algorithm</u>

This is the initial stage of the study of the activity. On this level the student obtains the information relating to the course of the activity, and the rule and algorithm of the activity from outside: verbally or in writing, or in an imitable model. Becaues of this he carries out the activity very slowly and with relatively many mistakes.

bb. <u>Level of the internal algorithm</u>

The student has learned the algorithm of the activity, and has stored it away in his memory. He can carry out the activity independently and free from efforts, and without instructions as to the course of the carrying out. Nevertheless, he does not simply only know the rule and the algorithm, but has also acquired a certain experience in carrying out the activity in accordance with the algorithm.

bc. Level of maximum efficiency

This is the highest level. After mastery at the level of the internal algorithm /as a result of practice/ the student still develops considerably in the carrying out of the operative activities: the quality of his performance improves substantially and the tempo quickens. The maximum efficiency means a precisely establishable quality and tempo.

One very important task of the measurement is to establish the qualitative and quantitative indices.

As regards the activities to be measured, it must be decided prior to the measurement as to the level on which knowledge is required. For example, it can be stated of the 8 types of elementary calculating ability that the students must know the operations of addition, subtraction, the multiplication table and the division table at the level of maximum efficiency. It is superfluous, on the other hand, to know all four types of substitution at the level of maximum efficiency, since of these the type $a+x=b$ plays by far the most prominent role: this is needed to carry out written subtraction and written division. For this reason, only this type need be practised to a maximum.

As indicated earlier, the tempo and time-factor appear in a different way in the activities of correct spelling and composition than in the functioning of elementary calculating ability. These two activities are more complex from this point of view too than the elementary abilities. As regards both correct spelling and composition, the curricular requirements are absolute correctness, that is they are maximum.

c. Arranging the means of measurement

Since the activities to be measured are strongly heterogeneous, a general course for compiling the tests can not be established. The compilation of every test measuring an individual type of activity requires a characteristic course of work. This will be demonstrated with the compilation of the means of measuring the three types of activities already introduced as examples.

ca. Compilation of a test measuring elementary calculating ability

First: It must be determined how many tests are necessary to measure the given activity. The number of tests depends above all on the time to be devoted to the measurement. The meassurement can not last longer than one teaching lesson of 45 minutes, and because of the preparation and the distribution of the test-papers not more than 40 minutes remains for the measurement. For other reasons, however, not even as long as this can be assigned to it in general. In the case of protracted work, the solution of elementary operations, for example, gives rise to monotony, a result of which is more rapid brain-tiredness. In our experience, the measurement of one type of elementary operations can not last more than 10 minutes.

Second: It must be determined as to which operations of the given problem type or operational system we wish to extend the measurement. It is worthwhile measuring five operational types of elementary calculating ability, which would result in 500 elementary operations. The operations with 1 and the extremely simple cases can be omitted from the tests, however, and so only 375 of these 500 elementary operations are necessary. Consequently, in each of the 5 operational types elementary calculating ability to be measured 75 operations remain. These must be represented on the test.

Third: The time required to carry out the total operations must be determined. For this a preliminary measurement is made: in the case of the elementary calculating ability this is done by having the students from consecutive years solve the 75 elementary operations from each of the five operational types in separate classes. These preliminary measurements show how much time is required for each operational type, and from this it is established how many operations the students can solve on average in 1 minute. On the basis of these data, it can be decided how many tests should be prepared, and how many teaching lessons should be devoted to the measurement.

A choice had to be made between two possibilities. One of these: each of the students carries out all of the 75 operations in every operational system. In this way, however, separate tests are required for the operational systems of addition, subtraction, substitution, multiplication and division, and this means that the developmental level of elementary calculating ability can be measured only in 5 teaching lessons. It is not possible to devote so much time to the measurement of one ability system. In addition, it has also turned out that in the lower forms not every student is able to carry out all 75 operations in a single type during one lesson. It was necessary, therefore, to give up the idea of having every student solve every problem.

A procedure was sought, whereby the students do not have to carry out every operation, but an accurate picture is nevertheless obtained as to the state and progress of the given ability. This can be attained by setting the time for the problems to be attempted, and during this time the students solve as many of the given operations as possible. On the basis of the data of the preliminary measurement a time is fixed upon, in which even the weakest students can solve 8–10 operations, while not even the fastest working students can solve all 75 operations. These times are as follows:

In classes 1 and 2 of the primary school: 5 minutes per type of operation;

in classes 3, 4 and 5 of the primary school: 2 minutes per type of operation;

in all other classes /in secondary school too/: 1 minute per type of operation.

Consequently, for the three types of operation /addition, subtraction, substitution/ learnt in class 1 of the primary school a total of 15 minutes is to be devoted to the solution of the problems on the test: in class 2 the time for the five types of operation is 25 minutes, in classes 3-5 10 minutes, and in all other classes 5 minutes. In the course of national measurements these time limits have proved to be correct.

Fourth: With the time determined in advance, it follows that the operations appearing at the beginning of the test are attempted by many more students than for the questions at the end, and this distorts the result. For this reason the test--papers are prepared in four variants, in which the order of the questions varies to eliminate this distortion. The four variants lead to the added advantage already mentioned in connection with the variants of the theme-closing tests, that the students can not look at each other's work.

Since it was also desired to learn the state of ability in addition, in subtraction, etc., all five types of operation appeared separately on the tests. Further, a conscious grouping of the operations was made within the individual types into simpler and more difficult groups /e.g. operations not exceeding 10, operations exceeding 10, etc./; this was done in such a way that in general every basic variation schould occur within 6-8 consecutive operations. Thus, even the slowest working and weakest students are confronted with every basic variation.

Fifth: In the interest of unambiguous correction and the facilitating of the work of the teachers, a correcting key is prepared. For the analysis and evaluation a note must be made on the tests of the total number of operations attempted and the number of incorrectly solved problems.

cb. We are faced with completely different problems in the measurement of the ability to spell correctly and of the composition technique.

First: It must be ensured that on the occasion of the measurement the activity is performed in a natural situation. It would not be correct, for example, to measure the state of correct spelling ability in the course of dictation, since the correct spelling must function as a means on the level of maximum efficiency, that is not under conditions where the student concentrates his attention during the writing on the rules of correct spelling. Hence, a composition problem is given to measure the state of correct spelling ability. This requirement is accompanied by the advantage that the levels of correct spelling and composition technique can be established by one and the same measuring device.

Second: It is regarded as a basic requirement that in the measurement of the composition technique everyone schould have something to say on the theme of the composition. Themes can not be given, therefore, which desire the expression of knowledge acquired during the course of school studies.

Third: It is presumable that the performance of the individual operations will exhibit different characteristics in the various types and forms of composition; for this reason, the state of the composition technique must be measured not on a single composition, but on several fundamental types. Thus, one descriptive and one narrative composition and one characterization must be written. This also has an advantage from the point of view of the measurement of the ability to spell correctly: it is more certain that all types of correct spelling phenomena will occur in the three different forms of composition. For example, a descriptive composition can be prepared so that only the present tenses of verbs figure in it; this no longer holds for a narrative composition. In the Hungarian language the correct spelling of the past tenses of the verbs is an appreciable problem.

Fourth: A measurement which is so complicated and requires so much work must be organized in the most economic way possible. On the basis of previous studies it has been established that the rates of development in correct spelling and composition technique are much slower than those of the more simple activities. Hence, it is not necessary to carry out measurements in every consecutive year. After careful consideration, the measurement is made in the four classes shown in the following Figure:

Figure 3

Class	Stage of schooling	Stage of development
primary 5	Middle of primary school studies	End of childhood
primary 8	End of primary school studies	Middle of puberty
secondary II	Middle of secondary school studies	End of puberty
secondary IV	End of secondary school studies	Middle of youth

It can be seen that two of the classes correspond to the middle of a given stage of scholing, and at the same time to the end of a stage of development, while the other two classes correspond to the end of a stage of schooling and the middle of a stage of development. /The stages of development are not to be taken too rigorously./

Since the state of composition technique is related with the level of development of thought, measurement in these classes permits the expectation of conclusions of more general validity beyond the problematics of the composition technique.

One 45-minute teaching period each is set aside for the writing of the three compositions.

Fifth: For the evaluation of the above-mentioned operations of both correct spelling and composition technique we have prepared a detailed description of the criteria, and have composed the necessary analysis sheets. The correction of every paper in correct spelling and composition technique is performed by the same group of teachers in Hungarian. In this way we hope to ensure a uniform evaluation of the criteria.

d. Elaboration of the evaluation system

As already indicated, the state of the operative activities can be expressed with two types of basic indices. One of these is the tempo, and the other the quality. A combined index too may be formed, which express several factors jointly. The calculation of these indices is quite simple for the elementary abilities, but for the more complex activities it requires a fairly involved calculation.

da. Indices of state of elementary calculating ability

These operations can be considered as homogeneous on the level of maximum efficiency, and thus it is not necessary to weight them.

The tempo of a performance composed of homogeneous elements is expressed by the number of operations carried out per minute. This requires simple division: the total number of elements solved is divided by the number of minutes used for the solution.

There are a number of procedures for the measurement of the quality of a performance composed of homogeneous elements. One of these is the correctness proportion, which is calculated by dividing the total number of elements solved by the number of incorrectly solved operations. This index expresses the proportion of operations in which the student makes a mistake on average. It is most practical for the teacher to calculate this and the tempo.

In the establishment of the national level, it is practical to employ an index which contains together values relating to both the tempo and the quality. The following combined index is used to measure elementary calculating ability:

$$O = \frac{n - i}{t}$$

where O = quality

n = number of elementary operations performed

i = number of incorrectly solved elementary operations

t = number of minutes devoted to the performance

This combined index therefore gives the number of operations solved correctly per minute. For example, if a student carries out 35 operations, 5 of them incorrectly, in a working time of 5 minutes, then the state of the ability is

$$\frac{35 - 5}{5} = 6 \text{ correct operations per minute.}$$

Both the tempo and quality of the calculating ability are contained in this index therefore. Because of its advantages it is recommended for general use.

With the procedure given the performance of every individual student can be demonstrated numerically. We dot not know, however, what the numerical values obtained are actually worth. How do they compare with the national performances for the given class, or with the performances of other classes? Nor do these numbers show how the elementary calculating abilities may be further developed, or the maximum which can be attained by man in general. We must therefore determine the national levels for all classes, while there is also a need for the numerical expression of the maximum performance.

The national level for each class and operation has been established by a representative measurement involving the entire vertical profile of the school-system, i.e. for every class from class 1 of the primary school to class IV of the secondary school. The results showed normal distribution everywhere, which permitted the mathematical determination of the maximum performances for each class. The results of the best students can be regarded as the maximum performance at a given age. Mathematically the average of the performances over twice the standard deviation were taken as maximum, that is the average performance of that 2 % of the students who achieved the best results.

Besides the class-maximum, there is also a need for an absolute index which indicates the extent to which the ability may be developed. The average performance of adults with intellectual occupations was taken as this index. The representative measurement was therefore extended to adults with intellectual occupations.

When these indicators have been obtained, the performances of individual students or classes can be compared as follows:

Let us assume that the performance of N.N., a student in class 1 of the primary school, is 3.2 correct operations per minute. The national average for class 1 is 4.8, the class-maximum is 10.6 and the average performance of adults is 50.5 correct operations per minute.

The performance of N.N. compared to the class level is:

$$\frac{3.2}{4.8} \times 100 = 66.7\ \%;$$

compared to the class-maximum is:

$$\frac{3.2}{10.6} \times 100 = 30.2\ \%;$$

and compared to the maximum is:

$$\frac{3.2}{50.5} \times 100 = 6.3\ \%.$$

In order that the teachers should not have too much calculating to do, we have calculated the values of all possible performances for all classes, compared to the national average and to the maximum, and have tabulated them. The performances of individual students can thus be simply read off from the table /Nagy, 1970. Appendix/.

db. <u>Indices of state of correct spelling</u>

We are faced here with different types of problem.

First: The time factor can not be taken into account.

Second: Since the state of correct spelling is examined in compositions, the extents of the performances to be evaluated differ. This is disregarded by comparing the number of mistakes per operation to the extent, which is determined by the number of words written.

Third: The operations are heterogenous, and so must be weighted.

We have made calculations only with an empirical weight. The level in every operation is the same /application/, and this is therefore not a controlling factor. A functional weight was not calculated on the basis of the results of the previous examinations, since this exhibits a strong correlation with the empirical weight.

Fourth: Apart from the operational indicators calculated with the empirical weight and the extent, a combined indicator requiring a fairly involved calculation has been worked out, which gives the state of correct spelling as a percentage value.

dc. Indices of state of composition technique

In principle we proceeded in a similar way in the evaluation of the composition technique as in that of correct spelling. Thus, we calculated only with an empirical weight of the heterogeneous while the quality index was obtained by comparison of the number of mistakes to the extent.

The index of the tempo, however, is of importance from another aspect.

Since the students devoted one 45-minute lesson to the writing of each composition, the tempo of the composition technique too is expressed by the extent and by the time used for writing. Of course, the index of the tempo is not of the same nature as in the elementary calculating ability; in that case the problem was for every student to carry out the same operations given on the test-paper, so that it indeed expresses only the speed of functioning of the ability. In the composition, on the other hand, the number of words written in unit time indicates not only the writing tempo and the speed of functioning of the composition technique, but also how much the students have to say about the given theme. Thus, in the measurement of the composition technique the average of the words at the same time expresses time factors too, that is it is also to be regarded as a type of quality indicator.

Let us now see how the performances of two students can be compared.

For example: there are 17 mistakes in a certain operation in the composition of student N.N., and the extent of his composition is 214 words. The mistake indicator in the given operation is therefore:

$$\frac{17}{214} \times 100 = 7.9 \%.$$

Student X.Y. makes 8 mistakes in the same operation, and the extent of his composition is 169 words. His mistake indicator therefore is:

$$\frac{8}{169} \times 100 = 4.7 \%.$$

X.Y.'s work is thus less faulty, or is better, than that of N.N.

Demonstration of the student's q uality on the basis of his faults, however, is in contrast with the general view, since the smaller number means a better performance, and the larger number a worse one. This is overcome by taking the perfect performance as 100 %. The quality indicator is then obtained by subtracting the mistake indicator from 100. Thus, in the above examples, N.N.'s quality indicator is 92.1 %, and that of X.Y. 95.3 %, and it is immediately seen which performance is the better.

It must be noted that this index has the result that the performance improves in direct proportion with the increase of the extent. It is now known, however, that this is not exactly true. To avoid distortion, we calculate not with the number of words, but with the square root of this number. The index for each operation is then calculated from:

$$E = /1 - \frac{m}{W}/ \times 100$$

where E = operational index,

m = number of mistakes,

and W = number of words.

In addition to the operational indices, however, there is also a need for a combined index, which can express the values of the compositions jointly, on the basis of the 18 operations. It was therefore necessary to form a common basis for comparison, such as that of the performance of adult intellectuals in the case of the elementary calculating ability. But what can be taken as maximum? The language of writers or poets? The language of the compositions of civil servants? The performance of scholars? To overcome the difficulties, we calculated the average of the indices of all the operations from the compositions of all the students who took part in the measurement, and this was compared with the average of the words of all the compositions. In this way we obtained a well usable basis of comparison.

IV
UTILIZATION OF THE RESULTS

1. The scope of the measurements

In principle it would be optimum if the measurements would extend to every subject. There are several obstacles to this, however. It must be stressed above all that the characteristics of the system of material in the taught subjects differ. Thus it can be said that the „natures" of the individual subjects are different. Accordingly, there are subjects which have systems of material practically „offering themselves" for measurement and are routinely measurable; those of others can be measured only with greater difficulty, and there are certain subjects in which the measurement is in opposition to the aims of the subject, so that measurement would be harmful.

a. Measurement can be carried out relatively easily in those subjects the material of which is composed in the vast majority by sets, laws, theories and rules. Examples of these subjects are mathematics, physics, chemistry, biology, and the grammar of the mother-language.

b. The measurement raises more difficult and qualitatively different problems in those subjects, the material of which is largely composed of individual concepts. Here a large mass of facts is taught, and the system of the concepts is less obvious; becauese of this we encounter measurement-methodological problems of a different nature. The principle of totality too may be made good only with difficulty in the measurement. Examples of such subjects are geography and history.

The methodology developed in our Department can be employed to measure the level of knowledge in subjects operating with sets.

c. There are subjects in which, in addition to reason, intellect and intuition, a very large, and perhaps decisive role is played by emotional factors; this is the case, for example, in literature and in the teaching of singing, music, drawing and the history of art. The carrying out of measurements here would oppose the aims of these subjects. The main target in the teaching of these subjects is for the children to become capable of accepting and enjoying works of art; to form their tastes and develop their emotional horizons; to arouse in them the need to make their lives more complete by means of aesthetics; to form their moral and ideological outlooks is such a way that the universe, and characteristic spheres of it which are less approachable merely by reasoning, are revelaled by the works of art as a living experience.

However, it is possible to measure in an exact way only the material assisting the understanding and acceptance of the works of art, and so the measurements would divert the attention of the teachers to this material, which would result in a distortion in the teaching. In our view, therefore, it would be dangerous and harmful to carry out measurements in these subjects. /The realization of the aims of these subjects is perhaps impeded by the fact itself that the syllabus includes them as subjects in the traditional sense, and this inspires the use of the same methods and similar requirements as in the other subjects. This problem does not belong to our present theme; we merely wished to point out why we do not deal with the measurement-methodological problems of these subjects even theoretically./

d. On the above basis, theme-concluding measurements were carried out in the following subjects: grammar of the mother-language, mathematics, chemistry, physics and biology.

Our plan to prepare a complete system of tests measuring knowledge in these subjects is progressively being effected. Teams of theoretical specialists and practising teachers, working separately from subject to subject, in each year in general compile theme-concluding questionnaires for one, or more rarely two classes in each subject. Only in the grammar of the mother-language was a test-system prepared for all primary school upper classes at the same time; these too were standardized in the school-year 1970/71. According to the present rate of this work, the preparation and standardization of the complete system of theme-concluding tests for the upper classes in the above-mentioned subjects will be completed in the school-year 1973/74.

For the time being, theme-concluding measurements are carried out in the secondary schools only in chemistry, biology and physics. The preparation of these tests too will be completed in the school-year 1973/74.

e. Our measurements on development have been extended to correct spelling, composition and arithmetical operations in the lower classes, and were completed in the school-year 1971/72.

2. Manner of publication

In order for the measurements to fulfil their most important functions as outlined in part I /continuous reporting back, objective comparison/, it is vital that the results be published.

From the point of view of school practice, it is of prime importance to publish the tests and the correcting keys containing the evaluation norms. For want of suitable apparatus, a university department can naturally not undertake this. At present, however, no test-preparing organization has been created either.

For the controlling organs and the research workers, on the other hand, a multidirectional analysis of the results is more important, containing separate volumes for each subject. The publication of a special series of these has now been started by the Tankönyvkiadó Vállalat /Test-book Publishers/ under the title „Eredménymérés az iskolában" /„Measurement of results in schools"/.

Since there is at present no organized method of ensuring that the tests are published and made available to schools, for the time being we are also publishing the tests in the analysis volumes, so that those schools which wish to make use of them may themselves duplicate the tests in accordance with their own requirements.

Consequently, the tests and the analysis of the measured results are published in separate volumes for each class in each subject. These volumes contain as many chapters as the number of theme-concluding measurements carried out in the relevant subject. At the beginning of every chapter a quantitative and structural analysis of the material of the thematic unit is presented; this is followed by the tests with the correcting keys, and finally come the results.

Each type of measurement of development is dealt with in a separate volume.

3. Interpretation of the results

a. In the individual measurement types we have developed uniform procedures and forms of the analysis.

It has proved possible to perfect the standardization of the interpretation in the theme-concluding measurements, while in the analysis of the results of the developmental measurements it became necessary to use not only agreeing procedures, but also methods characteristic of each individual type.

In general, quantitative, graphical and qualitative analytical procedures are employed.

Ouantitative analysis provides the necessary indeces. Graphical plotting is used mainly to give a clear picture of the calculated results, but in certain cases the graphical procedure also serves to disclose relationships /particularly in the measurement of development/ which are not at once obvious from the data themselves.

In addition to this, we also consider that there is a need for verbal, qualitative interpretation, especially in the initial volumes, since in general the teachers do not possess the mathematical, statistical knowledge necessary to interpret the data. Additionally, however, it is necessary to compose didactical conclusions of general validity involving syllabus theory in qualitative decisions too.

As an example we shall give a brief account of the manner of interpreting results of theme-concluding measurements.

We first give the average and level of significance for each variant, and then report the distribution in tables and graphically. The ranges of the marks too are depicted on a histogram. Following this the national percentage point averages attained for the individual knowledge elements are presented in a column diagram, and then, although the results are plotted in such a column diagram grouped according to unit content, every individual element is also given separately.

To illustrate this we present the data of one test variant /variant D, theme II. grammar, class 5 of the primary school/. /Only details of the column diagram are given./

Table 15

Summarizing data of variant II/D

Average	$\bar{x}$	60.1
Confidence interval	$\pm\Delta$	$\pm$ 2.6
Accuracy requirement		$\pm$ 4.3%
Standard deviation	$\pm$ s	$\pm$ 18.3
Relative deviation		30.4%
Median		60.0

Table 16 Distribution

Percentage points	Students /%/
0.0 - 5.0	0.0
5.1 - 10.0	0.0
10.1 - 15.0	0.5
15.1 - 20.0	2.1
20.1 - 25.0	1.0
25.1 - 30.0	2.1
30.1 - 35.0	4.8
35.1 - 40.0	4.8
40.1 - 45.0	3.7
45.1 - 50.0	8.0
50.1 - 55.0	10.1
55.1 - 60.0	11.7
60.1 - 65.0	10.1
65.1 - 70.0	6.9
70.1 - 75.0	7.4
75.1 - 80.0	12.2
80.1 - 85.0	5.8
85.1 - 90.0	4.2
90.1 - 95.0	2.6
95.1 -100.0	1.0

Figure 21. DISTRIBUTION OF THE RESULTS FOR VARIANT II./D

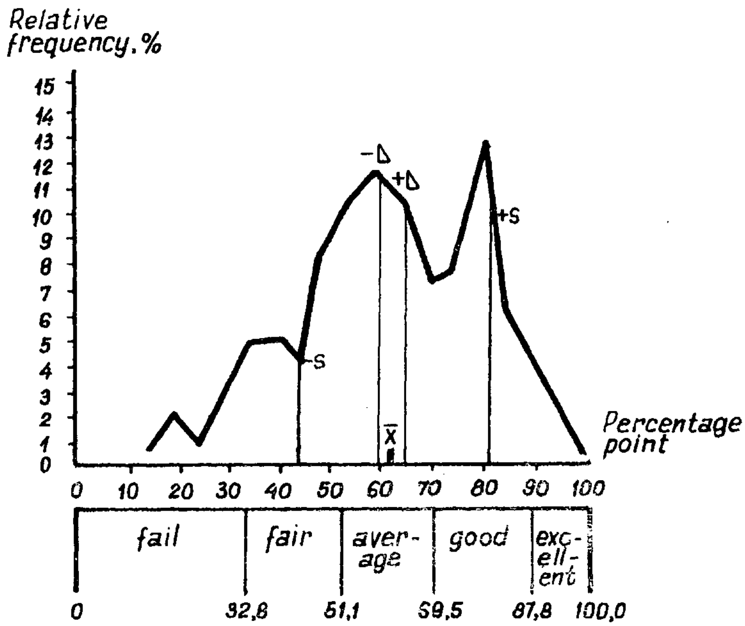
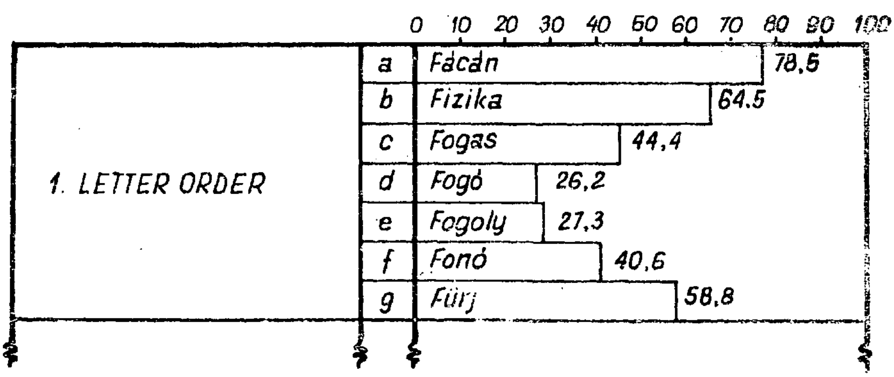


Figure 25. RESULTS FOR VARIANT II./D



b. On the occasion of national representative measurements the school director and all of the individual students taking part in the measurement are asked to fill in a data sheet.

The former provides data on the personnel and material provision of the school, its equipment, the conditions of the buildings, etc., while the students give information on their families, and the relevant economic, social and cultural conditions /occupation, income and scholastic achievements of parents; number of brothers and sisters; size and conditions of house, etc./ and on other factors /attitude to school, subject, etc./.

The representative measurement makes it possible for us to calculate the correlation between the causal data /see preceding paragraph/ and the study results, and hence to draw more general reliable conclusions, for example on pedagogical sociology, or with regard to the effectivity of the school system, or the factors affecting the effectivity.

References

Adams, G.S.: Measurement and Evaluation in Education, Psychology and Guidance. Holt, Rinehart and Winston. New York, Chicago, San Francisco, Toronto, London. 1964. 654 p.

Ágoston-Nagy-Orosz: Méréses módszerek a pedagógiában /Measurement methods in pedagogy/. Budapest. Tankönyvkiadó Vállalat. 1971. 372 p.

Nagy, J.: A témazáró tudásszintmérés gyakorlati kérdései /Practical questions of the theme-concluding measurement of knowledge levels/. Budapest. 1970. 126 p.

Nagy, J.: Az elemi számolási készség mérése /The measurement of elementary calculating ability/. Budapest. Tankönyvkiadó Vállalat. 1971. 188 p.

Orosz, S.: A fogalmazástechnikaerésmetodikai problémái és országos szinvonala /Measurement-methodological problems and national level of composition technique/. /In press/.

Orosz, S.: Témazáró tudásszintmérés és az országos szint elemzése magyar nyelvtanból az általános iskola 5. osztályában /Theme-concluding measurement of the level of knowledge and the analysis of the national level in Hungarian grammar in class 5 of the primary school/. /In press/.

Rummel, F.: An Intorduction to Research Procedures in Education. Harper and Brothers. New York. 1958. 414 p.

Sixtl, F.: Messmethoden der Psichologie. Theoretische Grundlagen und Probleme. J. Belz Verlag. Weinheim. 1967. 470 p.

Torgerson-Warren: Theory and Methods of Scaling. John Wiley and Sons. New York. 1962. /Third Printing./ 460 p.

Измерение знаний в единой школьной системе в Венгрии

Д-р Шандор Орос

На кафедре педагогики мы начали исследовать, разработать теорию и практику измерения подходящего к единой социалистической школьной системе 8-10 лет тому назад.

Эта статья дает обзор об этом исследовании.

Первая глава анализирует функции и роль измерения в единой социалистической школе: измерение как средство обратной всязи, управления учебным процессом, контроля для ученика и педагога и объективной оценки.

Вторая глава опишет теоретические и методические основы измерения, которые были разработаны на кафедре. Излагаются методы квантификации и перевода результата измерения в отметки. Наконец автор опишет два типа тэстов: тэсты для измерения развития навыков и умений и тэсты для заключения тем обучения.

Третья глава дает описание практики измерения. Анализ содержания и структуры тем; методы разделения тем на элементы и действия; определение уровня опроса; методы составления тэстов (комбинированная техника вопросов и задач).

Четвертая глава занимается проблемами анализа и воспользования результатов измерения.

Spezifisches Messungssystem von Kentnissen und Fertigkeiten der Schüler in Ungarn, angemessen dem einheitlichen Schulsystem

Sándor Orosz

Seit 8-10 Jahren arbeitet eine Forschungsgruppe am Pädagogischen Lehrstuhl der Szegeder Attila-József-Universität, um eine dem einheitlichen Schulsystem angemessene Theorie und Praxis der Messungen von Schülerleistungen auszuarbeiten. Die Studie gibt einen Überblick über diese Forschungen.

Im ersten Kapitel wird es behauptet, dass die Messung von Kentnissen und Fertigkeiten im einheitlichen Schulsystem als wichtiges Mittel der Steuerung der pädagogischen Tätigkeit gilt. Durch die Messungen wird 1. die ständige Rückkopplung über die Ergebnisse der Lehrtätigkeit für die Lehrer sowie die Schulverwaltung, 2. die Objektivität der Zensuren gesichert.

Im zweiten Kapitel werden die theoretisch-methodischen Grundlagen der Messungen bekanntgemacht. Es wird die spezifische Lösung von Problemen der Skalierung und des Gewichtens vorgetragen. Zuletzt werden die zwei Messungsweisen, die sog. themaschliessende und die Entwicklungsmessung dargelegt.

Das dritte Kapitel stellt die Praxis der Messungen dar. Der Verfasser legt die inhaltlichen und strukturellen Analysierensmethoden, die Art und Weise der Bestimmung und Zerlegung in alternative Elemente der zu messenden Tätigkeiten sowie die Festlegung des Tätigkeitsniveaus vor. Er teilt auch die Methoden des Zusammenstellens von Testblättern /"kombinierte Fragen-bzw. Aufgabentechnik"/ mit; er behauptet auch die zu lösenden Probleme der Verfertigung von verschiedenen Tests.

Im vierten Kapitel werden die Probleme des Analysierens und Benutzens der erhaltenen Ergebnisse bekanntgemacht.