STANDARD GRADE

József Nagy

In research connected with the measuerement of achievement the question of the grade is unavoidable. It is not coincidence that it is fequently necessary to turn to this problem. To date we have mainly attempted to provide an answer to the practical question of how practicable it is to convert the results of the measuerements to a grade[1]. In our earlier work, some features of the grade were naturally reported, but the theoretical basis has still not been explained. Since the grade, as one of the means of classification, has now come into the centre of attention as a result of certain distortions, it seems advisable to give an account of the theoretical principles.

It follows from the title and from the above that here only one of the many varied questions of the grade and the grading /which are virtually constant themes in the literature/ will be discussed: the grade as a scale and a means of measurement.

I.

The calculus, the grade, is based on the range scale. By means of the range scale the studied material can be arranged in an increasing or decreasing sequence according to a given property, and hence graded. The range scale is suitable only for the establishment of the range sequence. It does not tell us how much more, better or larger one value of the schale is than another.

The range scale denotes the values of the range sequence verbally, by symbols or by numbers For example: "I do not like it", "I like it", "I like it very much"; or: "weak", "good", "excellent". It is obvious that "excellent" is a higher-ranking classification than "good", but it would be impossible to determine how much poorer "weak" is than "excellent". In this respect, nothing is changed essentially if the verbal indication is replaced by letter symbols or by numbers. For example: "weak" = 1, "good" = 2, "excellent" = 3. Here, however, the value 3 is not worth three times as much as 1.

Accordingly, the numbers ascribed to the individual ranges, the order or range numbers, are not additive:

$$\begin{array}{ccccccc} \text{weak} & + & \text{good} & + & \text{excellent} & = & ? \\ 1 & + & 2 & + & 3 & = & 6\ ? \end{array}$$

It is not difficult to see that <u>the addition of range numbers has no objective sense</u>. Although formally the total in the above example is <u>six</u>, since the extent of the interval between the range numbers is unknown, this is only a formal numerical value. If these numbers must not be added together, then clearly they are not susceptible to any mathematical operations.

This also applies to averaging, and thus <u>the averaging of the grades is theoretically meaningless and unjustifiable</u>. This is particularly so if the scale originally consisting of five values is divided into tenths to give 50 values, or into hundredths to give 500 values, and it is thought that the data obtained with the scale of five values becomes ten or one hundred times more exact merely as a consegence of the averaging.

Even if we can average with appropriate reservations, it is at any rate desirable to round off the numerical values obtained to the original range numbers, that is to integral values. Accordingly, the earlier solution, whereby a pupil's average school achievement could be "excellent", "good", etc., was less formal than the averaging practice of more recent years, which expresses the average in tenths.

Although it is true that the range numbers of the range scales are not additive, this does not mean that the verbal indication is of the same level as the indication by numbers. The indication of the range values by numbers is qualitatively of a higher order than the verbal indication, because, although on a low-level scale, this is nevertheless a measurement, whereas the verbal designation is not. Range statistics offer excellent methods for the analysis of data obtained on the range scale. Further, the ranking of the verbal range includes many more uncertainty factors than the range numbers, and is much more cumbersome too.

It should be noted that the range scales are generally comprised of an odd number of range values /with a view to the better perceptibility of the middle value/, and most frequently have 3, 5 or 7 range values. Our grading system uses the range scale with 5 values, which /according to many measurement-theory experts/ permits the most reliable assessment.

The <u>objectivity</u> of a datum obtained with the range scale does not depend on the above properties of the range scale. The range is not subjective for the reason that measurements are made with a range scale. A datum is objective to the extent of the accuracy and concreteness of the definition of the measuring number. Such a definition must ensure that there is an unambiguous decision regarding the property of the given material: the range value which is to be attributed to it. If the grading is carried out by different persons, the given material should be graded identically by all of them.

It is well known that our grade does not possess a definition of the above level. This is the explanation of the <u>subjectivity</u> of the grading.

The question arises of why our definition is not sufficiently precise. Unfortunately, not because the authors did not formulate the section referring to the Procedure with the necessary circumspection. The situation is that the imprecision of the definition of the grades, as a series of numbers, arises from the function of the grade.

<u>The function of the grade is a double one</u>. /It should not be

forgotten that we are discussing it only as a means of measurement, and not its motivation and other functions./

Its first task is to indicate what proportion of the curricular requirements the pupil has mastered. According to this, for instance, a value of 5 denotes that the pupil knows the vast majority of the given curricular requirements. This function therefore attempts to assess the amount of knowledge compared to the requirements, or in other words the relation of the teaching matter and the pupil.

The second function serves to assess the relation between the pupils. The pupils are differentiated from each other according to the grade attained.

The investigations of Zoltán Báthory[2] and our own measurements too have shown that, although the distribution of the pupils according to grade is different from subject to subject, the grades in general fulfil this double function within certain limits of error.

Thus, when the result of assessment is a value of 5, this not only means that the pupil has mastered the vast majority of the curricular requirements, but also expresses the fact that the given pupil belongs in the group of the best of the pupils as a whole. This is so, independently of the natures of the school, the subject or the theme, the age of the pupil or the class attended.

The simultaneous fulfilment of the above two functions permits the considerable generlity and comparability of which the traditional grade is capable.

However, the price of the high order of generalizability and comparability is the loss of the requisite accuracy and objectivity.

With reference to a single exercise, for instrance, it is conceivable that the individual grades can be defined with satisfactory accuracy. It can be formulated exactly what conditions must be fulfilled for the achievement of a given pupil to be rewarded with a given grade. Thus, the grade would be of satisfactory accuracy and objectivity, therefore, but only within the assessment of the given exercise. However, the accuracy and objectivity obtained would be at the expense of the comparability outside the exercise. /A 4 for another, more difficult exercise might well be worth twice as much as a 4 for an easier exercise within the given subject./ In addition, there is the comparability between the different subjects and the different classes, which would become completely hopeless.

It can readily be seen that with the increase of the objectivity of the grade the assessment of the relation of the pupil and the curricular requirements may become more exact /in theory, of course, not considering here the sources of error in the assessment/, while the accuracy of the assessment of the relation between the pupils, and the more general level of comparability of the grades, may decrease substantially.

To summarize: the strength and advantage of the traditional grade lies in its high degree of general comparability, but this involves the necessary concomitant price of the relatively low level of the accuracy and objectivity.

It is clear from this that the grade taken in the traditional sense, which simultaneously provides the two basic functions reported above, when ensuring very general comparability, can not be satisfactorily objective. It is also obvious that it would be senseless to strive for "absolute" objectivity.

A grade permitting general /standard/ comparison can originate only from a measurement. But why is there a need for a grade, a "standard grade", if the measurement results are available anyway.

II.

Let us now disregard the practical facts that today we still have few standardized subject tests and use few test papers from which a standard grade can be obtained, while it is perhaps not possible, or even advisable, to measure all pupil performances with tests. In this respect, our present situation justifies the necessity of grading only from a practical point of view. /Of course, we must also pass over the perspective possibility of educational organizational forms which, at least in the fundamental primary school, would dispense with the need for the grade and assessment in their present conception. Such educational organizational forms can not be introduced either now or in the near future./

Accepting the present educational organizational forms as given, the question which arises is: what is the temporary justification of the expression of the measured results in standard grades?

The explanation of this lies in the nature of the psychopedagogic measurement. If we wish to measure in the interest of a more objective assessment, then a means of measurement is required. The basic data /the raw scores/ obtained with the means of measurement, however, can be compared with other scores only if these are obtained by the same means. As a consequence of the objectivization, therefore, the more general comparability of the results obtained has been lost. This is similar to the case of the more objective greades of the exercise mentioned above.

The raw scores of tests are particular, in themselves meaningless, useless data. They are not able to provide either of the two basic functions discussed in connection with the grade. For this reason, various methods have been elaborated to transofrm the raw scores in such a way as to lead to their comparability. Since the conversion of raw scores to a comparable form is one of the most important questions of psychometry, it is understandable that many methods have been devised. Nevertheless, all of these in effect are one or other variant of two basic possibilities, these being the two functions treated above. However, while the grade provides both functions simultaneously /the price of this being lower accuracy and objectivity/, there is no method which can transform the measured results to a single index that contains both functions.

Let us consider first the possibility of expressing the relation between the pupil's performance and the requirements.

The essence of this is that the total raw score attainable in the test is regarded as a 100 % performance, and the proportion of the requirement formulated in the test that is mastered by the pupil is expressed as a percentage of this maximum.

There are formally three possibilities for the expression of this proportion.

The most widely known solution is given by the ratio of the raw score attained by the given pupil to the overall possible raw score, multiplied by one hundred:

$$N_{\%} = \frac{p}{P} 100$$

where $N_{\%}$ is the performance attained compared to the requirement /level of attainment/, P is the total number of raw scores attainable, and p is the number of raw scores attained by the given pupil.

The second method, particularly readily used in measurements by the multiple-choice technique, is that exactly 100 elements /answer-selection questions/ figure in the test.

The third formal solution combines the advantages of the above two variants. This is attained by calculating the distribution ratios /percentage distribution/ from the total possible raw scores[3].

The basis of comparison is thus the total raw score of the perfectly solved test. This system of comparison well expresses the relation between the pupil and the curricular requirement /but at least the requirement formulated by the test/. This has the consequence, however, that the percentage performances thus obtained can not be compared with the performances in other themes and other subjects.

If, for example, the national average in one test is 60 %, and in another 80 %, then it is obvious that a performance of 70 %

attained by a given pupil in the former test is not of the same value as a performance of 70 % in the latter.

And since it is impossible to compile tests which are of the same degree of difficulty, independently of the theme, the subject and the class, comparability of a more general level by this means, with the percentage index, can not be achieved.

From the data of measurements with tests it is possible to create scales which permit a comparison independent of the theme, subject and year: we thereby come into possession of a standard scale.

The most fundamental characteristic of standard scales is that, since a scale beginning with the natural zero point is not possible in the world of psychopedagogic phenomena, the average of the measurements is taken as the starting-point. The average performance is converted to zero, and the individual performances are expressed in relation to this.

The starting basis is the standard z score /see Figure 1, second row/.

The transformation is performed as follows:

$$z = \frac{x - \bar{x}}{s}$$

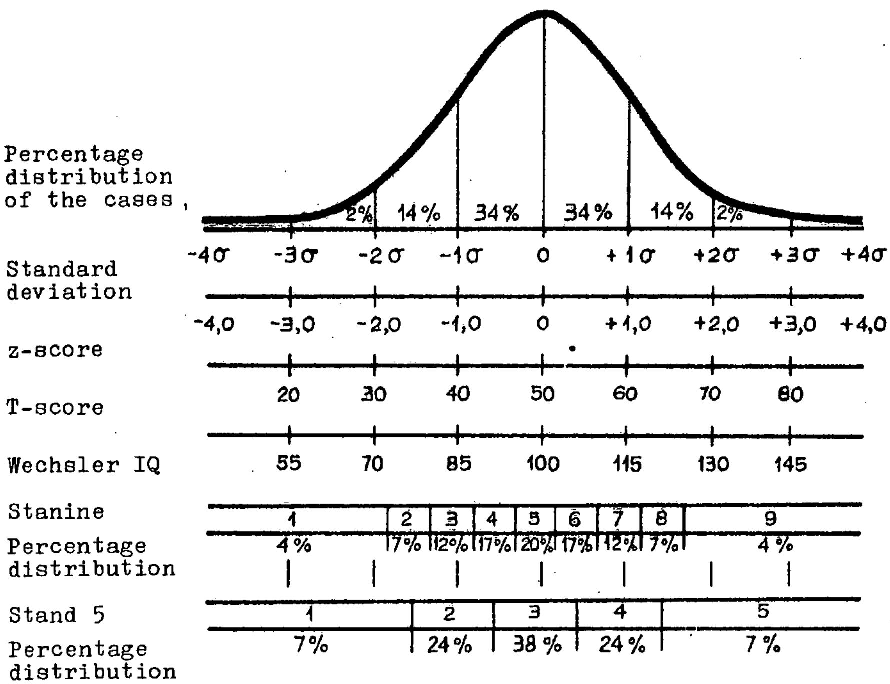
where x = the sum of the raw scores attained by the given pupil,

$\bar{x}$ = the average of the raw scores of the measurement,

s = the standard deviation of the raw scores.

Figure 1.

POSITION OF THE STANDARD GRADE /STAND 5/ IN THE SYSTEM OF STANDARD GRADES



For the reader not acquainted with statistics to sense the essence of the above formula exactly, let us take some examples. Let the avergae of the raw scores be 60 /$\bar{x}$ = 60/, and the standard deviation 20 /s = 20/. If the performance of a given pupil is 60, i.e. equal to the average, then the value of the standard z score is:

$$z = \frac{60 - 60}{20} = 0$$

If the performance of a given pupil be less than the average /e.g. 50/, then the value of z is negative, while if it is larger than the average /e.g. 70/, then z will be positive:

$$z = \frac{50 - 60}{20} = -0.5$$

$$z = \frac{70 - 60}{20} = +0.5$$

As can be seen in the Figure, the values of the standard z scores practically extend from - 4 to + 4. And as the Figure shows, the intervals between the values expressed by the numbers are equal, and hence these values are additive and are suitable for mathematical processing.

The standard z score thus expresses the relation to the average performance. Consequently, whatever the content of the test, and whatever the class the measurement is performed on, the standard z score ensures the comparability. No matter from what test it is found, for instance, that $z = 1.2$, this is equal to a standard score of $z = 1.2$ from any other test. In both cases the performance is at a distance of +1.2 times the standard deviation from the average. It is better by this amount than the average performance. <u>We are faced here, therefore, with the much debated fact that the role of the gauge is occupied by the average pupils</u>, and the others are related to this.

The questions of what the given average performance is worth, what level it means, and what relation it is in with the requirements, do not even arise here. This is dangerous for the reason that it may appear that there are no such questions, or that if so, then they are not of importance. The average may be predistined to appear as the central figure, but the essential question is nevertheless the extent to which the pupils on average /and the individual pupils concretely/ have mastered the curricular requirements, and not where they stand in comparison to the average performance, or to the average pupil.

We have seen that the content, the percentage relation, gives an answer to this question, but it is not comparable with the results of other tests; the relation to the average at the same time permits comparability of a high order, but it gets out of touch with the requirements, the teaching material.

Since these two points of view can not be united into a single index without the loss of the objectivity, it appears advisable and desirable to use both indices.

The index obtained with the content, the percentage relation, is readily understandable for everyone, and can be well used with reference to the given test. It unambiguously shows what the pupil knows and what he must still learn. However, if we wish to compare the performances attained in several consecutive themes and to express them with a single index, the values expressed as percentages can not be added and can not be averaged. /In this respect one should consider the performances of 70 % obtained in each of the two tests mentioned above./ For this reason, therefore, the standard index discussed earlier is indispensable.

Because of the negative numbers and the cumbersomenes of the values expressed in tenths, it is customary to transform the standard z scale to derived standard scales.

Figure 1 shows the most frequently used transformed standard scales.

In the case of the T scores, the zero point of the z score, expressing the average, is taken as 50, and its standard deviation as 10:

$$T = 10\ z + 50$$

For the above example, with a performance of $z = +\ 1.2$:

$$T = 10 \times 1.2 + 50 = 62$$

The scores of the widely known intelligence quotient /IQ/ are also derived from the standard z scores. Here, the average is 100, and the standard deviation 15 or 16. /These standard IQ's should not be confused with that obtained as the ratio of the mental and chronological ages./ Thus:

$$IQ = 15\ z + 100$$

$$\text{or}\quad IQ = 16\ z + 100$$

The stanine /standard nine/ scale, with nine range values, was developed during the Second World War by the American Air-Force. As a result of its advantages, which cannot be given in detail here, its use has spread very rapidly in recent years.

$$\text{Stanine} = 2\ z + 5$$

If, for example, $z = 1.2$, then the value in the stanine is:

$$\text{Stanine} = 2 \times 1.2 + 5 = 7.4 = 7$$

As can be seen in the Figure, the two extreme values are open. A value of even 11 may be obtained from the calculation, but all values larger than 8 are assumed as stanine = 9.

As a consequence of the greater scale of the stanine, each number represents such a large interval that it is advisable to determine what percentage of the pupils are associated with each value. According to custom, therefore, the proportions of pupils expected for the individual stanine values have been given below the stanine scale.

For instance, the stanine = 9 shows the relation to the average /5/, and also indicates that this performance is so high that it can be achieved by only about 4 % of the pupils. With this performance, therefore, a given pupil belongs among the best 4 % of the pupils.

If the distribution is not close to the normal, i.e. it does not resemble the bell-shape to be seen in the Figure, then this advantage of the stanine is naturally lost. /This question will be returned to later./

On the basis of the Figure, the reader will certainly already have discerned that there is no obstacle to the transformation of the raw scores to any optional standard scale, whereby we can attain general comparability with the simultaneous retention of the objectivity.

In practice, however, in addition to those used so far, there are still three reasonable possibilities on the analogy of the stanine: these are standard scales with 7, 5 or 3 range values. Since a range scale with 5 values is in use for grading in Hungary and many other countries, the task is in fact to accommodate to this, and to fit the standard scale of the results obtained by measurement into this system.

On analogy with the stanine, the standard scale with 5 range values is termed standard 5, or in short <u>stand5</u>.

The solution of stand5 may be of various forms, depending on the value assumed for the standard deviation. The average is given by 3, the middle of the 5 range values on the scale.

In the studies mentioned in the introduction, experiments were made with two types of solution. Gradually accumulating experience and theoretical and practical considerations show that it is reasonable to select the variant to be seen in the Figure.

In the case of this variant, the intervals for the values 1 and 5 each contain about 7 % of the pupils, which is close to Hungarian practice. This solution gives the proportion of the pupils with a value of 1 below 1.5 times the standard deviation, and above 1.5 times the standard deviation the proportion of the pupils with a value of 5. $2 \times 1.5 = 3$ times the standard deviation is divided into three equal parts for the standard grade values of 2, 3 and 4. it follows from this that these values each cover intervals with a standard deviation of 1. Thus:

$$\text{Stand5} = 1 \times z + 3 = z + 3$$

Accordingly, only the average value of 3 need be added to the standard /$\underline{z}$/ score to obtain the stand5 value /after the normal rounding, of course/.

In the above, the conclusion was reached that it must be established from the raw scores what proportion of the requirements has been mastered by the pupil. This is given by an index expressed as a percentage. In addition, there is also a need for some standard scale or index which permits general comparability. Of the many types of possibility starting from the standard /$\underline{z}$/ score, in Hungary it appears that a standard scale with 5 range values can reasonably be fitted into the present classification system.

It has been seen that both indices are obtained from the raw scores. However, since the percentage index value is obtained by linear transformation, it is irrelevant whether the standard /$\underline{z}$/ score is calculated from the data expressed in raw scores or in percentages. The value of $\underline{z}$ will be the same in both cases.

The possibility of further simplification arises from the circumstance that, in the case of a sample of given average and standard deviation, the individual stand5 grades are comprised of definite intervals.

Let us consider a test in which the national average is 60 % /$\bar{x}$ = 60/ and the standard deviation is 20 % /$\underline{s}$ = 20/, and in which a pupil $\underline{x}$ achieved a performance of 75 % /$\underline{x}$ = 75/. The question is: what mark should be awarded for this performance from stand5?

On the basis of the relations reported above:

$$z = \frac{x - \bar{x}}{s} = \frac{75 - 60}{20} = \frac{15}{20} = 0.75$$

From this the stand5 is:

$$\text{stand5} = z + 3 = 0.75 + 3 = 3.75 = 4$$

It would be tedious to perform this calculation separately for every pupil. For this reason, it is worthwhile determining the limiting values of the intervals. It can be seen from the Figure that the limit of the "fail" mark, for instance, begins at a score of 1.5 times the standard deviation. Since the standard deviation in the above example is $\underline{s} = 20$, this value is 30, and this is the distance from the average of the score where the limit of the "fail" begins. Deducting 1.5 times the standard deviation from the average, we obtain the upper limit of the 1, which is at the same time the lower limit of the 2, i.e. 60 - 30 = 30.

We know that the three central values of stand5 embrace 1.5 times the standard deviation both upwards and downwards from the average, i.e. in all 3 times the standard deviation. An interval of unit standard deviation therefore falls to each of the three middle marks, and thus a very simple solution results. The value of the standard deviation is added to the limit of the 1 three times, one after another.

$\bar{x} - 1.5\ s$	the upper limit of standard 1
$/\bar{x} - 1.5\ s/ + s$	the upper limit of standard 2
$/\bar{x} - 1.5\ s/ + 2s$	the upper limit of standard 3
$/\bar{x} - 1.5\ s/ + 3s$	the upper limit of standard 4 and at the same time the lower limit of standard 5

In our example the limit of the "fail" is 30, while the value of the standard deviation is 20. From these data the limits of the other standars are:

1	below 30
2	30 - 50
3	50 - 70
4	70 - 90
5	above 90

We have thus obtained a key to the given test, by means of which the performance of the given pupil can be expressed in stand5. The performance of pupil $\underline{x}$ above was 75.

By calculation it was established that this is equivalent to 4. This can be read off from the above key without calculation. The calculation of the key naturally involves the condition that the values of the standard deviation and the average be known, while its use is that it gives the relevant key to all standardized tests. With the above method the limiting cases can not be decided. Should a performance of 70 be a 4 or a 3? If it is considered advisable, the problem can simply be eliminated according to the rules of grouping, but in the end a decision as to which grade a pupil receives in the limiting cases should preferably be based on pedagogic considerations, rather than on some "absolutely exact" computational viewpoint. Thus, the use of the above solution is satisfactory.

It was mentioned earlier that in the event of normal distribution the standard grade expresses what percentage of the pupils belong to each grade /see the Figure/. However, if the distribution is not symmetrical /not approximately bell-shaped/, then these proportions are distorted.

It is possible to take the asymmetry into account. Such a procedure has been described in the book already reffered to: "A témazáró tudásszintmérés kérdései /Questions of the theme-concluding measurement of achievement/", pp. 71-74. It is unnecessary to give an account of this method again here. We merely wish to point out that if the skewness is taken into consideration this resolves to a certain extent the extreme distribution of the proportions of pupils in the individual grades, but at the same time the equality of the intervals of the individual grades undergoes a distortion.

All this, however, is a problem only in the case of extremely high /above 75-80 %/ national averages. In the other cases, a situation approximating to that seen in the Figure is obtained; that is, the standard character of the grade remains. The distorting effect of the extreme averages and standard deviations somewhat distorts the standard character of the grade, no matter what method is used to try to eliminate the consequences of the distortion.

To summarize, it may be stated that the performance in the standardized tests, expressed as a percentage, shows the extent to which the pupils in general, or the individual pupils, have mastered the subject matter. However, this index is not a standard, and can not be compared with the percentage indices of other tests. Accordingly, it is absolutely necessary to find some standard index too. Since a grading system with five range values is used in Hungary, the stand5 is the most appropriate for this purpose.

The tests used for university and other examination purposes, and also the most varied measurement and problem forms, are not standardized. Nevertheless, if the means of measurement are good, then the results obtained are objective. It is necessary to express the performance in these cases too, given as a percentage, on a scale permitting a more general comparison; i.e. to convert them to a grade corresponding to our conditions. Since the assessment system of these tests can not be based on a previous representative measurement, the use of the method reported above can naturally not result in a standard grade. It would be in vain to calculate the average performance of the pupils taking part in the measurement, and the value of the standard deviation. Since these data are not of general validity, the standard character of the standard grades calculated from them would also become uncertain.

In spite of this, in place of the various /often decidedly dangerous/ methods of conversion to a grade, it is desirable to intorduce a better, "more standard" solution. A Table was reported in the Appendix to the above-mentioned book, for the conversion of the results of such tests to a grade. This Table is now reported in a further developed and simplified form.

It is a condition of use of this Table that we calculate the average of the performances, expressed as a percentage, of those taking part in the measurement. If the corresponding value is found in the "Average" column, the series of data associated with this gives the key to the conversion to the grade.

Table 1

TABLE FOR THE SELECTION OF THE KEY OF THE PARTIALLY STANDARD GRADE

Average	Limits of grades								
	1		2		3		4		5
37,4 and below		15		20		40		65	
37,5 - 42,4		19		36		53		70	
42,5 - 47,4		21		38		55		72	
47,5 - 52,4		23		40		57		74	
52,5 - 57,4		26		43		60		77	
57,5 - 62,4		29		46		63		80	
62,5 - 67,4		31		49		67		85	
67,5 - 72,4		33		55		75		90	
72,5 - 77,4		35		60		80		90	
77,5 and above		40		70		80		90	

The reasons that the grades thus obtained are partially standard grades are that the measurement is not representative, and that only the average is taken into consideration. The standard deviation is established by a relative standard deviation of about 30 %.

At the same time, the lower limit of the "excellent" here is fixed at 90 % for the high averages, since in such tests the uncertainty factor is more.

At present, and in the future, therefore, three types of grade will be in use: a standard grade based on the results of standardized tests, a partially standard grade from the results of non-standardized tests and measurement and problem forms, and the grade obtained with a range scale by the traditional means. These tree types of grade can be conceived as a unified system, in spite of their objectivities being of different levels.

While the vast majority of grades are obtained in the traditional way, the standard and partially standard grades make relatively little change or improvement in the objectivity and standard character of the grading. The increase of the proportion of the standard and partially standard grades /as a consequence of the graudal spreading of assessment based on measurement/ may clearly progressively increase the objectivity of the grading, and the more clear-cut comparability and astandard character of the grades.

References

[1] Nagy, József: Az elemi számolási készségek mérése és fejlettségének országos szinvonala /Measurement and national level of development of elementary calculating abilities/. Tankönyvkiadó, Budapest, 1971. pp. 21-32. Ágoston-Nagy-Orosz: Méréses módszerek a pedagógiában /Measurement methods in pedagogy/. Tankönyvkiadó, Budapest, 1971. pp. 103-111. Nagy, József: A témazáró tudásszintmérés gyakorlati kérdései /Practical questions of theme--concluding measurement of achievement/. Tankönyvkiadó, Budapest, 1972. pp. 71-74 and a transformation table on pp. 142-143.

[2] Báthory, Zoltán: A tantárgyi osztályozás néhány mai jellegzetessége /Some current characteristics of subject grading/. Pedagógiai Szemle, 1968. No. 12. pp. 1077-1083.

[3] See in greater detail in: Nagy, József: A témazáró tudásszintmérés gyakorlati kérdései /Practical questions of theme-concluding measurement of achievement/. Tankönyvkiadó, Budapest, 1972. pp. 63-70.

Стандартная отметка

ЙОЖЕФ НАДЬ

В Венгрии существует пятибаллная система для оценки учеников. В прошлом десятилетии началось распространение использования тестов. Автор указывает на то, что шкала, используемая на квалификацию, должна ответить двум требованиям. Выполнение, соотносимое к требованиям, показывает процентный показатель. Но, это несравнимое с результатами других тестов. Поэтому нужны и стандартные показатели, дающие возможность сравнивания. Автор ознакомляет читателя с сущностью стандартных шкал и вырабатывает пятибаллную стандартную шкалу, которая органически связывается с ныне существующей системой.

Он создал таблицу к нестандартизированным тестам, которая даёт возможность для перечисления процентных показателей на так называемую частично-стандартную отметку /см. таблицу/.

Standardskala

Dr. J.Nagy

In Ungarn werden die Schülerleistungen mit Hilfe einer Fünfgradskala zensiert. Objektive Tests sind hur in den letzten zwei Jahrzehnten verbreitet worden. Der Verfasser behauptet, dass Qualifizerungsskalas zwei Funktionen zu erfüllen haben. In einem Test wird die Leistung durch einen prozentuellen Index mit den Fächern verglichen. Dieser Index kann aber mit den in anderen Tests dargestellten Leistungen nicht verglichen werden. Um einen allgemeinen Vergleich verwirklichen zu können, sind standarde Indizes notwendig. Der Verfasser beschreibt die Hauptcharakteristiken von Standardskalas und stellt den Typ einer standarden Fünfgradskala dar, die in das gegenwärtige ungarische Bewertungssystem von Schulleistungen integriert werden könnte. Für nicht standardisierte Tests wird eine Tabelle angegeben, mit derer Hilfe prozentuelle Indizes in eine sogenannte partielle Standarskala transforiemrt werden können. /s. die Abbildung/