Standardized mathematics testing in Sweden: The legacy of Frits Wigforss

Jeremy Kilpatrick and Bengt Johansson

Developed more than 50 years ago, the Swedish system of standardized testing as a means of moderating marks (or grades) is about to be replaced by a criterion-referenced measurement scheme. The principal developer of the original system, Frits Wigforss, was a psychologist and mathematics educator who understood the complex issues raised by any marking system and who attempted to use testing not to replace but to improve teachers' judgment. A close examination of the history of standardized mathematics testing in Sweden reveals the magnitude of Wigforss's contribution as well as its subsequent eclipse by the elevation of measurement technique over mathematical substance and a serious absence of attention to the educational and social consequences of changes in the system.

> *Det är hemskt att lägga ner så mycket arbete och så rinner det bara bort.*
> *[It is terrible to put in so much effort and have it just slip away.]*
> (Vera Wigforss, December 1983)

Sweden is one of very few countries in the world (the United States is another) that does not use a formal final examination to signal the completion of secondary school. Instead, employers rely on students' marks (or grades) for an indication of their academic preparation. Universities and other institutions of higher education use marks, work experience, and a university aptitude test in the admissions process. The aptitude test has been used since 1977 to admit students 25 years of age or older to the university (Henricson, 1987). Since 1991, the test has been open to all students, and a number of places in the university are available for

Jeremy Kilpatrick *is professor at the Department of Mathematics Education, College of Education, University of Georgia, Athens, USA.*
Bengt Johansson *is senior lecturer at the Department of Didactics, Göteborg University, Sweden.*

students wishing to use the test to compensate for low marks. (See Fägerlind, 1992, and Swedish Ministry of Education and Science, 1993c, for details of the Swedish educational system.)

The Swedish scheme of continuous assessment through course work, tests, and the marks given by teachers is regulated by means of standardized tests in certain subjects that are administered periodically through the year in elementary and secondary schools. The standardized tests are used not to give each student a mark but rather to permit the teacher to compare the performance of the entire class with the performance of the nation as a whole and thereby adjust his or her scale when it comes time to give a final mark. The use of standardized tests in this way has had remarkable success in stabilizing the marking system so that differences in the scale of marks across schools are quite small (Fägerlind, 1992, p. 83).

Standardized tests in mathematics (as well as in Swedish and English) have played a key role in making Swedish teachers' judgments about their students' performance a useful and valued part of educational decision making. Under a plan to begin in 1994, the regulation of marks by means of norms from a national sample will be replaced by a goals- and achievement-related governing scheme (i.e., management by objectives) (Swedish Ministry of Education and Science, 1993c, ch. 3; Utbildningsdepartementet, 1993a, 1993b). The new scheme has been proposed in part as a response to long-standing complaints about an assessment system commonly seen to be focused on ranking students and stimulating competition rather than encouraging students to work toward goals and standards.

As Sweden and other countries move from well-established, functioning systems of norm-referenced tests in mathematics toward new and untried schemes of criterion-referenced measurement, it may be helpful to examine how the Swedish system of standardized testing in mathematics developed and came to be used.

The current tests

> ***Regeringens bedömning:*** *Jämförbarheten i ett nationellt reglerat betygssystem upprätthålls genom riksgiltiga prov.*
> ***[The government's judgment:*** *Comparability in a nationally regulated mark system is maintained by nationally valid tests.]*
> (Utbildningsdepartementet, 1993a, p. 81) [1]

1 In this paper, all translations from documents in Swedish have been done by the authors.

Standardized test for compulsory school

The nine-year compulsory school was introduced in Sweden in 1962 after having been tried out for about 12 years. Swedish students begin school in Grade 1 when they are 7 years old and finish the compulsory school at 16. Standardized tests have been given to all students – at first in Grades 3, 4, 6, and 8 and later only in Grade 8 or 9 – as a means of regulating the marks they receive at the end of their schooling. Since 1991, there have been three versions of the mathematics test: one for students choosing the special course in mathematics that begins in Grade 7, one for students choosing the general course, and one for students in schools that do not have separate courses. The test is now compulsory for all students, and a new form is constructed each year.

The test has two parts, both scored by the student's teacher. The first part consist of short-answer items. There are from 24 to 30 items to be answered in 35 minutes. Answers are scored as right or wrong (except for a few questions in the version for the special course). Students are not allowed to use calculators on this part of the test.

Part 2 consists of application and modelling problems organized around a theme related to other subjects in the national curriculum. For example, in 1993, the theme was "A Week in the Scandinavian Alps". There were 10 problems to be solved in 75 minutes, and students were allowed to use calculators. The problems were scored from 0 to 2 points (except for one item for the general course). This part of the test is generally seen by teachers as a clear improvement over earlier tests (B-O. Ljung, 1991, p. 89; see Grevholm & Nilsson, 1992, for copies of the 1990 versions).

Standardized test for upper secondary school

Beginning in the mid-1960s, the formal final examination at the end of secondary school, which had been a feature of Swedish education since 1864 (Sjöstedt, 1963), was gradually abolished, to be replaced by a standardized test given during the school year. The first such test was administered in the 1967-1968 school year, and in 1970 the last group of students took the final examination. The purpose of the test is the same as for the test in the compulsory school. Scores from a national sample of students are used to regulate the mark scale. The test has two versions: one for the natural sciences and technology program; the other for the social sciences and economics program.

The test consists of 5 problems scored 0 to 2 points each and 5 scored from 0 to 3. Students have 3 hours and 45 minutes to complete the test. The test covers all sections of the syllabus; it does not have problems related to a single theme and involving other curriculum subjects. Scientific calculators are allowed throughout. (For details of the test, see B-O. Ljung, 1991; G. Ljung, 1991).

Development of the original test

> *Det får inte vara så, att läraren fordrar, att barnen skall förstå, hur han vill tänka, medan han själv inte förstår hur barnen vill tänka. [A teacher must not demand that children understand how he thinks when he himself does not understand how they think.]*
>
> (Wigforss, 1957, p. 178)

Frits Wigforss

The national tests and the accompanying system for standardizing teachers' marks in Sweden that has lasted with modifications for half a century are essentially the creation of one man, Frits Wigforss, an all-but-forgotten pioneer in mathematics education who was one of the first and most insightful Nordic educational psychologists and measurement experts. Wigforss was born in 1886 in Halmstad, Sweden. He was educated at the University of Lund, where he received a bachelor's degree in 1908. After several years as a secondary school teacher, he became a lecturer at the Rostad Teachers College in Kalmar. By 1919, Wigforss was senior lecturer in psychology, pedagogy, and mathematics, a position he held until his retirement in 1951. He died in 1953.

Wigforss contributed to many facets of education. He conducted the research and development work needed for the construction of diagnostic tests and of the national achievement tests in mathematics and Swedish. He also conducted various research studies using these tests as instruments to compare group performance. Among his important positions on the national educational scene, Wigforss chaired two committees on the mark system from 1937 to 1942, as well as the subject matter committee for the Swedish School Commission of 1946. The last committee outlined the curriculum guidelines for a unified compulsory school that eventually replaced the system of parallel schools separating students in the academic stream from the general stream after Grade 4. He was also asked to develop the first syllabi in mathematics for the new compulsory school.

Wigforss was a versatile and prolific textbook author. In addition to developing a series of textbooks and diagnostic tests for Grades 1 to 8, he wrote textbooks on mathematics education for teachers of the elementary and middle grades, a classic text on educational statistics, a textbook on how to measure achievement, and a book How to Play

Chess considered one of the best ever written in Swedish and still in print in its sixth edition.

Background of the tests

Wigforss had shown his interest in elementary school mathematics in a textbook for teachers he published in 1925, a book obviously influenced by Johannes Kühnel's classic German text Neubau des Rechenunterrichts (excerpts from which had been translated into Swedish and discussed by Dahlgren in 1923). When Wigforss began developing elementary mathematics tests some time before 1931, he had little to guide him. The first work on standardized testing in mathematics in the Nordic countries had been done in 1924 by Henning Meyer in Denmark (Meyer, 1926; Østlyngen, 1944). Wigforss knew about Meyer's work. He also knew, primarily through his colleague Carita Hassler Göransson who worked on tests of the Swedish language, about the work on testing that was being done in the United States, England, Germany, and Norway. [2] In particular, these Nordic researchers knew of the work of Walter S. Monroe (1918) in the United States and Philip B. Ballard (1923) in England, who had developed arithmetic tests. Wigforss tried to find a test used abroad that could help Swedish educators compare the performance of their students with that of students from other countries. His review of foreign tests, however, led him to conclude that they did not have the qualities he sought, nor did they fit the conditions needed for standardization.

Wigforss developed tests in computation and in problem solving. In the article setting forth his computation tests, he cited six antecedents of his work:

- Monroe's General Survey Scale in Arithmetic
- Monroe's Diagnostic Tests in Arithmetic
- The Compass Survey Tests
- The Compass Diagnostic Tests in Arithmetic
- A test by Philip Ballard
- A test by Henning Meyer (Wigforss, 1931, pp. 87-88)

He presented an analysis of the two tests from Monroe and said that the reasons not to use the other tests were the same. His main criticism of

2 Based on an interview on 10 october 1993 with Hassler Göransson, aged 102, who was Wigforss's collegue at the Rostad Teachers college and who, along with Wigforss and Ossian Åström, wrote the 1942 report of the committee on the mark system.

the Monroe General Survey was that it did not give specific information about students' skills in the different branches of arithmetic. The Diagnostic Tests were closer to what he was looking for, but his interest was limited to whole number arithmetic.

The national curriculum in mathematics of 1919 had set forth a view of prerequisites for learning that Wigforss saw as justifying the need for his mathematics tests:

> Every previous topic must be learned before a new topic is introduced. . . . Therefore it is important that the instruction proceed slowly and that the teacher carefully test whether the child can do each topic mentioned. (Wigforss, 1925, p. 7)

Wigforss relied on quantitative methods for measuring student knowledge and skill because the success of those methods in the natural sciences had led many Nordic educators to conclude that they would be of comparable value in education. He saw testing as an important aid to teachers both in marking the students' work and in making their teaching more effective. By means of what Wigforss termed a standard measure, the teacher could see whether the class was at the same level as the mean across all students taking the test.

Most important for Wigforss, however, was the information the test gave not for the class as a whole but for individual students. He saw that test results can help a teacher individualize instruction. After a test has been administered, it can be used for a thinking aloud interview in which the student tells how he or she has worked the items.

> Such an individual interview can be carried out by letting the student do his [or her] calculation aloud. "Let me hear how you calculate this item." If the teacher has enough patience, he [or she] can gradually get the student to think aloud and then often obtain interesting information about the student's method of working. As an example, the following experience can be related:
>
> On a test given in Grade 4 on multiplication and division combinations, there was a boy who did the items especially slowly. The slowness was remarkable because the boy had very good reasoning ability and had scored considerably better than average on the problem-solving test. An interview on how he managed to get the answers to the items showed that he often used highly complicated methods of calculation. Thus, he treated the item $53 \div 8$ as follows:
>
> First he noted that since $10 \div 8$ was 1, $50 \div 8$ would equal 5. But since $10 \div 8$ gives 2 as a remainder, the remainder of $50 \div 8$ must equal 10. Therefore, the remainder of $53 \div 8$ must be 13. But because $13 \div 8$ is 1 with 5 as a remainder, $53 \div 8$ must be equal to 6 with 5 as a remainder.

> Because the boy had not memorized the multiplication table, he was unable to use it in estimating a quotient. Through a brilliant method of calculation, he was nonetheless successful in getting right answers. His solution method simultaneously demonstrated his general ability and explained his low score on the test. The discovery of his calculation method led to energetic efforts to help him memorize the multiplication table. (Wigforss, 1939, p. 102)

Wigforss believed that a standard test was most important for use in mixed-age classes, where it is very difficult to judge whether a student's skills are good or poor. Because standard tests can be repeated again and again, they also can give students and teachers an opportunity to see the progress the students are making.

The computation tests

Table 1 (adapted from Wigforss, 1931, p. 90) contains information on the computation tests Wigforss constructed. The 13 tests were to be used in Grades 2 to 6. They were grouped into five forms for administration in successive class periods. Second graders took two forms, third graders three, and the rest five. Although the tests were speed tests, the time limits were more generous than those of comparable tests. For example, the Addition I test had a limit of 3 minutes in Grade 2, and the time was reduced at subsequent grades down to 1 1/4 minutes in Grade 6. As another example, Division II had a limit of 12 minutes in Grade 4, 10 in Grade 5, and 8 in Grade 6.

The time limits were set so that most of the children would be able to finish the first quarter of a test by the time the fastest students had finished all of it. Therefore the first part of the test was constructed so that students would meet everything of importance in it. The same principle was used in constructing the remaining three quarters of each test. The result was reasonable coverage of important content by each student and an approximately normal distribution of correct responses.

The items were constructed on the basis of a detailed analysis of all possible combinations for each operation, with a careful sampling of the numbers used. Wigforss took pains to distinguish multiplication items in which the standard algorithm yielded combinations such as $5 \times 7 + 4$ from those with $5 \times 7 + 8$, and he distinguished 31/7 from 29/7. He was sensitive to the processes students used to work the items (e.g., carrying to or borrowing from the next decade in the algorithms for multiplication and division).

The tests do not cover all parts of whole number arithmetic. For example, there is no item on multiplying two-digit numbers. The reason was that Wigforss was concerned about the amount of time each item

Name of test	Sample item	Number of items
Addition I	4 + 5	88
Addition II	7 + 8	100
Addition III	46 + 7	100
Addition IV	7 + 5 + 4 + 3 + 2 + 8 (given in vertical form)	40
Subtraction I	9 − 6	140
Subtraction II	13 − 8	100
Subtraction III	362 - 175	60
Multiplication I	7 x 8	180
Multiplication II	543 x 4 (vertical form)	64
Multiplication III	3876 x 8 (vertical form)	56
Division I	56 ÷ 8	172
Division II	60 ÷ 8 (items with remainder)	160
Division III	169 ÷ 21	32

Table 1. Wigforss's tests of arithmetic computation.

required. One of his criticisms of tests like those developed by Monroe was that individual items often took so long that a student could only do a few of them in the time available. The student's score would then be unstable. Monroe's tests were also too speeded (some had time limits of 30 seconds) and too difficult for students. Wigforss remarked that the mean score of 0.9 for one of Monroe's tests (less than one item correct) made the test more of a curiosity than a practical tool for educators.

In Wigforss's view, a standard test must be reliable. It should consist of many short items, most of which can be worked correctly by most students. The test should cover the topic so that good or bad scores do not occur by chance, a problem that is especially serious for students with low scores. Wigforss tried to make his tests suitable for less able students. According to Wigforss, Monroe's failure to provide for such students is reason enough not to use his tests.

Wigforss argued that the parts of the Monroe Diagnostic Tests dealing with rational numbers and decimals are not appropriate for a standardized calculation test because students can increase their performance dramatically simply by memorizing how to invert a fraction or where to put the decimal point. A standard measure of such skill would have little value. Wigforss believed that only those skills that students develop slowly after hard work are appropriate for standardized tests. He did not rule out the possibility of constructing a test in the arithmetic of rational numbers; he simply thought that Monroe's approach was inadequate.

In his first paper on the tests, Wigforss (1931) ends his comments by saying that his tests certainly do not preclude other types of tests and should be used instead as a complement to them. Many tests are best when they are not speeded. The criterion teachers should use for their usual tests – in contrast to standardized tests – is that "all items are correctly solved".

The problem-solving test

From the beginning of his work on test construction, Wigforss had planned a test of arithmetic problem solving, and in 1932 he began work on it. The problems he used were word problems of the sort commonly used in elementary school instruction. He did not attempt to distinguish among problem types. In his view, a problem-solving test

> should give information more about the student's general ability in problem solving than about how he [or she] can solve special types of problems. (Wigforss, 1934, p.135)

He argued that students who know the meaning of the four operations and how to handle them should be able to solve every problem. He tried to use familiar content. He wanted children without any skills in calculation with fractions to be able to solve the problems, and to solve most of them easily by mental arithmetic.

The items were ordered according to increasing difficulty. The first were to be so easy that most first graders could solve them; the last, so difficult that almost no sixth grader could solve them. The same test was given to all children in Grades 2 through 6. Sample items are the following:

> 1. Anna gives Greta and May 15 badges. She gives 6 badges to Greta. How many does May get?
>
> 10. A plank 3 meters long is laid across a ditch 19 decimeters wide. The plank overlaps 6 decimeters on one side. How much does it overlap on the other side?
>
> 20. What is the largest number that gives 2 as a remainder when both 212 and 296 are divided by it?

By using the same test at each grade, a teacher could compare the performance of children in different grades. Wigforss was aware that the reliability of the test was lowered by having only one test and that some children would be solving problems that were too hard or too easy for them. He thought that teachers could easily deal with that situation by telling the students that the same test was being given at all grades.

The younger students would be proud to have the same problems as the older ones, and the older ones would be more careful in their work if they know that younger students were being given the same problems. He claimed that although the reliability of the test was not adequate for making judgments about a single child, it was satisfactory at the class level. He could see the importance of having a different test for each grade and later was able to do that for Grades 2, 4, and 6.

Wigforss was also aware of the difficulty of devising a test with items having appropriate levels of difficulty. He conducted a pilot study with more than 15 000 students in Grades 2 to 6 in which he administered three tests of 25 items each. He then constructed two versions of the final test that were of approximately equal difficulty, each having 20 items. The students were given 45 minutes to do the test. Two versions were developed so that teachers could keep students sitting close together from helping one another.

Additional studies

As part of his developmental work on the computation tests, Wigforss (1934) investigated the relative difficulty of all possible combinations in the tables (from 0 to 10) for the four arithmetic operations. The rankings of difficulty were presented in lists for teachers to use in their instruction. In the same investigation, he studied the relation between speed and accuracy in mechanical arithmetic. He concluded that the two were so strongly related that teachers need not separate them when giving marks. His concern that the students' speed of responding to the computation items might be hampered by their slowness in writing led him to examine as well the relation between calculation speed and writing speed. He found that writing speed had only a small influence on calculation speed.

Wigforss (1937a) also investigated the types of errors students made in adding numbers from 1 to 10. He saw the importance of skill with these combinations for work with larger numbers:

> The more fundamental knowledge teachers have about the difficulty of and error types for the different number combinations, the better they can plan their teaching and the more likely they are to uncover the weaknesses they must try to remedy. (p. 17)

Wigforss classified the errors into seven categories and demonstrated, among other things, a dramatic change in the pattern between Grades 1 and 6. For example, counting-on errors fell from 39% of the errors in Grade 1 to 7% in Grade 6 and zero-combination errors from 25% to less than 1%, whereas errors in choice of operation rose from 7% to 74% (Wigforss, 1937a, p. 30).

In 1946, Wigforss published the results of an extensive study of children's arithmetic when they enter school. The readiness tests he developed for arithmetic, and also for Swedish, were widely used in Sweden for several decades thereafter. This work on readiness was influential all across the Nordic countries (Meyer & Gregersen, 1951).

Standardization and marks

Problemet att fixera de särskilda betygsgradernas betydelse erbjuder svårigheter av såväl teoretisk som praktisk natur. . . . Det är framför allt kravet, att kunskapsbetygen skola vara jämförbara, som vållar svårighet. Om all fordran på jämförbarhet olika elever emellan toges bort, skulle det problem, som här diskuteras, upphöra att existera.
[The problem of establishing the meaning of the specific marks presents both theoretical and practical difficulties. . . . In particular, the requirement that the marks for knowledge be comparable causes difficulty. If the demand for comparability among different pupils were removed, the problem discussed here would cease to exist.] (Statens offentliga utredningar (SOU), 1942, p. 50)

The marking system

The first Swedish school act of 1571 did not attempt to regulate marks (SOU, 1942, p. 11). In the 17th and 18th centuries, some rules were developed for evaluating applicants' qualifications for entrance into secondary grammar school (Läroverk), but not until the 1820 curriculum for that school were marks discussed and prescribed. The marks, derived from Latin phrases, were A, B, C, and D. Each student was to receive a mark for manners and behavior and for each subject a mark for knowledge and effort. The marks for knowledge and effort were characterized as follows:

A = passed with great distinction in insight and diligence,
B = passed with insight and diligence,
C = defensible insight and diligence, and
D = not enough insight and diligence.

In the 1859 secondary grammar school curriculum, a mark between A and B – AB – was introduced and described, and in 1897 a 7-level system of assigning marks was prescribed for the elementary school (folkskola). In 1905, the government introduced into Swedish secondary grammar schools a similar 7-point scale: A, a, AB, Ba, B, BC, and C, with BC and C as failing marks. The scale was given a set of numerical equivalents (A = 3, a = 2.5, AB = 2, Ba = 1.5, B = 1, BC = 0.5 or 0, C = 0) in 1933. These numbers were used in averaging marks but were normally not reported to students.

Wigforss was concerned about the way in which marks were being given by different teachers. A student who received a mark of A from one teacher might have received a B from another. Wigforss thought that teachers' marks should be set with reference to an objective scale. He standardized his tests on a national sample of Swedish students so that teachers could use them in giving marks:

> One of the most important goals of the standardization effort is to create a standard meaning for the marks. (Wigforss, 1932, p. 8)

Wigforss's approach was to devise tables that would allow teachers to convert a score on each of his tests – whether it was a single student's score or the average for the class – into a mark. The purpose would be not to dictate the assignment of marks to any student but rather to show teachers how the marks they were giving compared with those of other teachers. Wigforss acknowledged the difficulty of combining marks from a set of tests to yield a single mark. The tables he constructed were to provide guidance:

> Without standard tables it would certainly be extremely hard to remedy the confusion about setting marks. (Wigforss, 1933b, p. 184)

The standardization process

In 1932, Wigforss published an article in which he gave tables for converting computation test scores into marks for A Form classes (classes homogeneous in age) in Grades 2 to 6. (B Form classes contained students of different ages.) The norms were based on data from 16 839 students. There were two types of tables, one for individual students' marks and one for classes.

Wigforss noted that in many studies in which frequency tables had been calculated, the distribution was approximately normal and that few scores deviated from the mean by more than three or four standard deviations. For each of his tests, he calculated the mean score and the standard deviation and transformed the raw scores so that they had a normal distribution. Then he defined an interval from plus to minus 1/2 standard deviation from the mean as a transformed score equivalent to the mark Ba. The remaining marks were defined by intervals 1 standard deviation in width, with C defined by any score less than 2 1/2 standard deviations below the mean and A by any score more than 2 1/2 standard

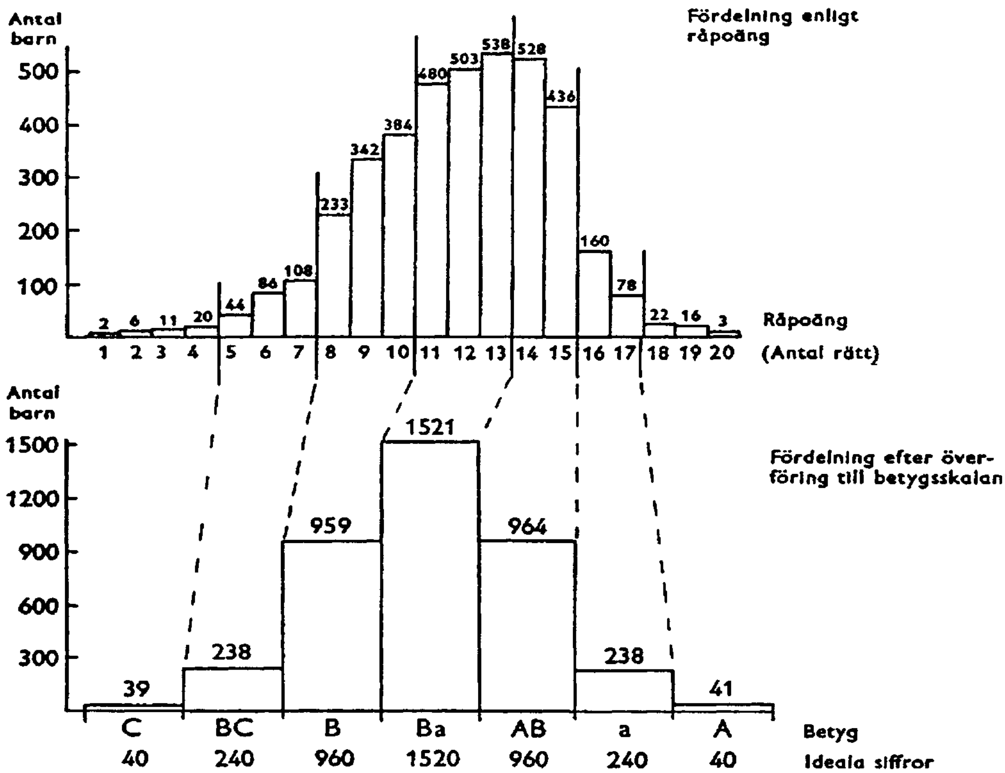

Figure 1. Distribution of marks.

deviations above it (see Figure 1, from Husén, Björnsson, Edfeldt, & Henrysson, 1956b, p. 32).

This distribution of marks, according to Wigforss, fit well the practice of many teachers, and he argued for using the same system for individual students and for classes. He also expressed the hope that in 10 years a new study would show an increase in students' scores and the need for a new standardization. Unfortunately, when that time came, the students' scores had not risen enough to make re-standardization necessary (Henricson, 1987).

In scoring the students' responses, Wigforss did not distinguish a student who got, say, 40 right answers and no wrong answers from a student who got 40 right answers and 20 wrong answers. He was aware of the problems this procedure caused but said that they were not so severe in tests having many short, relatively easy items as compared with tests composed of long, difficult problems. He found a difference between boys and girls on the computation tests that would later be observed by some other researchers: The girls had a higher mean, and the boys had a larger standard deviation.

Wigforss recommend that teachers give the computation tests at least twice a semester, or at least give a sample of the tests. He argued for the diagnostic value of the different tests and for the use of various combinations of tests. At the same time, however, he was aware of the time needed for scoring them. For example, 35 students taking 13 tests each that together contained 1 292 items meant that their teacher would have to score 45 220 items. Wigforss estimated it would take a teacher 6 or 7 hours to score that many items. His solution was to suggest that the students change papers and score each other's tests, with the teacher reading the answers.

After Wigforss had standardized his computation tests on A Form classes, his next step was to study the mixed-age classes. In B-1 Form classes, the students' ages differed by one year, and in B-2 Form classes, they differed by more than a year. Wigforss found that at all grades the performance in the A Form was superior to that in the B Forms, with the largest differences occurring in the higher grades. He described the differences in terms of the backwardness of Form B compared with Form A. For example, the greatest backwardness for Form B-1 was at Grade 6, where the difference was 0.61 years. At all grades, the backwardness of Form B-2 was greater than that of Form B-1.

The standardization of the problem-solving test was begun in 1933. It involved more than 20 000 children in Grades 2-7. In 1938, Wigforss published the level of difficulties for every item in the first problem-solving test at Grades 2, 4, and 6. A Norwegian study (Ribbskog, 1936) using the same test had yielded almost identical levels of difficulty. Wigforss concluded that the test was acceptable even though it did not contain items at every level of difficulty. Good items are more important than an even distribution of item difficulty (Wigforss, 1938a, p. 4).

Tests and the marking system (to 1953)

Already in 1933 – 3 years before the government authorities began to be interested in the matter – Wigforss had explained how he thought his computation and problem-solving tests should be used to control the marks scale on the national level (Wigforss, 1939). He had previously studied the marks given in different classes:

> Investigation of the mark scales and the giving of marks shows how chaotic the situation is and how something needs to be done to improve it. The need for standardized tests is very great. (Wigforss, 1933b, p.176)

In an article published in a journal of the Rostad Teachers College, Wigforss described the "mark situation" in Sweden using strong language such as "confusion", "miserable", and "chaos" (Wigforss, 1934, p. 10) and recommending that some order be imposed.

In Wigforss's opinion, students' marks in arithmetic should depend on both their computational skill and their problem-solving ability. But how were teachers to combine the two marks into one? Wigforss (1934) argued that good teachers were giving more weight to problem solving than to computation. He suggested that "we let the problem-solving mark weigh twice as much as the other mark" (p. 10). By using the numerical equivalents, a teacher could calculate a weighted average to arrive at the final mark:

> My suggestion is then that the teacher calculate the mark by taking the problem-solving mark times 2, add the mark for calculation, and divide the sum by 3. (pp. 10-11)

On the question of how to make the marks objective, Wigforss (1934) proposed the following:

> The standard test shall regulate the marks scale. From the standard test, the teacher gets information as to whether the scale he [or she] has used is normal, too high, or too low. Thus, the standard test shall not decide a single student's mark, but shall indirectly influence it; namely, if the test shows that the whole marks scale must be changed. The teacher will thereby get information by comparing the mean of the class on the standard test with the mean of the marks he [or she] was intending to give the students....
>
> But beside the correction in terms of the class mean, another correction of the mark scale may also be needed. . . . If a teacher, after having corrected his [or her] marks with respect to the mean, finds that the deviation is quite different from the deviation in the marks the students received on the standard test, the marks need to be adjusted up or down until the difference in deviations is no longer so great. (pp. 11-12)

In his article he gave concrete examples of how the two types of modifications could be done for a class. He concluded the article with the following claim:

> It seems to me that the present proposal could be realized without too many problems. I have used it myself in practice and found that it functions in a satisfactory way. And the possibility of getting objective marks seems already to be in our hands. (p. 14)

In 1934 and 1935, Wigforss made an extensive analysis of the results of the problem-solving tests, but he did not have time to write up his findings (1939, p. 45). He had become involved in 1936 in a series of studies for the Swedish Society for a Psychological-Pedagogical Institute that were connected with an International Examinations Inquiry, sponsored by the Carnegie Corporation, the Carnegie Foundation, and the International

Institute of Teachers College, Columbia University, and conducted from 1931 to 1939 (P. Monroe, 1931, 1936, 1939).

Wigforss (1937b) first looked at the value of the national entrance examination for predicting school performance in the 4- or 5-year lower secondary school (realskola, below the 3- or 4-year gymnasium), using data over 7 years from Kalmar Lower Secondary School. He found that the test made a large number of errors in prediction, and his data suggested that certificates from the elementary school might do a better job of prediction, especially if variations among teachers in the marks on those certificates could be reduced or eliminated. Because the data from the 4-year lower secondary school were meager, he began a more extensive follow-up investigation, but the results were never published. Instead, he conducted and reported an investigation of the effects of an attempt he had begun in 1937 to get uniformity in the certificates being awarded by different teachers in the elementary schools of Kalmar (Wigforss, 1941).

Wigforss found considerable within-teacher consistency but little between-teacher consistency in the average certificates given across elementary school classes. He argued for the use of standard tests, not to replace the teacher's judgment but rather as a means of modifying whatever scale the teacher had used in awarding the certificates. He had used existing tests in his investigation but proposed that a series of improved mathematics tests be constructed, noting the dangers as well as the advantages of "building standard tests on special parts of the curriculum" (Wigforss, 1941, p. 131).

In 1939, the government appointed Wigforss to chair a committee to conduct an inquiry on whether marks in the six-year elementary school could replace the entrance tests for the lower secondary school. The committee consisted of Wigforss, Hassler Göransson, and Ossian Åhström. The inquiry, which took 3 years (SOU, 1942), drew heavily on research Wigforss conducted between 1931 and 1941 and on an earlier report on marks he had done for the government (SOU, 1938). Hassler Göransson also studied the history of the marking system. Her work showed that the marking practice in schools often departed substantially from the prescribed system, mostly through the use additional marks beside the official ones. When teachers thought that too few marks were available for them to use, they inserted marks in between. A survey done as part of the inquiry showed that in 1940, 90% of the elementary school teachers were using an unofficial marking scheme, with + and –, that had 13 levels instead of 7 (SOU, 1942, p. 62).

From the beginning of this century, marks in Swedish schools had been unified in expression but not in meaning. The meaning of the marks in secondary school was controlled in part by national examinations given at the end of the lower and the upper divisions. In the elementary school, which Wigforss looked at, there was no such control. The committee discussed the feasibility of a compulsory examination system for the end of the elementary school but concluded that the price would be too high in terms of the pressure it would put on instruction. They proposed instead that a nation-wide, voluntary system of standardized tests be developed to help elementary school teachers adjust their marks, which could therefore be used to replace the lower secondary school entrance tests. They also proposed doing away with failing marks.

In a special supplement to their report (SOU, 1942), the committee presented a proposal for the tests for elementary school that in the 1943-1944 school year became the first standardized tests in Sweden approved by the National Board of Education to be used in the whole country. Table 2 gives an overview of the mathematics tests (Østlyngen, 1944, p. 53).

Table 2. First official standardized mathematics tests in Sweden.

Test	Grade	Number of items	Time in minutes
A. Tests in Applied Arithmetic			
I. Problem Test 1	2, 4, and 6	20-21-20	35-40-40
II. Problem Test 2	2, 4, and 6	16-20-24	15 (for 4 and 6)
III. Arithmetic Achievement Test	4 and 6	15-16	40
IV. Coin Test	2	20	10
V. Cube Test	4/6	15	5
B. Tests in Mechanical Arithmetic			
VI. Computation Test	4 and 6	25-30	40
VII. Addition Test	2, 4/6	48-42	8-6-5
VIII. Subtraction Test	2, 4/6	54-60	6-10-6
IX. Multiplication Test	4/6	60	8-6
X. Division Test	4/6	24	14-12
XI. "Which Answer Is Correct?"	4 and 6	30	30

We take the description of the tests from Østlyngen (1944):

> Altogether there are 20 different tests, and all except Problem Test 2 for Grade 2 exist in two equivalent forms. Five tests are designed for Grade 2, 10 for Grade 4, and 10 for Grade 6; 5 tests are the same for Grades 4 and 6.

Problem Test 1 is aimed at assessing not the special mathematical knowledge the students have achieved through instruction but rather their general capacity to think mathematically. It consists of application items with easy calculations. The items are ordered by increasing difficulty, and the last items are so hard that very few students in that grade can solve them. The students are allowed to use scratch paper, which is handed in with the test.

Problem Test 2 has a similar aim as the previous test, but assesses much more of the students' capacity for speed in thinking because the time is much shorter. All calculations are to be done mentally. In the Grade 2 test (Mental Arithmetic Test), the items are to be read aloud to the students, and the students just write the answer.

The Arithmetic Achievement Test differs from the problem tests in that the items treat specific topics taught in the particular grade. It is more similar to the tests many teachers use. The items are ordered by difficulty. The children are allowed to use scratch paper.

The Coin Test is like a multiple-choice test. The items are of the type "What 2 coins give 60 öre altogether? What 3 coins give 55 öre altogether?" The student is to put the correct number under one or more headings for 50-, 25-, 10-, 5-, and 1-öre coins.

The Cube Test contains pictures of cubes piled in different ways, and the student is to find out how many cubes there are in each pile. The teacher shows the students an example in advance using real cubes – the test did not meet the committee's expectations.

The Computation Test consists of a mixed items on the four operations. The students are given relatively generous time (40 minutes) to do the test, and they are to use scratch paper.

The tests in addition, subtraction, multiplication, and division are "speed calculation tests". The students are to show how fast they can calculate with reasonable accuracy. All solutions are to be made on the test paper, and the student is not to do any item twice. All items in the same test are of the same kind, but they are grouped. In the addition tests, the items are increased by 1 addend from one group to the next, and in the multiplication tests, the number of digits is increased by 1. Thus, the Addition Test for Grade 2 starts with 3 addends and ends with items consisting of 10 addends.

"Which Answer Is Correct?" (estimation test in mechanical arithmetic) consists of mixed items from the four operations. The student is to put a line under the one of four choices that is closest to the correct answer. Only mental arithmetic is to be used. (pp. 53-54)

Wigforss made a minor revision of the standardized tests that went into effect in 1947. In 1951, when he checked the standardization, he found that no further changes were needed (Henricson, 1987, p. 1).

Subsequent developments (1953-1993)

> *Av flera skäl . . . var det nödvändigt att göra denna omarbetning helt från grunden. . . . För att få bättre grepp om vilka målsättningar, som f.n. är implicerade i den konkreta undervisningen i berörda ämnen, fick vissa faktoranalytiska undersökningar göras.*
> *[For many reasons . . . it was necessary to do this revision completely from the ground up. . . . For a better grasp of what goals are currently implied in the concrete instruction in the subjects concerned, some factor analytic investigations had to be done.] (Husén et al., 1956a, p. 10)*
>
> *Med ett par undantag har alltså inga principiella förändringar av standardprov och centrala prov ägt rum sedan starten på 40- respektive 60-talen.*
> *[With a few exceptions, there have been no major changes in the standardized tests for the compulsory school and for the upper secondary school since the tests were begun in the 1940s and 1960s, respectively.] (Henricson, 1987, p. 31)*

When Wigforss died, the responsibility for the development of the standardized tests was given to Torsten Husén at the Stockholm Institute of Education. According to Husén et al. (1956a), new standardized tests were needed for the new curriculum being developed for the elementary school. The first step was to develop a parallel test for each of the existing tests at Grades 4 and 6. When the 9-year compulsory school was established in 1962, it was divided into three key stages, Grades 1-3, 4-6, and 7-9. It was natural, therefore, to develop a new set of standardized tests, at Grades 3, 6, and 8 or 9. With the increased use of computerized data processing, Husén and his colleagues were able to developed a quick method for sampling and norming (B-O. Ljung, 1991).

In 1965, the job of developing and standardizing the tests was shifted to the Swedish National Board of Education. Then in the mid-1980s, partly because of criticisms of the Board's inability to do research on the test, the Ministry handed the responsibility for the tests back to the Stockholm Institute of Education (B-O. Ljung, 1991).

A five-level mark scale had been instituted in the compulsory school in 1962 (Kungliga Skolöverstyrelsen, 1962) and in the upper secondary school in 1965 (Kungliga Skolöverstyrelsen, 1965) to replace the seven-level scale. As before, the scale followed the normal distribution:

Mark:	1	2	3	4	5
Percent:	7	24	38	24	7

This system is still in use in the upper secondary school (for the natural science program the mean has been increased by 0.3), but in the 1980 curriculum for the compulsory school the system was changed slightly:

> The marks still are to be from 1 to 5, and the mean should be 3 in the country as a whole. Specific percentages will not be set for the different mark levels. Normally, however, the number of 4s and 2s in a class should be more than the number of 5s and 1s, respectively. (Skolöverstyrelsen, 1980, p. 39)

One sees in this movement away from the normal curve an attempt to hold on to some aspects of norm-referencing while abandoning others. Fixing the mean is a normative procedure, but failing to set percentages leaves open the question of what mathematical model is being used. Each teacher is presumably free to choose any distribution of marks whatsoever as long as most teachers' marks satisfy the prescribed inequalities and as long as the mean across the country is 3. This proposal has been termed "criterion-referenced" by some, but it appears to be more a combination of loose local and national norm-referencing. The effort by Wigforss and his colleagues to help teachers put their marks on a common scale with fixed central tendency and fixed dispersion has been steadily eroded.

A new evaluation system

> *As part of the policy of stressing quality and making school more demanding the marking system has been revised by the Commission on Marks. . . . Marks will be described in terms of qualities instead of norms, implying a progression of knowledge. The present system, where marks are group-referenced, will thereby be replaced by a system where the marks are goal-referenced and thus related only to the achievements of the individual assessed in relation to criteria that are clearly accounted for in the new curricula and syllabi. (Swedish Ministry of Education and Science, 1993c, pp. 136-137)*

A goal- and achievement-related marking system

Under the proposed Swedish school curriculum to begin in 1994, the syllabi will specify the criteria for awarding marks (Swedish Ministry of Education and Science, 1993a, 1993b). In most grades, schools themselves will decide when marks are to be awarded and in what form. The system Wigforss began in 1944 had gradually eliminated failing marks after 1962. In the new system, for the first time since 1970, some Swedish students will be receiving grades of not passed.

The compulsory school will use a 6-point scale from A to F, with F signifying not passed. Teachers will be given criteria for A, C, and E, and will be expected to use B and D as marks between A, C, and E. The criteria for a mark of E will set the pass level. Although the system is not supposed to be normative, teachers are told that in principle all pupils should reach the pass level (Utbildningsdepartementet, 1993a, pp. 75-77).

The upper secondary school will use a 4-point scale (passed with high distinction, passed with distinction, passed, and not passed). In the upper secondary school, marks will be "awarded on the basis of courses and not individual subjects or for each year" (Swedish Ministry of Education and Science, 1993b, p. 9). The criteria are explicitly assumed not to be group related, that is, norm-referenced (Utbildningsdepartementet, 1993b, pp. 54-55).

Proposed changes in assessment

The syllabi for the compulsory school specify goals to be achieved by the end of Grades 5 and 9. National tests in mathematics will be developed for administration at the end of Grades 5 and 9 in all public sector schools. The tests at Grade 9 will be used to regulate the awarding of marks. Diagnostic tests in reading, writing, and arithmetic will be administered at the end of Grades 2 and 7 (Swedish Ministry of Education and Science, 1993a, pp. 9-11).

In the upper secondary school, so-called central tests will be developed to support the awarding of marks in certain subjects (Swedish Ministry of Education and Science, 1993b, p. 9). Students will be allowed to retake these tests to achieve a higher mark in subjects studied previously.

Lessons Wigforss taught us

The Swedish system for moderating students' marks by using a set of nationally standardized tests to help teachers construct comparable marking scales was an innovative and highly successful attempt to address a problem many countries face today. No effort to moderate marks can yield perfect comparability, as Wigforss and his colleagues knew well, but the system offered teachers a workable scheme in which their professional judgment could play a major role. Educators and policy-makers who are trying to help teachers become more informed and more confident evaluators of their students' work in mathematics should examine the Swedish system carefully. It was the product of much thought and experience.

Wigforss understood the limits of the standardized test system. He knew that one test cannot do everything. It cannot simultaneously provide detailed diagnostic information that teachers need for instruction and the succinct information that is needed for the construction of national norms. He also recognized how important it was to take time to investigate the effects of introducing a new system of assessment. He saw that teachers needed to understand what they were being asked to do and that sweeping changes in educational practice cannot be legislated, nor do they occur overnight.

Moreover, Wigforss could see the value of creating a strong rationale and mechanism for change. The system he and his colleagues devised was built on population norms not because he had no criteria for good performance, but because he saw clearly that any criterion for judging students' learning is socially determined. One cannot establish absolute standards for assigning marks. Marks take their meaning from the context within which they are used. Wigforss knew that no conceivable system can allow for unambiguous interpretation, and that comparability across groups demands some means of equating. One can quarrel with his choice of the normal distribution or with the way in which the marks were set, but the system itself was simple and elegant. It gave testers and teachers alike the freedom to adapt to changing circumstances.

Unfortunately, after Wigforss passed from the scene, the system became dominated by technical considerations and not by attention to the substance of the tests. Although the composition of the items was changed periodically to reflect new syllabi, there was no deep, thorough-going, and intellectually powerful investigation of how the test structure might be modified to reflect and support the mathematics students were expected to learn. The garden Wigforss had planted was not cultivated. The bureaucrats and technicians managed the system; they did not improve it.

No one can question the desirability of linking achievement in school to clearly specified goals and objectives. So far, however, a detailed system of standards referenced only to goals and not to the performance of student groups has still to be established. The work of Frits Wigforss reminds us that before we mandate new educational visions, we ought to see whether we can develop, try out, and implement what we are talking about.

References

Ballard, P. B. (1923). *The new examiner.* London: Hodder & Stoughton.

Dahlgren, T. (1923). Om undervisning i räkning [On teaching arithmetic]. *Skola och samhälle, 4,* 163-213.

Fägerlind, I. (1992). Beyond examinations: The Swedish experience and lessons for other nations. In M. A. Eckstein & H. J. Noah (Eds.), *Examinations: Comparative and international studies* (pp. 79-87). Oxford: Pergamon.

Grevholm, B., & Nilsson, M. (1992, May). Gender differences in mathematics in Swedish schools. Paper presented at the Colloquium on Differential Performance in Assessment in Mathematics, University of Birmingham.

Henriksson, S-E. (1987). *Skolöverstyrelsens provverksamhet – en översikt 1965-1985* [The administration of tests by the National Board of Education—An overview 1965-1985] (Rapport R 87:3). Stockholm: Skolöverstyrelsen.

Husén, T., Björnsson, C.H., Edfeldt, Å.W., & Henrysson, S. (1956a). *Standardproven. En redogörelse för konstruktion och standardisering* [Standardized tests: A report on construction and standardization]. Stockholm: Almqvist & Wiksell.

Husén, T., Björnsson, C.H., Edfeldt, Å.W., & Henrysson, S. (1956b). *Betyg och standardprov. En orientering för föräldrar och lärare* [Marks and standardized tests. An orientation for parents and teachers]. Stockholm: Almqvist & Wiksell.

Kungliga Skolöverstyrelsen. (1962). *Läroplan för grundskolan* [Curriculum for the compulsory school]. Stockholm: Author.

Kungliga Skolöverstyrelsen. (1965). *Läroplan för gymnasiet* [Curriculum for the upper secondary school]. Stockholm: Author.

Ljung, B-O. (1991). Prov i matematik [Tests in mathematics]. In T. Lindblad (Ed.), *"Allt är relativt". Centralt utarbetade prov i svenska, matematik, engelska, franska, tyska, fysik och kemi – beskrivning, analys, förslag* ["Everything is relative": Nationwide tests in Swedish, mathematics, English, French, German, physics, and chemistry—descriptions, analyses, suggestions] (Rapport 1991:03, pp. 85-96). Göteborg: Göteborgs universitet, Institutionen för pedagogik.

Ljung, G. (1991). *Centrala prov i matematik på NT-linjerna 1985-1989. Analys av innehåll och resultat* [Standardized tests in mathematics in the natural and technical science programs at the upper secondary school 1985-1989] (Rapporter från PRIM-gruppen No. 6). Stockholm: Högskolan för lärarutbildning i Stockholm, Institutionen för pedagogik.

Meyer, H. (1926a). Om standpunktprøver [On standardized tests]. *Vor Ungdom,* pp. 37-50.

Meyer, H. (1926b). Nogle Standardprøver i de fire Regningsarter [Some standardized tests on the four arithmetic operations]. *Vor Ungdom,* pp. 51-84.

Meyer, H., & Gregersen, T. (Eds.). (1951). Skolemodenhedsproblemer. 15 artikler av nordiske forfattare [School readiness problems: 15 articles by Nordic authors]. *Psykologisk-Pædagogisk Bibliotek. Bind XIII.* København: Ejnar Munksgaard.

Monroe, P. (Ed.). (1931). *Conference on examinations* (Eastbourne, England, May 23-25, 1931). New York: Columbia University, Teachers College, Bureau of Publications.

Monroe, P. (Ed.). (1936). *Conference on examinations* (Folkestone, England, June 8-10, 1935). New York: Columbia University, Teachers College, Bureau of Publications.

Monroe, P. (Ed.). (1939). *Conference on examinations* (Dinard, France, September 16-19, 1938). New York: Columbia University, Teachers College, Bureau of Publications.

Monroe, W. S. (1918). *Measuring the results of teaching.* Cambridge, MA: Riverside Press.

Ribbskog B. (1936). *Standpunktprøver i regning. Norsk standard av Rostads standardprov i problemløsning. En undersøkning av ca. 16000 elever i folkeskolens 2-7. klasse* [Standardized tests in arithmetic: Norwegian standardization of the Rostad standardized tests in problem solving—A study of about 16000 students in Grades 2-7 of the elementary school]. Oslo: Komiteen for pedagogisk forskning.

Skolöverstyrelsen. (1980). *Läroplan för grundskolan* [Curriculum for the compulsory school]. Stockholm: Liber Utbildningsförlaget.

Statens offentliga utredningar. (1938). *Intagning av elever i första klassen av de allmänna läroverken* [Intake of students in the first class of the lower secondary school] (1938:29). Stockholm: Ecklesiastikdepartementet.

Statens offentliga utredningar. (1942). *Betänkande om betygssättningen i folkskolan* [Government report on mark setting in the elementary school] (1942:11). Stockholm: Ecklesiastikdepartementet.

Sjöstedt, C. E.(Ed.). (1963). *Studentexamen 100 år* [One hundred years of the final examination for the upper secondary school]. Stockholm: Natur och Kultur.

Swedish Ministry of Education and Science. (1993a). *A new curriculum for the compulsory school.* Stockholm: Author.

Swedish Ministry of Education and Science. (1993b). *A new curiculum for the upper secondary school.* Stockholm: Author.

Swedish Ministry of Education and Science. (1993c). *The Swedish way towards a learning society.* Stockholm: Author.

Utbildningsdepartementet. (1993a). Prop. 1992/93:220. *En ny läroplan för grundskolan och ett nytt betygssystem för grundskolan, sameskolan, specialskolan och den obligatoriska särskolan* [A new curriculum for the compulsory school and a new marking system for the compulsory school, Lapp school, special school, and compulsory school for the mentally retarded]. Stockholm: Author.

Utbildningsdepartementet. (1993b). Prop. 1992/93:250. *Ny läroplan och ett nytt betygssystem för gymnasieskolan. komvux, gymnasiesärskolan och särvux* [A new curriculum and a new marking system for the upper secondary school, municipal adult education, the special secondary school, and adult education for the mentally retarded]. Stockholm: Author.

Wigforss, F. (1925). *Den grundläggande matematikundervisningen* [The foundations of mathematics instruction]. Stockholm: Magnus Bergvalls.

Wigforss, F. (1931). Rostads standardprov i sifferräkning [The Rostad standardized test in numerical arithmetic]. *Rostads elevförbunds årsskrift, **28***, 86-100.

Wigforss, F. (1932). Rostads standardprov och tabeller i mekanisk räkning samt några råd för deras användande [The Rostad standardized tests and tables in mechanical arithmetic and some advice on their use]. Särtryck ur *Rostads elevförbunds årsskrift, **29***, 1-12.

Wigforss, F. (1933a). Färdigheten i mekanisk räkning i folkskolan enligt Rostads standardtabeller [Skills in mechanical arithmetic in the elementary school according to the Rostad standard tables]. *Skola och samhälle, **14***, 191-222.

Wigforss, F. (1933b). Rostads elevförbunds räkneundersökning [The Rostad student union arithmetic study]. *Rostads Elevförbunds årsskrift, **30***, 167-168.

Wigforss, F. (1933c). Färdigheten i mekanisk räkning i B:1—skolorna i jämförelse med A-skolorna [Skill in mechanical arithmetic in B:1 schools compared with A-schools]. *Rostads Elevförbunds årsskrift, **30***, 169-184.

Wigforss, F. (1934). Rostads räkneundersökningar [Rostad arithmetic studies]. *Särtryck ur Rostads elevförbunds årsskrift, **31***, 1-43.

Wigforss, F. (1937a). Hur reda sig skolbarnen med additionsuppgifterna inom talområdet 1-10? [How do schoolchildren do addition combinations with numbers from 1-10?] *Rostads elevförbunds årsskrift, **34***, 14-31.

Wigforss, F. (1937b). *The entrance examination in view of later school performance.* (Publications of the Swedish Society for a Psychological Pedagogical Institute, I). Stockholm: Norstedt & Söner.

Wigforss, F. (1938a). En svårighetsskala för räkneproblem [A difficulty scale for arithmetic problems]. *Rostads elevförbunds årsskrift, **35***, 1-5.

Wigforss, F. (1938b). *Matematisk statistik för pedagoger* [Mathematical statistics for educators]. Stockholm: P.A.Norstedt & Söner.

Wigforss, F. (1939a). Rostads pedagogiska forskningsarbete [Educational research at Rostad]. *Rostads elevförbunds årsskrift, 36*, 1-5.

Wigforss, F. (1939b). *Kunskapsprövningar* [The assessment of knowledge] (Pedagogiska skrifter, Häfte 165). Lund: Sveriges allmänna folkskollärarförenings litteratursällskap.

Wigforss, F. (1941). *Marks and certificates in the primary school.* (Publications of the Swedish Society for a Psychological Pedagogical Institute, III). Stockholm: Norstedt & Söner.

Wigforss, F. (1946). Barnens färdighet i räkning vid skolgångens början [Children's arithmetic skills on entering school] (S.A.F.:s psykologisk-pedagogiska undersökningar 2) (*Pedagogiska skrifter 191*). Göteborg: Sveriges allmänna folkskollärarförenings litteratursällskap.

Wigforss, F. (1957). *Den grundläggande matematikundervisningen* [The foundations of mathematics instruction] (6:e uppl.). Stockholm: Bergvalls.

Østlyngen, E. (1944). Standpunktmåling i regning i de nordiske land [Standardization in arithmetic in the Nordic countries]. *Norsk Pedagogisk Årsbok for 1944*, pp. 48-62.

Standardiserade matematikprov i Sverige: Arvet från Frits Wigforss

Sammanfattning

För mer än 50 år sedan utvecklades i Sverige ett system av standardiserade prov för nationell reglering av betyg. Detta system skall nu ersättas av ett s. k. mål- och kunskapsrelaterat betygssystem med tillhörande betygskriterier. Utvecklingen av det ursprungliga grupprelaterade betygssystemet leddes av Frits Wigforss – en psykolog och matematikdidaktiker som förstod den komplexa problematik som omger varje betygssystem. Han försökte använda nationella prov för att utveckla och stödja, inte ersätta, lärares egna bedömningar. En närmare analys av de standardiserade provens historia i Sverige visar på omfattningen av Wigforss bidrag samtidigt som den visar mätteknikens överordnade ställning i förhållande till matematikämnets innehåll. Den visar också på allvarliga brister när det gäller att uppmärksamma de utbildningsmässiga och sociala konsekvenserna av förändringar i systemet.

Författare

Jeremy Kilpatrick är professor i matematikämnets didaktik vid University of Georgia, Athens, USA.

Bengt Johansson är universitetslektor i matematikämnets didaktik vid Göteborgs universitet, Sverige.

Adresser

Jeremy Kilpatrick, Department of Education, College of Education, University of Georgia, 105 Aderhold Hall, Athens, Georgia 30602, USA.

Bengt Johansson, Matematikavdelningen, Institutionen för ämnesdidaktik, Göteborgs universitet, Box 1010, S-431 26 Mölndal, Sverige.
