Educational Studies in Japan: International Yearbook
No.3, December, 2008, pp.27-40

The Origins of the Standards Movement in the United States: Adoption of the Written Test and its Influence on Class Work

MIYAMOTO, Kenichiro*

Every state in the United States, under the NCLB act, has set state standards and is testing all students in grades 3–8. Students are given printed questions to which they write answers with a pencil on an answer sheet. These written tests are usually given to determine the academic achievements of students.

This paper traces the early history of the written examination and the change in the meaning of "standards" from the middle of the 19th century to the early 20th century. Although the meaning of standards was ambiguous a century ago, by focusing on the written examination for promotion, we can investigate the meaning of standards at different points in time, i.e. what educators expected pupils to achieve at primary school.

It has been stated that (1) the written examination for promotion has forced teachers to accept external standards given by educational administrators; (2) it has also widened the distance between examinations and regular class work, and (3) "standard" had approximately the same meaning as "norm" in the educational measurement movement in the early 20th century.

However, we need to examine the works of educators such as Emerson E. White and Joseph Baldwin, who, in the late 19th and early 20th century, were opposed to written examinations for promotion, and tried connecting written examinations and regular class work. Frank McMurry insisted at the time that standards should be the means for improving the curriculum and instruction at schools.

Educational measurement did not continue for a long time afterward, and was criticized by many educators, especially progressives. The standards movement arose in the 1980s, in the midst of heated controversy surrounding exam-based testing that has lasted from the middle of the 20th century. Yet supporters of the standards movement are still struggling to solve the same problems that White, Baldwin, and McMurry tackled a century ago, most notably, how best to connect instruction and examination.

The answer to this problem lies in unification of the two. Educators now possess various methods and media for connecting examination and instruction. However, the early history of the standards movement implies that educators do not necessarily re-establish interaction and co-operation in class

*Kobe Women's University
e-mail: kmiyamot@suma.kobe-wu.ac.jp

between themselves and pupils as long as academic achievements of children are compared with standards external to teachers. Instruction does not occur without participation of teachers in setting standards and aims of instruction.

1 Introduction

A national survey of academic achievement was conducted in April, 2007 after 46 years' interval in Japan. It was also done in April, 2008, and is supposed to be conducted annually for all pupils of 6th grade and 9th grade. The revival was caused by public anxiety concerning the decline in academic standards of Japanese children. They were said to have lost high rank in international comparisons of academic performance as shown by their scores on PISA in 2000, 2003 and 2006.[1] The Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) revised the national course of study in March, 2008 and is encouraging every school to make efforts to raise academic standards and performance.[2]

The situation is similar to that of the United States since the 1980s. The standards movement, sometimes called "standard-based reform," has been ongoing in the United States.[3] The federal government recently began drawing up policy regarding setting and raising standards.[4] After the publication of *A Nation at Risk* (1983),[5] there appeared a great deal of literature accusing public schools of academic laxity, and calling for a more rigorous academic curriculum and higher academic standards in public schools.[6] There have been three influential documents: *AMERICA 2000: An Education Strategy* announced by the 41st president, George Bush, in 1991; Goals 2000 enacted by the Clinton administration in 1994; and the No Child Left Behind Act of 2001, the legislation reauthorizing the Elementary and Secondary Education Act by the George W. Bush administration.

These three documents contain the common assumptions that higher academic standards are necessary for the nation, and that achievement tests should be used to ascertain the level of academic performance of students and schools. Though the federal government cannot intervene directly in educational matters in individual states, it can have an enormous influence on state education policy by requiring states to implement statewide accountability systems. Under the NCLB Act, each state has set state standards, and must annually test all children in grades 3–8 in reading and mathematics, based on these standards. Not a few educators are trying to set national standards. These standards are not within the jurisdiction of teachers in that they are decreed to teachers from the top down, and educational achievements are tested and connected with rewards and sanctions.[7]

Although the federal government had paid scant attention to educational attainment before the legislation of the Elementary and Secondary Education Act of 1965, American educators have been concerned with the educational standards of public schools since the common school system was established in the 19th century. Teachers have been asking what students should learn in the schoolroom or in the classroom since early times. Educational administrators, who appeared in the late 19th century, have been asking what teachers as well as students should know. In the 20th century, all educators interested in the standards movement were asking how students' and teachers' achievements should be measured, evaluated, and assessed. These are momentous questions for most advocates of the recent standards movement in the United States.

Recent research points out some perils of the standards movement. Thomas Sobol argues that setting higher standards does not necessarily improve educational outcomes without support for teachers and necessary resources.[8] Linda Darling-Hammond emphasizes that higher standards undermine access to education for low-achieving students rather than enhancing it.[9] Kennon M. Sheldon and Bruce J. Biddle maintain that it entails high-stakes tests, coupled with rewards and sanctions, which deprive children of intrinsic interest in subject matter.[10] Patricia Graham points out in her historical studies that teaching standards has become the same as "teaching to the test."[11] Scott Thompson pays attention to the fact that standards-based reforms are sometimes confused with test-based-reforms.[12]

Most research assumes that students' academic achievements can be ascertained by written tests. When we try to assess academic achievements, we usually use paper-and-pencil tests, or written examinations,[13] as is shown in the examinations used in the Programme for International Student Assessment (PISA), National Tests in England, National Achievement Tests in Japan, and the National Assessment of Educational Progress (NAEP) in the U.S.A. Although in recent years other types of examination have appeared, for example, performance-based assessments,[14] written examinations are dominant in most cases. We must be conscious that the written test measures only one aspect of academic achievement, nevertheless it constitutes the most important part in high-stakes testing. We need to know the effects of using written examinations as a way of assessing students' achievement.

This paper traces the early history of the written examination and the change of the meaning of "standards" from the middle of the 19th century to the early 20th century. It will show the emergence of external standards given to teachers, and teachers' responses to external standards. By focusing on the written examination for promotion, it is possible to examine the meaning of the standards, i.e. what educators expect pupils to achieve at primary school.

2 Adoption of the Written Examination

1. Boston School Survey in 1845

In the 19th century, there were two kinds of examinations, oral and written. Oral examinations included recitations, vocal expression, question and answer, or disputation; written examinations were essay-type, problem-type, and some other objective achievement tests.[15] The latter one has spread rapidly since the late 19th century.

Oral examinations were the most prevalent method of ascertaining student achievement before the steel pen and the pencil became commonly used in the late 19th century.[16] Teachers tested students' memory in a recitation to find out whether or not they had achieved the required standard and proficiency. Oral examinations were not impartial or objective in measuring achievement, but teachers could judge ability, based on their observation of pupils in their daily work. Teachers "knew no qualms in those days about reliabilities or validities or comparability." It was the "Age of Innocence."[17]

Instead of oral examinations, written examinations, though not popular in the middle of the 19th century, came into common use in most schools by the end of the 19th century. Horace Mann (1796–1859), State Superintendent of the Public Instruction of Massachusetts, was the first administrator who emphasized written examinations. In 1845, the Boston School Committee conducted a survey on the achievements of all students in Writing and Grammar Schools in order to prove

the effectiveness of public education and to show that Boston's schools required no state supervision. Oral examinations were usually used for examination at that time, but it was difficult to give oral examinations to over 7,000 students. "Written examinations were given, and scored under uniform conditions, and the tabulated results published in detail, question by question and school by school."[18] With all the ineffectiveness revealed by the survey, Boston School Committee published every figure.[19] It might be "the first survey of an American system."[20]

Horace Mann became interested more in the way of inspection than in the results. He criticized oral examinations and emphasized the merits of written examinations "as a potential useful supervisory tool for his office."[21] He maintained,

> When the oral method is adopted, none but those personally present at the examination can have any accurate or valuable idea of appearance of the school…Not so, however, when the examination is by printed questions and written answers. A transcript, a sort of Daguerreotype likeness, as it were, of the state and condition of the pupils' minds, is taken and carried away, for general inspection. Instead of being confined to committees and visitors, it is open to all; instead of perishing with the fleeting breath that gave it life, it remains a permanent record. All who are, or who may afterwards become interested in it, may see it.[22]

Caldwell and Courtis summarized his arguments on the superiority of written examinations over oral examinations as follows:

1. It is impartial.
2. It is just to the pupils.
3. It is more thorough than older forms of examination.
4. It prevents the "officious interference" of the teacher.
5. It "determines, beyond appeal or gainsaying, whether the pupils have been faithfully and completely taught."
6. It takes away "all possibility of favoritism."
7. It makes the information obtained available to all.
8. It enables all to appraise the ease or difficulty of the questions.[23]

The superintendent was seeking impartiality in measuring students' attainments and efficiency in administering examinations. He was anxious that "the different committees (would) have different standards of excellence," and that "different standards of judging (would be) in the minds of different men."[24] Written examinations were almost wholly responsible for stopping teachers from judging pupils' achievements to be arbitrary and bringing about administrative standards. The Boston School Survey was the beginning of objective measurement. It had suggested the *possibility* that standards would become external to teachers. Answer sheets could be read by someone other than teachers, and filed outside of the classroom.

Written examinations did not become widely popular at that time. However, fifty years later, Joseph M. Rice, Edward L. Thorndike, and some other educationists, who advocated science of education, promoted written examinations, changed the way of scoring, and used them for measuring students' academic achievements.[25]

2. Development of Writing and Scoring Tools

Written examinations were not popular until students began using the steel pen instead of the quill pen. The first factory for the mass production of the steel pen was built in New Jersey in 1870. By 1890, students had become used to using the steel pen and to writing answers on paper in written examinations.

Mass production of pencils began in 1866 when Joseph Dixon was granted a patent for his machine. Demand for pencils grew, and more than 20 million pencils were being sold annually by the 1870s. Prices of pencils declined and quality improved in the late 19th century. By the beginning of the 1900s, the pencil was commonly used in the classroom. Writing on paper with a pencil was much easier than scratching on slate surface with a special slate pencil. Students could write rapidly and continue writing for a long time in class.[26]

At the turn of the century, when such writing tools became popular in the classroom and used in examinations, the character of the written examination changed. There existed several types of achievement tests in the 1930s. In addition to oral examinations, they were essay-type examinations, problem-type examinations, and objective achievement tests (new-type).[27] Though ninety percent of actual examinations were still essay-type, the use of the new-type objective achievement tests was rapidly growing.[28] It would not have been possible without the new writing tools. The invention of an electronic-scoring machine by the International Business Machines Corp. in 1936 made new-type examinations much easier to use in many schools.[29]

We must keep in mind that the development of these technical devices brought about written tests.[30] They made it possible for a large number of children to take the same examination at the same time. External examiners, who were not necessarily teachers, could collect students' answer sheets and score them objectively and mechanically. Examiners were thus prohibited from changing marks due to favoritism for certain students.

3 Examinations for Promotion

While grading and promotion gradually spread in city school systems in the 1860s,[31] examinations began to have a new function. They became important means to classify children according to a certain standard of achievement, or "grade." Students who did not reach the standard were not promoted, and were sometimes demoted. High-stakes testing, as it is called today, came into being. By the end of the 19th century, examinations for promotion were commonly in use in elementary and secondary schools. Both students and teachers could not but be conscious of what students should know and could do to pass the examination for promotion.

According to *The Cyclopedia of Education* published in 1877, graded schools were usually defined as "schools in which the pupils are classified according to their progress in scholarship as compared with a course of study divided into grades, pupils of the same or a similar degree of proficiency being placed in the same class," and "promotions should always be based upon a careful examination; and, in a graded school, care should be taken that every grade is passed through in a legitimate manner, that is, without hurry or cramming."[32] (Underline added by the author). At that time, "grade" meant a child's degree of academic proficiency, and "promotion" was decided by an examination. We can think of "grade" as having almost the same meaning as "academic standard," the topic which we are now discussing. Though examinations were not confined to written tests, they were more likely to be used as examinations for promotion.

Until the 1860s, teachers conducted examinations during regular class work. There existed neither examinations for promotion nor grade standards, because schools were ungraded. In the late 19th century, however, as schools were graded, the examination for promotion was introduced and given at the end of each grade. The promotion standard was born with the invention of grading.

Several educators were opposed to examination for promotion. William J. Shearer, Superintendent of Schools in Elizabeth, said, "The teaching test is a necessary part of all true teaching, but the promotion examination prevents broad and intelligent teaching, makes out of the teacher a grind [*sic*], and turns out machine pupils. It is not a good test either of the ability of pupils or of teachers. It is a great temptation to deceit and causes many mental wrecks."[33] He emphasized that teachers should observe students' daily work carefully, and that the teacher's evaluation was a matter of utmost importance for promotion.

The examination for promotion had less relation with daily class work than previous exams. It was quite different in character from that conducted by teachers during daily class work. J. L. Pickard (1824–1914), Superintendent of Public Schools of Chicago, emphasized the distinction between the roles of the superintendent and the teacher.

> A superintendent's examination should be directed rather to the judgment than to the memory of the child. Questions can be so framed as to call for opinions rather than facts. The teacher's examination calls for facts more than for opinions. … The frequent examination by teacher will test memory—the rare examination by the superintendent will measure mental grasp. … The teacher's examination reviews a small portion of text-book work. The superintendent's examination covers a wider field, and fastens upon the child's mind the connection between its parts. … In arithmetic, the teacher asks *how*, the superintendent rather *why*—one calling for solution of examples, the other for statement of principles: the teacher confines himself to illustrative examples; the superintendent more to principles illustrated.[34]

Teachers and superintendents assumed different roles in examining children's academic abilities. The teacher's role was to test children's memory and give them techniques, while the superintendent's role was to develop their minds.

The principal or the superintendent generally made the decision on a pupil's promotion. Pickard pointed out the evils of divergent judgments between principals, and proposed the "unification of work of different principals." He continued,

> Much may be done toward this unification of effort by means of the superintendent's frequent conferences with the principals. But after all the theorizing before an association, he can only determine the practical results by such an inspection of their work as a common examination will afford—an examination of such a character as shall present, in questions prepared, the outcome of his theorizing. A single test may suffice for several years if it may be made to cover the entire ground traversed by the classes presented for promotion. This test will be of incalculable advantage as a means of instruction to teachers.[35]

The superintendent chose promotion standards, and teachers were supposed to accept minor roles in deciding whether or not a child would be promoted. The standards given by the superin-

tendent were external to teachers so far as they did not participate in setting them.

Joseph M. Rice (1857–1934), famous as a popularizer of educational measurement, recognized the significance of teachers' roles, but his words implied distrust of teachers. He said, "A school system must be judged not by what particularly energetic teachers are, of their own accord, willing to do, but what each teacher is required to do in order that she may retain her position."[36] Bureaucracy was appearing in the school system.

In 1912, David Snedden (1868–1951) found the general tendency that "A highly mechanical system tends to introduce external examinations as a basis for promotion and graduation." As examples of external examination, he referred to the "payment by results" plans in England and in American cities during the period from 1870 to 1895.[37]

4 Criticisms of the Written Examination for Promotion

1. Teaching Examination: E. E. White (1829–1902)

When the written examination for promotion came into common use in the 1890s, many educators, mostly superintendents, began to criticize its defects severely at annual conferences of the National Education Association and at other conferences. Some of them conducted a survey on examinations, or wrote books and articles on school management to exemplify their arguments. They pointed out that the written examination was measuring only verbal accuracy; it would tempt pupils to dishonesty; the severe tax of using a pen or pencil on students could lead to physical injury; teacher would drill facts rather than foster ideas.[38]

Being conscious of these defects, Emerson E. White, Superintendent of Cincinnati, Ohio, regarded the distance between the matters dealt with in pupils' daily lessons and examinations for promotion as a serious problem. He made a survey on the situation of the promotion system in leading cities in the U.S. in around 1890. He found that "the use of written examinations as a basis of promotion (had been) well-nigh universal in graded schools."[39] There were two factors which determined pupils' promotion: examination results and teachers' judgment. He gave attention to the fact that "the reported union of these factors (was), in some cases, only nominal, most pupils being promoted on examination results or class record, the other factor being considered only in doubtful cases."[40] White was not dissatisfied with dismissal of teachers' judgment, but he was anxious about the distance between daily class work and the promotion examinations.

He was opposed to promotion examinations and proposed promotion without examinations. He thought that daily evaluations recorded by teachers should be the basis of promotion, and said:

> The special purpose of these recorded estimates is to pass judgment on the fidelity and success of pupils during the period for which they are made. They are not primarily judgments respecting the fitness of pupils for promotion, but they afford an intelligent basis for such a judgment when the time for promotion comes.[41]

For him, "a teacher who does not know the fidelity and success of his pupils in daily work is not competent to teach, and should be retired."[42]

While he criticized the promotion examination, he emphasized the significance of the teaching examination:

> Whatever of testing may be needed to determine the proficiency of individual pupils should be done mainly by the teacher. If he is competent to teach, he is also competent to test the results, and these he must personally know in order to teach…. He must know to what extent his instruction has been appropriated by each pupil in his class.[43]

Superintendent White considered the teaching examination "as essential a part of teaching as instruction itself."[44] Teachers must set grade standards and make teaching examinations for themselves. Teachers must read essays written by pupils. He would not put the management of examinations into the hands of educational administrators. He further stated that, "When the examination of pupils becomes an outside business, the high office of teacher may sink into the trade of preparing pupils for examination."[45]

2. Written Recitation: Joseph Baldwin (1827–1899)

Joseph Baldwin, professor of pedagogy at the University of Texas, was also critical of the promotion examination, and tried to abolish "the periodic examination in our schools as the test of scholarship and as the standard for promotion." He substituted the written recitation for "the dreaded test examination,"[46] namely, the promotion examination.

The written recitation consisted of oral questions asked by a teacher and answers to be written by students during regular class work. This would occur whenever the teacher deemed it helpful. Pupils did not need to prepare for the recitation, but were always ready for written work on their tablets. The teacher read the papers submitted by the pupils at his leisure. The papers were returned to the pupils with the mark S (satisfactory), or U (unsatisfactory). The written recitation was not so much a test examination as regular work in class.

The written examination for promotion had no place in Baldwin's ideology. He argued that promotion should follow satisfactory work, saying:

> It (promotion) does not depend on per-cent marks or on test examinations. Pupils study to know. They become interested in the work and advance with their classes. As the days go by, pupils found able to work more profitably with higher classes are promoted, and pupils who prove unable to work with their classes are demoted. As the terms go by, pupils who do satisfactory work go forward with the class. The teacher knows. There can be no possible excuse for per-cent marking or for test examinations as conditions for promotion or graduation; such devices occasion a world of trouble and do incalculable injury. Good management and good teaching remedy these evils.[47]

Baldwin was especially critical of per-cent marking, as his words above show. It included prizes, honours, test examinations, per-cent record, and per-cent reports. He thought that per-cent marking was a hurtful device, and adamantly insisted that it should be abandoned.

> It (per-cent marking) is a low incentive. It magnifies success at the expense of fidelity. It fosters a brood of school vipers, such as honours, prizes, and hurtful emulations. It abounds in the exact ratio of poor teaching; the poorer the teaching the more perfect the marking. Too often per-cent marking proves an antidote to high thinking and moral teaching.[48]

Baldwin did not dispel the examination or the marking. It is true that the "test examination"

and "per-cent marking" were objects of his hatred, but he was more interested in good teaching. Per-cent marking, which was often thought to be an incentive to study, would only encourage students to aim at irrelevant targets. He believed that good teaching would be sufficient to give children incentives to study.

3. Regular Class Work and Promotion Examinations

White and Baldwin had two propositions in common. One was that the written examination could be a means for improving regular instruction. They had strong interest in improving regular class work, and promotion could not be the aim of instruction. Though White and Baldwin criticized the written test for promotion, they did not deny the significance of the written test used in class. On the contrary, they emphasized the effectiveness of the teaching examination and written recitation in instruction.

The other proposition they had in common was that teachers could more or less decide students' grades and eligibility for promotion. They insisted that promotion should be based on daily class work. Naturally, teachers' judgments were the most important factor in the decision. Promotion standards could not be outside the jurisdiction of teachers.

David Snedden in 1912, referring to the situation of grading and promotion after 1895, observed, "Teachers' judgment of the pupil's ability to proceed enters as a factor, as do also formal records made of a term's work."[49] These words are one piece of evidence to suggest that teachers' judgement and regular class work were important factors for in deciding promotion in the early 20th century. Whether his diagnosis on grading and promotion reflected the reality of the times is uncertain, but it predicted the coming dispute on measurement and standardized testing.

5 The Standards for Efficiency: Educational Measurement

1. Standard as norm

In 1912, Stephen S. Colvin (1869–1923) divided teachers in elementary and secondary schools into two classes, those who liked to mark and those who did not.[50] It was too simplistic a classification. Neither White nor Baldwin disliked marking. However, when written examinations for promotion proliferated and Edward L. Thorndike and some other educational psychologists developed numerous standardized tests in "the age of standard, 1890–1915,"[51] criticisms of examinations and standards also had considerable influence on classroom teachers.

It was inevitable that grading actually caused *retardation* in public schools. Retardation meant at that time the failure to be promoted regularly from grade to grade. The number of retarded students, sometimes called laggards, became an indicator of the efficiency of school systems. The fewer the number of laggards, the more efficient was the school system. L. P. Ayres invented an "Index of Efficiency" based on numbers of promotion and retardation, and made a list of cities arranged in the order of the Index.[52]

In the late 19th century, the age of Social Darwinism,[53] societal reform was always directed toward efficiency. Superintendents, as professional educational administrators, were eager to make the school system more efficient. At that time, it was believed that when the levels of efficiency between school systems were compared in a competitive society, the standards for promotion should be equal in terms of difficulty. Hence, objective measurement and standardized tests were developed.

Advocates of standardized tests did not have confidence in teachers' judgment of students' academic ability. Stephen S. Colvin criticized the marking system of the time for being "unsystematic," as pupils were marked "for the convenience of the teacher and principal." Instead, he expected "men of scientific ideals and scientific training" to create the methods of objective measurement."[54] What most interested him was efficiency of the school system. He asserted that,

> The school system must be efficient, and how can we know that it is, unless we have standards to measure such efficiency? The importance (of this question) … is shown by the prominent place that the question of securing standards to measure educational practices was given at the recent meeting of Superintendents in St. Louis.[55]

His usage of the word "standard" is confusing. He did not discriminate between "standards to measure …efficiency" and "standards to measure educational practices." When he asked for "greater exactness" and "completeness" in measurement, standards were the means for acknowledging efficiency of the school system, not "for the convenience of teacher and principal." Standards were to be made by "men of scientific ideals and scientific training," i.e. "educational experimentalists." He did not refer to what pupils should know or what teachers expected pupils to achieve.

George D. Strayer (1876–1962), professor of educational administration at Teachers College, more positively acknowledged the relation between standards in business administration and measurement of pupils' achievements. He wrote in *A Cyclopedia of Education* edited by Paul Monroe in 1913,

> With the derivation of scales for measuring the achievement of pupils, and with the development of standards in business administration and in organization, it will be possible to test adequately the efficiency of a school or system of schools. When such a system of tests and standards has been derived in all fields, there will be no excuse for the objection to examinations or tests, which one hears so commonly now, because of their inadequacy or unfairness. It is to be expected too that the efficiency of our schools will be greatly increased by the application of such measures.[56]

When we look into the *Encyclopedia of Educational Research* (revised edition, published in 1950), we can find "Standards" in its index list, but we are directed to read the item "norms." "Norm" is a statistical term. By the middle of the 20th century, the word "standard" and "norm" were often used interchangeably in educational measurement.[57] Standards were norms to be referenced to measure efficiency of school systems.

2. Standards in Curriculum

Although neither Colvin nor Strayer paid much attention to standards in the school curriculum, Frank M. McMurry (1857–1929), professor of elementary education at Teachers College, famous as a Herbartian, was extremely interested in curriculum and instruction in relation to standards. He wrote *Elementary School Standards: Instruction—Course of Study—Supervision* (1913) as one of the *School Efficiency Series*, edited by Paul Hanus (1855–1941). McMurry set standards in relation to aims of instruction. He defined a good curriculum as follows,

> The curriculum will be good to the degree in which it contains problems—mental, moral, esthetic, and economic—that are socially vital and yet within the appreciation of the pupils; and its method of presenting that curriculum will be good to the degree in which it exemplifies the methods of solving problems found most effective by the world's most intelligent workers.[58]

He made no mention of promotion examinations, but he discussed the graded course of study in detail. Each grade had positive standards. Promotion occurred naturally after pupils achieved grade standards. He thought standards were valuable "as tests of habits"; "as tests of thoroughness of knowledge"; "as tests of instruction in the three R's"; "as sources of suggestion for improvement"; and "as tests of the curriculum."[59] The merit of standards, he thought, was to fix attention "(1) on *what the children are doing*, and (2) on the value of it as judged by *its relation to the purposes of instruction*."[60] (Italics are in the original).

McMurry thought that standards for judging instruction should include four factors: motive on the part of the pupils; consideration of values by pupils; attention to organization of ideas by pupils; and initiative by pupils. He discussed ways of applying standards to course of study and supervision, and described them in detail. In addition, he gave some recommendations.[61] He held the firm conviction that standards should be the means for improving curriculum and instruction.

McMurry's view on standards was not congruent with the preface attached to his book by the series editor, Paul Hanus. Hanus was dissatisfied that McMurry did not apply scientific measurement in judging the value of instruction. Hanus wrote in the preface,

> Scientific measurement in education is, indeed, as yet too little developed to be applied to more than a very limited portion of the work of the elementary schools. Except for arithmetic and penmanship, 'standard scores' or standard achievements are not available for measuring the quality of the results actually attained by the schools; and even for penmanship and arithmetic, the standard measures for each grade are not yet firmly established.[62]

Hanus recognized that "the methods of exact measurement (were) still in process of evolution," and wrote McMurry's conclusions would be "confirmed by exact measurement, so far as it was possible to apply exact measurement, by Stuart A. Courtis."[63] Hanus did not realize that his usage of the word "standard" was quite different from McMurry's. In McMurry's argument, the standards were not "standard scores," or "standard achievements," but were "sources of suggestion for improvement (of instruction)."[64] Hanus must have thought of Frank McMurry as a giant of the old days, as probably did Colvin and Strayer. It was "the age of standards."

6 Conclusion

We have traced the history of standards and the written examination from the 19th century to the early 20th century. As has been made apparent, the standards movement did not begin with the publication of *A Nation at Risk*. We can find several antecedents in the 19th century.

The written examination, which is now the most dominant means for testing students' academic achievements, was first adopted in the Boston School Survey in 1845. Superintendent Horace Mann found the written examination to be impartial, objective, and free from arbitrary judgments

of teachers. It was the beginning of objective measurement of academic achievements. External standards were ready to materialize.

The common use of the steel pen, pencil, and paper, which occurred in the late 19th century, popularized the written examination. Written examinations were used to test students' achievements in grading and promotion. Hence, external examinations and promotion standards were developed. They were made by educational administrators and teachers were expected to accept them.

Some educators were vehemently opposed to written examinations for promotion. What they feared most was the distance between instruction and examination. They were interested in improving regular class work. They thought examinations for promotion did not contribute to class work and actually hurt pupils. Promotion standards were closely connected with their daily class work.

The period between 1890 and 1915 was the age of standards. Educational measurement, in the name of science of education, increased rapidly. Many school surveys were conducted. Some educational administrators made lists of cities ranked in the order of the "Index of Efficiency." Standards became the means to measure efficiency of education. They had a close relationship to achievement as measured by the standardized test, but lost their relation to the aims of instruction. Standard had almost the same meaning as "norm."

The educational measurement movement did not continue for a long time. It was criticized by many educators, especially progressives. There is numerous research concerning the consequences of the measurement movement.[65] Long after the heated controversy in the middle of the 20th century, the standards movement once again arose in the 1980s. Yet, supporters of the standards movement are still struggling to solve the same problems that White, Baldwin, and McMurry tackled a century ago, most notably, how best to connect instruction and examination. The answer to this question may well lie in unification of the two.

The experience of educators alive a century ago shows us that the written examination created distance between instruction and examination and made standards external to teachers. We must now re-establish the connection that previously existed between them. How can this be possible?

We are now using various media and methods for assessing students' academic achievements. These can be useful means to connect examination and instruction. By using computers, e-mail and other writing, reading and communicating tools, we can easily know students' responses in instruction. Performance-based assessment can show students abilities which paper-and-pencil tests cannot measure.

However, the early history of the standards movement implies that these new media and methods do not necessarily re-establish personal relationship between the teacher and the pupil, as long as academic achievements of children are compared with standards external to teachers. Instruction does not occur without participation of teachers in setting standards and aims of instruction.

References

1 Koji Tanaka ed. *Atarashii Gakuryoku Tesuto wo Yomitoku* (The Interpretation of New Academic Achievement Tests, Tokyo: Nihon Hyojun, 2007)
2 See *The Japanese Journal of Educational Research, Special Issue: Academic Achievement Policies and Schools*, Vol.75, No.2 (June 2008) (*Kyoikugaku Kenkyu* in Japanese).
3 Diane Ravitch, *National Standards in American Education: A Citizen's Guide* (Washington, D.C.: Brookings Institute

Press, 1995); Ravitch, "National Standards: '50 Standards for 50 States' is a Formula for Incoherence and Obfuscation," *Education Week*, Vol.25, Issue 17 (Jan. 5, 2006), pp.54, 56, 58; Susan H. Fuhrman, ed. *From the Capitol to the Classroom: Standard-based Reform in the States*, 100th Yearbook of the NSSE, Part II (2001).

4 Christopher Mazzeo, "Frameworks of State: Assessment Policy in Historical Perspective," *Teachers College Record*, Vol.193, No.3 (June 2001), pp.367–397.
5 The National Commission of Excellence in Education: *A Nation at Risk: The Imperative for Educational Reform* (Washington, D.C.: U. S. Government Printing Office, 1983).
6 See Diane Ravitch, *Left Back: A Century of Battles Over School Reform* (New York: Simon & Schuster, 2000) chap.11.
7 Linda Darling-Hammond, "Standards, Accountability, and School Reform," *Teachers College Record*, Vol.106, No.6 (June 2004), pp.1047–1085.
8 Thomas Sobol, "Beyond Standards: The Rest of the Agenda," *Teachers College Record*, Vol.98, No.4 (Summer 1997), pp.629–636.
9 Linda Darling-Hammond, "Standards, Accountability, and School Reform," *Teachers College Record*, Vol.106, No.6 (June 2004), pp.1047–1085.
10 Kennon M. Sheldon & Bruce J. Biddle, "Standards, Accountability, and School Reform: Perils and Pitfalls," *Teachers College Record*, Vol. 100, No. 1 (Fall 1998), pp.164–180.
11 Patricia A. Graham, *Schooling America: How the Public Schools Meet the Nation's Changing Needs* (Madison, New York: Oxford U.P.) p.186.
12 Scott Thompson, "The Authentic Standards Movement and Its Evil Twin," *Phi Delta Kappan*, Vol.82, No.5 (Jan. 2001), pp.358–362.
13 In this article, the words, "test" and "examination" are interchangeable.
14 Scott J. Cech, "Showing What They Know," *Education Week*, Vol.27, Issue 42 (June 18, 2008), pp.25–27.
15 C. W. Odell, *Traditional Examinations and New Type Tests* (New York: The Century, 1928), Foster Watson & I.L. Kandel, "Examinations," in Paul Monroe ed. *A Cyclopedia of Education*, Vol.2 (New York: Macmillan, 1911) pp.532–536.
16 Adrienne D. Coles, "Pencil Leaves Its Mark," *Lessons of a Century: A Nation's Schools Come of Age* ed. by the staff of Education Week (Maryland: Editorial Projects in Education, 2000), p.160.
17 Max McConn, "Examinations Old and New: Their Uses and Abuses," *The Educational Review* Vol.16 (Oct. 1935), p.375.
18 Otis W. Caldwell & Stuart A. Courtis, *Then and Now in Education, 1845–1923: A Message of Encouragement from the Past to the Present* (New York: World Book Co., 1923), pp.6–7.
19 Chrisopher John Mazzeo, "Examining Examinations: Assessment Policy in Historical Persective, 1865–1935," Ph. D. diss. Stanford University, 2000, p.26.; Caldwell & Courtis, op. cit. p.7.
20 Caldwell & Courtis, op. cit. p.7.
21 Mazzeo, op. cit. p.28.
22 Calewell & Courtis, op. cit. pp.43–44.
23 Ibid. p.37.
24 Ibid. p.43.
25 Ibid. p.8.
26 Robert M. W. Travers, *How Research Has Changed American Schools: A History from 1840 to the Present* (Kalamazoo, Michigan: Mythos Press, 1983), pp.96–98; Adrienne D. Coles, op. cit. p.160; For history of the pencil, see Henry Petroski, *The Pencil: A History of Design and Circumstance* (1989), translated by Watanabe & Okada, *Enpitsu to Ningen* (Shobunsha, 1993).
27 McConn, "Examinations Old and New," p.378.
28 Ibid. p.379.
29 Coles, op. cit. p.161; Mark Walsh, "Quiz Biz," in *Lessons of a Century,* pp.166–172.
30 We can find many examples of new-type tests in C. W. Odell, *Traditional Examinations and New-type Tests*.
31 J. C. Boykin, "Class Intervals in City Public Schools," *Report of the Commissioner of Education for the Year 1890–91*, Vol.2, pp.981–1009.
32 Henry Kiddle, & Alexander J. Schem eds. "Graded Schools," and "School Management," in *The Cyclopedia of Education* (New York: E. Steiger & Co., 1877), p.375, p.766.
33 William J. Shearer, "Elizabeth Plan of Grading," *Addresses and Proceedings of the NEA* (1898), p.443.
34 J. L. Pickard, *School Supervision* (New York: D. Appleton and Co., 1892), pp.97–98.
35 Ibid. pp.105–106.
36 Joseph Mayer Rice, *Public-School System of the United States* (New York: Century, 1893), p.7.
37 David Snedden, "Grading and Promotion," in *A Cyclopedia of Education*, ed. by Paul Monroe, Vol.3 (1912), p.127.
38 Agnes N. Lathe, "Written Examinations—Their Abuse and Their Use," *Education* Vol.9 (1889), reprinted in John Laska & Tina Juarez, eds. *Grading and Marking in American Schools: Two Centuries of Debate* (Springfield, Ill:

Charles C. Thomas, 1992), pp.117–120.
39 Emerson E. White, *Promotion and Examination in Graded Schools, Circular of Information*, U.S. Bureau of Education, No.7 (1891), p.28.
40 Ibid. p.30.
41 Ibid. p.51.
42 Ibid. p.44.
43 Ibid. p.57–58.
44 Ibid. p.58.
45 bid. p.57.
46 Joseph Baldwin, *School Management and School Methods* (New York: D. Appleton and Co., 1907/1897copyright), p.185.
47 Ibid. p.194.
48 Ibid. p.189.
49 Snedden, "Grading and Promotion," p.127.
50 Steven S. Colvin, "Marks and the Marking System as an Incentive to Study," *Education*, Vol.32 (1912), reprinted in Laska & Juales, op. cit. p.85.
51 Max McConn, "Examinations Old and New: Their Use and Abuse," *The Educational Record*, Vol.16 (Oct., 1935), p.381.
52 Leonard P. Ayres, *Laggards in Our Schools: A Study of Retardation and Elimination in City School System* (New York: Charities Publication Committee, 1909), pp.175–184.
53 Richard Hofstadter, *Social Darwinism in American Thought* (Boston: Beacon Press, 1955).
54 Colvin, op. cit. pp.86–87.
55 Ibid. p.87.
56 George Drayton Strayer, "Tests," in Paul Moroe ed. *A Cyclopedia of Education* Vol.5 (1913), p.570.
57 John Martin Rich, "A Philosophical Analysis of Educational Standards," *Educational Theory*, Vol.17, Issue 2 (April, 1967).
58 Frank McMurry, *Elementary School Standards: Instruction—Course of Study—Supervision* (New York: Yonkers-on-Hudson, 1913), pp.3–4.
59 Ibid. pp.13–18.
60 Ibid. p.17.
61 Ibid. pp.5–12ff.
62 Ibid. p.viii. (preface by Paul Hanus).
63 Ibid. p.viii.
64 Ibid. p.17.
65 See Robert Travers, op. cit.