Assessment in historical perspective

Ben Wilbrink

abstract

A historical perspective on assessment is presented, organized around seven themes, and located in the time window between the early middle ages and the end of the 19th century: ways of questioning, the graded school, medieval university examinations, the disputation, ranking and marking systems, examinations and personnel selection, and the possibly facilitating influence of Chinese mandarin examinations on the development of meritocratic assessment. It is concluded that the essential characteristics of current assessment practice were already established by the end of the 19th century, in contrast to the received view on 'educational measurement' being a direct descendant of the innovative work of Galton and Binet.

Assessment in historical perspective

Historical study has the power to illuminate current assessment practice, because that practice is for the greater part traditional practice, and it is a powerful instrument to stimulate reflection on assessment. However, a systematic historical treatment of the subject is not available. Some authors do come close, such as Smallwood (1935) in her study on the history of examinations in the U.S.A., Prahl (1974) in his dissertation on the history of examinations in Western Europe's universities, or Hanson (1993) who, from the perspective of the anthropologist, treats critically the place testing has in modern American society, tracing its roots in witch trials and medieval education. The history of assessment has to be assembled from information hidden in many different monographs, school histories, and studies on this or that aspect or period of assessment practice, as will be clear from the references used in this article. The purpose of this article is to facilitate reflection on critical aspects of current assessment practice by tracing their possible roots in history. The search may uncover some unsuspected facts, as for example the existence of an earlier continental analogue of the Mathematical Tripos, the 18th century Cambridge competitive examination.

The sheer age of some assessment traditions shows them to be relatively immune to changes in cultural environment. Indeed, the university as an institution is one of the oldest of the western world, and university examinations are as old as the universities of Bologna and Paris. The concept
of school organization in forms (the graded school), and so the concept of curriculum, is only two centuries younger. University examinations and the idea of the school form are obvious subjects for historical analysis. While examinations in the 13th century share characteristics with modern examinations, it is not to be assumed they had the same function and meaning to the actors involved as they have in the 20th. Assessment practice must be studied in its historical context, in order to understand how a particular practice was a solution to problems and tasks as perceived by historical actors.

The reverse case is just as interesting, the solutions of the past still being thought valid in current education even though the original problems have long since ceased to exist. It is quite conceivable that our ineradicable habit of ordering and ranking students is such a solution to a problem that no longer exists, or that it is no longer a legitimate solution to an original and still existing problem.

In this article assessment is to be understood in a generic sense, in contrast to the specific approach known as 'educational measurement' which shares its 19th century roots with those of psychological testing. The concept of assessment will be intentionally left undefined, leaving it to the historical analysis to give it shape and content. That it would be misleading to attempt to define assessment on the basis of contemporary practice is illustrated by the fact that in the medieval university the disputation was a prominent part of the examination, with no parallel or analogue in current examinations. We have to settle with 'family resemblances' to illuminate the concept of assessment in its historical and current manifestations.

**What does it mean to know something?**

Medieval education can be characterized as 'teaching' students to learn sacred and other texts by heart. To know something was to know it by heart (Riché, 1989, p. 218). In the early middle ages the texts to be learned were religious texts, and learning took place mostly in monasteries and convents. There was an urgent motivation to learn the Holy Scripture and other religious texts, because doing so made it more likely after one's death to be admitted to heaven. Not only the scarcity of manuscripts forced the monks to learn the scripts by heart; medieval manuscripts being difficult to read one had already to know the text by heart in order to be able to read it (Bolgar, 1954, p. 111). Muslim manuscripts were ambiguous because they consisted only of consonants, therefore the Muslim student had to give proof by recitation that he 'knew' the text. Only then would his master authorize him to teach the text (Berkey, 1992, p. 29).

The medieval monk was confronted with a double task: learning Latin grammar in order to be able to learn Latin texts. Meditation, consisting of the recitation of religious texts, was an important activity for the monk. Holy texts, of course, were written in Latin, so one had to study Latin grammar in order to learn to understand and to speak Latin. The study of grammar consisted in the learning by heart of famous grammars dating from the Roman Empire, or simpler textbooks used for beginners. These grammars were written in the style of questions-and-answers, which was a familiar style in antiquity, see in the Bible the questioning of Adam and Eve by the Lord. Memory could use some support, so many manuscripts had illustrations that served as mnemonics. The 'art of memory' (Yates, 1966) was practised widely, the Jesuit Matteo Ricci even tried to convince the Chinese of its usefulness in preparation for their exams.

Assessment under these circumstances of necessity took the form of having students recite, answer the questions as posed in the grammar that was used, or question each other. The arts examinations at the medieval universities consisted mainly of very simple questions and answers (Lewry, 1982, p. 116). Questioning and answering was the dominant didactic form in teaching and learning. Knowing the right answers to questions about religious texts was extremely important. Out of this kind of questioning grew the catechism, and in its wake the catechetical method. These archetypes of assessment were still dominant in education as late as the 19th century (Foden, 1989, p. 12). Only in the second half of the 19th century did the American Colleges replace the recitation...
method by lectures or 'group discussions.' The recitation method was a combination of learning and examining, but in the American colonies the examining part was in fact non-existent: 'The colonial college student was essentially ungraded and unexamined. (...) public oral examinations were gestures in public relations and therefore not designed to show up student deficiencies' (Rudolph, 1977, p. 145). The first written examinations in Oxbridge in a sense followed the catechetical method, because no questions were put that allowed different interpretations: 'the way to achieve more accurate and certain means of evaluating a student's work was to narrow the range of likely disagreement and carefully define the area of knowledge students were expected to know' (Rothblatt, 1974, p. 292).

It is the experience of almost every living adult in developed countries that even today a substantial part of all questioning and assessment in education is recitation and giving the 'right' answers to known types of questions. Most standardized tests only count the proportion of 'right' answers. The difference between modern and medieval testing seems to be mainly that not the salvation of one's soul but that of one's career depends on producing the right answers. Irony apart, the question-and-answer paradigm deserves the critical looking after it is getting now from many different quarters.

**Joan Cele: 14th century originator of Western style education**

Important principles of curricular and school organization were developed by Joan Cele, rector of the Latin school of Zwolle, a Hanze town in the Low Countries, in the period ca. 1375 to 1415 (Frederiks, 1960; Codina Mir, 1968). Cele, a famous teacher, had to run a school with 800 to 1000 students in a town with only five thousand inhabitants. Many of these students came from Utrecht, Liège, Flanders, and the German countries. Cele solved the organizational problems posed by the sheer number of his students by imposing a new and strict division of the students in eight classes, as well as the curriculum in eight different forms. Cele hired two Parisian masters in the arts to teach philosophy in the highest two classes. However, most of the students were in the lower classes, learning Latin and its grammar. Still being confronted with classes of up to hundred students, Cele introduced a subdivision in groups of ten students called *decuriae*. Each group had a leader who was responsible for learning and discipline; leadership was changed every week. Twice a year Cele held examinations for promotion to a higher form. In the lower forms the exam consisted of a recitation to check on the achievement of the task posed in that form; in the higher forms Cele also looked for the student's insight (*sententia*) into the meaning and message of the Latin texts that were translated.

Cele's innovations were important because his students introduced the new didactic principles in schools all over Europe, among them the university of Paris. The Jesuits, whose *Ratio Studiorum* was inspired by this didactic *modus parisiensis*, definitely established this pedagogy in Europe's schools and universities (Codina Mir, 1968). Joan Cele single-handedly established the European model of the graded school, examinations for promotion, and ranking of students on the basis of merit. The historical importance of Cele's innovation was only recently revealed by the work of Post (1954), Frederiks, and Codina Mir. As late as 1960 Philippe Ariès could still present a meticulous study on the evolution of the graded school, being unaware of the source of the innovations lying in Zwolle in the 14th century.

The medieval class contained pupils of different ages, who were yet in the same form for possibly quite different durations. Current school organization is certainly based on the ideas of Cele, but in the 18th and 19th centuries classes came to be constituted bureaucratically according to age and with a fixed duration of stay, grade retention coming to be a consequence of bureaucratic rules, whereas earlier students used to be promoted on the basis of their learning potential (Paulsen, 1921, p. 621; Ingenkamp, 1972, p. 24, 42). The educational philosophy legitimating this modern school organization after the model of the standing armies of the newly formed states, was already formulated by Comenius in the 17th century (Ingenkamp, 1972, p. 16). This evolution in the
principle of grouping pupils in classes on the basis of age instead of learning is extremely important, for now pupils came to be assessed in comparison to their peers in age, not in comparison to their peers in learning.

**Examinations at the medieval university of Paris**

The university of Paris in the middle ages was an organization of masters, in contrast to the university of Bologna that was an organization of its wealthy students. Most of the Parisian masters, however, were masters of arts, being at the same time students in one of the 'superior' faculties of law or theology. In the medieval university one first had to study grammar and philosophy in the four-year arts curriculum, only 'masters of arts' being admitted to the 'superior' faculties of medicine, law or theology. Nobody could be a student in Paris without having a master, so the first thing the newly arrived student had to do was to seek himself a good master (Thorndike, 1944, p. 30). The master was responsible for his students, he saw to it that they spent their time in study and not in idleness, he set daily exercises and heard their recitations. The master put students in competition to each other giving explicit praise to the student with the best achievement of the day, and blaming the student who blundered worst by giving him the cap with earflaps or asinus. Assessment was part and parcel of the daily life of the medieval student. In the early German universities the 'propaedeutic' arts examination tested students on questions and answers that were extensively practised in the years before. The level of this arts examination was surpassed by that of many schools, such as the schools of Brethren of the Common Life (Schwinges, 1986, p. 336, 356). To help students in the preparation for their exams there were, already in de 13th century, 'examination compendia' available (Lewry, 1982), the same kind of book with questions and answers and model-poems from the civil service examinations that was a selling success in China (Hu, 1984, p. 13).

A major responsibility of the master was to nominate his students for examinations, but only if he deemed them ready. Examinations were public and formal events, failing a candidate was an extremely rare event and even then the reason would be the moral behaviour of the candidate (Schwinges, 1992, p. 235). The candidate was questioned on his knowledge of the prescribed books, he had to deliver a lecture on a text that was only hours before stated to him and he had to take part in a public disputation. The candidate had to give a proof of what the examination would qualify him to do: to lecture.

What really is new and characteristic for the university as a new institution, is the examination by a committee of masters acting on behalf of the representative of the Pope, the chancellor of Paris (see Weijers, 1995, on research on these examinations). 'More particularly, [the universities] were the only institutions—and this was one of the great innovations of the medieval university system—;to link teaching and examinations closely together' (Verger, 1992, p. 43). The successful candidate received the 'licence to teach,' a certificate that enabled the licentiate to teach anywhere in the Christian world and to attract students. Until the rise of the university the authority to teach was self-declared or based on a written statement from one's own master, and the license to teach was temporarily given by the local representative of the church. The genesis of the university examination coincides with the loss, about 1200, of absolute autonomy of the individual master. The individual master became dependent on his examining colleagues: only they could recommend his pupil for the 'licence to teach.' Another way to describe the introduction of the examination is to say that the chancellor of Paris lost his autonomy in the appointment of university teachers because now there had to be an examination of the candidate by a committee composed of masters of the university. Gradually there grew a distinction between the examination and the appointment as a master. Still later, examinations did qualify one for a certain profession, but did not give entry to that profession because an academic grade was only one of many qualifications, descent and wealth being the more important ones (Moraw, 1992). The new institution and its examinations for the first time in Western history defined what knowledge was, thereby also encouraging the new
The university examination was a new institution, having no model in the past, nor in any other country. Webber (1989, p. 36) suggested that the sudden appearance of examinations was influenced by contacts with the Chinese. There are two problems with this hypothesis. At the time, just before Dzenghis Khan established his realm, there were no direct contacts with the Chinese, and Marco Polo went on his journey to China when the examinations were already in place. More important is that the then existing Chinese examinations did not particularly resemble the new university examinations. Another possibility would be that the university examination was copied from practices in higher education in the Muslim world, but there the individual masters were strictly autonomous in licensing their disciples, leading Makdisi (1981) to the conclusion that the organizational forms of the Western universities and their examinations were real innovations.

The methods of lecturing and studying made it necessary to 'hear' the lecture series on a particular book more than once, before one had a reasonably sure knowledge of the text and its commentaries. The regulations of the university stipulated the minimum number of times to hear the lecture series on every book in the examination, making repetition a natural characteristic of education in the universities as well as in the schools, and contributing to the very long duration of studies.

Order of merit in the middle ages was based on one's position in society. The right order was extremely important, even in sitting positions at daily lectures; rich students could buy themselves a place in the 'noble bench.' Also the order of merit at examinations, the *locatus*, was first of all an order of social merit (by birth), and was only in second place determined by criteria such as the length of study (the longer the stay, the higher the place) (Rashdall, 1895 i, p. 459; Schwinges (1986, p. 355; 1992, p. 234). Many students, however, did not even have the intention to go for the arts examination. The conclusion is: yes, there was an honours list for every examination, but the place on the list had little or nothing to do with academic merit.

In the medieval university merit in the modern sense of academic achievement was important in daily practice, but was not explicitly recognized in the examinations in the way of a ranking order. The medieval university examination has always been an important model for examinations ever since, but the competitive examination definitely is a later development.

**The disputation: a lost examination format**

The disputation is the high mark of medieval education. Famous are the disputations between Abelard and William of Champeaux; Abelard describes in his autobiography the flavour of the times and the details of his contests with William (Thorndike, 1944, p. 3). These disputations attracted large numbers of 'students,' and marked the beginnings of what would become the university of Paris. There are, of course, many different forms of disputation, and over the centuries there have been important developments in techniques and traditions. A disputation was a major event: all other activities in the university were cancelled so as to give everybody the opportunity to attend. The pièce de résistance of the disputation was a theorem or problem posed by the master who chaired the disputation. The position of the master was to be defended by one of his students (the *respondens*), and could be opposed by other masters and students. The disputation could last the better part of the day, or even the whole day. The next day the master would give a summary of the arguments pro and contra, and indicate why the opposition failed and what the conclusion or solution (*determinatio*) of the problem should be. For the *respondens* participation in the disputation was part of the fulfilment of his examination requirements.

In rare cases the problem posed was a sincere problem eagerly waiting for a solution; here the disputation was a method of finding new secure knowledge. In the middle ages the disputation was the *only* method to develop new knowledge, and to critically analyze newly translated or discovered theories. In the Muslim world, in the 11th century the disputation was an important instrument in the
development of Muslim law, and for that reason an important method in higher education; Makdisi (1981), in good disputational style, posits the primacy of the Muslim disputational form over that of the later European universities. In the development of logic the disputational method was crucial, as described by Kretzmann & Stump (1988, p. 6).

There is an extensive body of literature on the disputation. Many reports have been preserved in the particular literary form of the report as authorized by the master. McDermott (1993) presents in his anthology of the works of Thomas Aquinas a number of quaeestiones disputatae. In this anthology there is also a lecture of Thomas; in this lecture one can find the same elements as used in the disputation: arguments and counterarguments, conclusions and refutations. In the field of logic a number of disputations and an introduction to the genre are to be found in Kretzmann, Kenny & Pinborg (1982). Lawn (1993) treats the disputation in medicine and science, shows its essential place in the development of science, and gives some examples. References to the literature can be found in Weijers (1987).

Most of the time, however, the disputation was exercises intended to sharpen the wits of the participants, and as such they were related to the didactic form of questions and answers. Little is known about the role of the disputation in the instructional process, 'about how students were taught,' but Perreiah (1984, p. 85) gives details about how ca. 1400 'trial disputation' were delivered: under very strict rules specific to trial disputation, the obligation or the insoluble, and of course under the rules of logic. In the context of trial disputation Perreiah, following Aristotle (Topics 159a 250), speaks explicitly of an instrument to test the knowledge of the participants.

In Jesuit schools the disputation was an instrument to rank students according to merit: the lower ranked student could 'win' the rank of his adversary, and vice versa (Compère, 1985, p. 83). Winning or losing was determined by the number of errors made by each contestant. This was also the practice in the Latin school of Sturm (Codina Mir, 1968, p. 173). This kind of ranking by competitive disputation was also known in late Antiquity (Lim, 1995), and in the Muslim world about 1000 (Makdisi, 1981).

The disputation kept a prominent place in university curricula and examinations until the 18th century, when they became more farcical and finally gave way to modern forms of examination in the 19th century. In Leiden, early in the 17th century, disputation took place once in every two weeks, the very rude discussions regularly resulting in a serious scuffle (Schotel, 1875, p. 332). In Oxbridge, in the 18th century students and faculty no longer took disputation seriously, but only halfway the 19th century they disappeared altogether (Rothblatt, 1974).

The disputation is the only major type of exercise on which one's intellectual agility could be assessed not to survive as such in modern times. The disputation was a public event, and because of that the participants must have been highly motivated to do a good job and to give a good public impression. Assessment in this case was self-assessment as well as assessment by one's public. The disputation has been replaced by examinations in question-and-answer style, yet it might be that there is a modern equivalent to the disputation: scientific research and all the preparation for it that goes into modern secondary and higher education. To be able to do scientific research in the late 20th century demands extensive preparation in mathematics, statistics, discipline-specific research methods, and in the peculiar stylistic scholastics that has developed around reporting and publishing research (e.g. in psychology: Madigan, Johnson, & Linton, 1995). The assessment characteristics are like those of the disputation: reporting is public, and standards for good practice are explicit and objective.

Punishment or reward? Ranking and marking systems

A perennial problem in education is to keep the student's attention on his educational tasks. Punishment is traditionally used to this purpose, often taking the form of punishment for
non-disciplinary behaviour. The heads of medieval schools and universities were entitled to punish their students, even for crimes committed outside the school. For the medieval student punishment was a daily routine. In the 11th century Egbert, a teacher in Liège, criticized the harsh punishment in the schools of his day, and 14th century Joan Caele was known to be mild in his punishments (Fortgens, 1956, p. 36; Frederiks, 1960, p. 56). The humanists propagated competition and reward instead of punishment to motivate students (Bot, 1955). Scaglione (1986, p. 13) sees a connection between the emergence of these new ideas and practices and the innovations of Joan Caele; he also points out that in the Renaissance there was an extraordinary eagerness to learn, in contrast to the periods before and after (o.c., p. 93). The influence of the humanists led to a system of prizes for the best students of the class that dominated Western education until deep in the 19th century.

In order to be able to reward the best student one should know who he is, and one should have some rules to rank students for this purpose. The prize mechanism led to bookkeeping of points or notae throughout the academic (half) year, points being earned by good behaviour, or lost by making academic mistakes as well as by bad behaviour. The prize system is a driving force behind the development of systems of points and 19th century marking systems. The schools of the Brethren in the late middle ages already had an elaborate system of ranking students according to merit, examinations being used to determine their ranking. Students could challenge the rank given to them, in which case a contest between the challenger and the next better ranked student was held (Codina Mir, 1968, p. 173). Haskins (1923, p. 74) gives the example, from a 15th century student manual, of the daily disputation held by the master with his own pupils, where a prize as well as a symbolic punishment (asimus) was given for keeps until the following dispute; the same practice existed in 1559 in Calvin's Academy of Geneva. There was then already a practice of keeping a record of earned points or notae. 'Classes were divided into decuriae not by age or social rank but by merit and achievement. The decurio supervised all work, and punishment for intellectual sluggishness could take the typical form of nota asini "the ass's mark" or nota sermonis soleocismi, "the mark of bad Latin" (Scaglione, 1986, p. 47).' Centuries earlier, in the Muslim world, the same practice existed of ordering of pupils according to merit (Makdishi 1981, p. 81, 91).

In Jesuit schools competition and ranking by academic merit was the core of the educational program. 'The Jesuits, as educators general before modern times, did not formally grade students' homework or even tests, but by their results they listed the students publicly in order of merit (Scaglione, 1986, p. 74).' From the 17th century lists have survived where every student at the end of the school year was graded according to his achievements and capacities. (Compère, 1985, p. 83).

There have always been objections to the prize system. In the middle ages Italian parents objected to the leniency of the system: they preferred punishments. A frequently stated objection was that the many students who would never be able to earn a first or second prize were in fact neglected by this system of rewards. Also there were objections against certain moral problems in the wake of the competition for prizes: fraud, malicious delight, stress, and lying.

In England, during the latter part of the 18th and the first half of the 19th century, the university climate grew competitive, written examinations replacing the orals, and candidates being ranked according to achievement on lists of 'honours' candidates that were made public. Low achieving candidates could hide their shame by taking a 'pass', in which case they were not ranked and their names were not made public. At Cambridge the participants in the Mathematical Tripos were until 1910 ranked according to achievement, the best achievement being honoured with the title of Senior Wrangler, the least one with a title as well as a man-sized attribute: the Wooden Spoon. Competitive examinations in Oxbridge, in the early 19th century, put the students under great pressure (Rothblatt 1982). Competitive examinations were also known on the continent, where already in the 17th century at Leuven there was fierce competition between students from its four colleges, called pedagogies (Vanpaemel, 1986, p. 33). Leuven also had its own variant of Cambridge's classes, called lineas; the best candidate was called the primus and he was highly honoured. The great pressure on students that Rothblatt mentions manifested itself at Leuven already before 1675. The resemblance
between the examinations in Leuven and Cambridge does seem to have escaped the attention of historians.

Exactly why and how ranking systems were in the 19th century replaced with marking systems is not known, but surely the 19th century belief in the power of measurement (Kula, 1986) must have been involved. Ranking of students was in the first half of the 19th century still the dominant practice in secondary education. According to Compère (1985, p. 83), before 1850 there was no marking system in use in France. In early 20th century Germany, class ranking was possibly still in general use: Stern (1920) compared the scores on his new intelligence test with the rank in class, not the marks obtained. In the Netherlands the gymnasium of Groningen was probably the last school to substitute a marking system for the system _notae_ and ranking lists, doing so only in 1901 (Van Herwerden, 1947, p. 41). For the United States the history of grading systems in higher education is described by Smallwood (1935). Notwithstanding the replacement of the believed oldfashioned ranking system with the marking system, high marks were still as scarce a good as the first or second place in the class order of merit, because they were artificially made scarce (Deutsch, 1979, p. 393).

In England the first case of marking examination papers is found in the Mathematical Tripos of 1836: 'Earlier examiners and moderators tended to rely on impression' (Rothblatt, 1982, p. 14).

In the ranking system rank was determined by the summed scores (= notae) of all the students in the form. For that purpose notebooks were kept; in Groningen, for example, every student had a notebook wherein all notae were jotted down, not only those of himself, but also those of all other students (Van Herwerden, 1947, p. 41; Rudolph, 1977, p. 147, for a parallel at Harvard). The notebooks in Western education resemble the _Books of Merit and Demerit_ in China, in the 16th and 17th centuries (Brokaw 1991), but there is probably no link between the two systems. There must be some kind of relation, however, between ranking systems and their replacements, the marking systems that are still used all over the world; knowing that relation might shed some light on the reasons for adopting marking systems.

A short description of the emergence of the marking system in England is given by Rothblatt (1993, p. 44); competitive examinations in Oxbridge demanded objective assessment, and credible objectivity demanded the curriculum to be narrowed so as to be able to assess by using marks. This is an important clue, that marking served purposes of ranking, especially to legitimize the judgments being made of the examination papers, and that curricular content was adjusted to make this kind of assessment possible. In France the marking system seems to have evolved from the ranking system: Chervel (1993, p. 136 ff.) shows how juries for the French _concours d'agrégation_ gradually change a complex ranking procedure into a marking system. Instead of simply ranking the candidates from the worst (number one) to the best achiever (equal to the number of candidates), candidates came to be ranked on a fixed range from one (worst) to ten (best), allowing ties, or breaking ties by using halves. The change was made complete by not using the extreme numbers when the impression was that candidates were not good or bad enough to 'deserve' them. Marking systems differ from country to country, while the basic idea underlying them is the same everywhere in the Western world: the system of ranking stripped of its prizes, and pseudo-objectified by evaluating achievement directly on a marking scale. With hindsight, the problem in the new marking systems is the lack of rules or standards that could make the translation from the number of errors to the assigned grade an objective one.

**Competition and the state**

Modern examinations were formed in the critical period of the late 18th and early 19th century, this formation having much to do with the rise of modern states in Europe. In fact it was state influence that was the crucial factor in most countries, England being a special case because of the autonomous nascence of Oxbridge competitive examinations, and the U.S.A. not yet participating in this process of state formation.
University enrolment in the 17th and 18th century was low and in many countries examinations did not exist any more, or what was called examination was farcical. ‘All through the 18th century, the examination for the B.A. had been a purely formal ritual of answering standard questions known in advance, and reading a "wall lecture," so called because the examiners would generally leave during the reading of the lecture. It was purely a formal requirement that the lecture be read and the examiners were not required to judge its quality’ (Engel 1974, p. 307). ‘For most of the 18th century undergraduates and collegiate fellows were bored’ (Rothblatt, 1974, p. 247). In continental Europe the general trend in the 17th and especially the 18th century was that the state tried to get a hold on the universities and its examinations in order to control the numbers and qualities of its civil servants (Frijhoff 1992). Where earlier one's family, wealth and relations were decisive to get attractive government positions, now merit was becoming the prime criterion. This did not mean that other factors now became unimportant, or that elite positions were threatened by newcomers (Fischer & Lundgreen, 1975). The importance of merit also did not mean that positions were now in fact open to all talented: the costs involved in reaching competitive positions in education were so high that only the established elites and wealthy merchants could bear them, as was the case in the middle ages also (Schwinges 1986, p. 5). Only the 20th century would see the combination of merit and more equal opportunity.

**England**

The development of 'modern' examinations in England begins already in the first half of the 18th century with the institution of the Senate House examination at Cambridge, later to become the Mathematical Tripos (Gascoigne, 1984). The why and how of this development is unknown, but Rothblatt (1974) presents many relevant facts and interesting speculations. Roach (1971, p. 12) affirms the decisive role the English university examinations played as a model for the civil service examinations that were established in the middle of the 19th century. The pervasive influence of the university examinations is described by Rothblatt (1982, p. 15): ‘The Oxbridge model was followed in the schools, in military academies, in the system of local examinations and in the various branches of civil services, excepting the Department of Education and the Foreign Office. Different career phases became linked together by the same examinations (...).

**France**

Present-day France knows the educational contest, the *concours*, for entrance to prestigious institutions and colleges; this tradition has its origin in a legate of Louis Legrand, who started a yearly contest between 10 Parisian colleges in 1747 (Palmer, 1985, p. 24). In the later 18th century more examinations began to be used, and in a more stringent manner, for recruitment to technical institutions for the army (*École du Génie*) and the government (*École des Ponts et Chaussées*), after the revolution the *École Polytechnique*, an institution that was much followed after by other European countries. The whole point of the concours is that admission to a *grande école*, for example, will practically guarantee a prestigious job. In France it was the government that made examinations, for the first time in French history, decisive for many a state career; for this purpose it instituted examinations that did not exist before in this form.

**Prussia**

Prussian rulers in the 18th century built the most efficient bureaucracy of Europe. They instituted the earliest civil service examinations with the intention to break the monopoly of the aristocracy in high government positions (Prahl, 1974, p. 300). In the 18th century a course preparing for government jobs was instituted next to the traditional faculties of theology, law, and medicine. To regulate numbers there came restrictions, also for the other faculties, taking the form of a final examination of the *Gymnasium*: the *Abitur*. In the 19th century students in government tracks and of limited means,
the so-called *Brotstudenten*, were cramming for their state examinations; this group was not sold on to the Humboldtian ideal of the university. Growing numbers of students in the 19th century led to the bureaucratization of state examinations themselves as well, strengthening the natural tendency of *Brotstudenten* to cram for their exams (McClelland, 1980). In these strong developments taking place in the 18th and 19th century the form and function of assessment in Germany was definitely set.

The characteristic development in the 18th and 19th century is that assessment has become a serious matter. No longer was it only a question of honour to win the prize, now one's future career depended on it. No wonder that competitive examinations were going to dominate the educational scene: assessment now served many other lords and interests besides those of transmission of cultural heritage. Assessment served no longer any didactic purposes, instead it dictated them in the form of the necessity of cramming for narrowly defined examinations. Rothblatt (1982) studied the stress that Oxbridge students experienced in their years of study early in the 19th century. From now on for most students only counted what would ultimately be tested. Because so much now depends on the outcome of examinations, the pressure is in the direction of kinds of questions that do not divide assessors, and on procedures of counting errors or assigning marks that give the impression of exactness. Assessors now stand on the side of state interests or of the professional association, no longer on the side of the student like the medieval master did. Merit assessment has its price: an objectifying distance between assessors and assessed. Yet, the same meritocratic procedures, once in place, made it possible in the 20th century to really offer educational and career possibilities to the talented from all classes in modern society, even though in the eyes of some this may have been a mixed blessing (Ringer, 1979, voicing this feeling).

**Chinese mandarin examinations a model for the West**

Imperial Chinese examinations are the first known written examinations in history; they gave entry to civil service and were very selective. They were held once in every few years, in halls dedicated to these examinations. Examinations were thoroughly meritocratic, reflecting the Confucian philosophy on the place of merit in China's hierarchic society. Examinations had different forms and functions in different periods, as documented for the examinations of the Ming and Tsing dynasties, from the 14th century until 1905, by Ho (1962) and Miyazaki (1976). Miyazaki's title, 'China's examination hell,' adequately depicts the character of these examinations. The main characteristics of these examinations are the written form, the elaborate measures taken to ensure objective assessment, their literary content based on the Confucian classics, the possibility of open participation to all, chances of success more often than not only one in a hundred, unlimited opportunities to participate again even in high age, and the frequency of once in every three years. Examinations were the most important opportunity to become a civil servant, a function in very high esteem and well paid. With the exception of the reign of the Khan's, who abolished the examinations, examinations played a crucial role in the stability of the empire, curtailing the power of the aristocracy and the military, and legitimizing the favoured position of civil servants.

Changes in the examination 'culture' in Europe between the early 18th and the late 19th century were manifold, and in the direction of the chief characteristics of Imperial Chinese examinations and bureaucracy: from oral to written examinations, from inconsequential examinations to explicit selection for civil service, from formal ceremonies to competitive examinations, from small numbers to numbers of participants many times higher than the numbers of available places. The resemblance of the European developments during the Enlightenment with the situation in China in the 18th century adored by many intellectuals in Europe suggests some influence of the Chinese model. During the 18th and 19th century many factors influenced the development towards competitive examinations in Europe, among them the achievement of free trade, a principle that also could be of use in government and education. Among the numerous factors mentioned in the literature, the

availability of the model of Chinese civil service examinations deserves special mention. It was widely known in Europe, and examinations modeled after this Chinese examination format were propagated by, for example, Adam Smith in his *Wealth of nations*; see Têng (1943) and Guy (1963) for details on the way the Chinese model influenced European thinking on examinations and their societal role. In their turn, European examinations influenced developments in Japan; its Meiji government instituted meritocratic civil service examinations after the Chinese model, with strong Prussian influence (Spaulding, 1967; Rohlen, 1983, p. 61).

That the Chinese model might have served as a kind of magnet for developments in Western Europe should strengthen our reflective mood regarding the dominant presence of examinations in our daily life. The Chinese civil service examinations were just what the name suggests: a means for selection of civil service personnel, not an educational system. Imperial China never developed an adequate educational system, although in the Sung period a serious effort was made. The suggestion from the Chinese experience is that a strong examination system threatens the quality and even the existence of the educational system. Selection is not a productive process, for it does not of itself produce qualifications; a society that takes the productivity of its educational system seriously should keep education and assessment and selection in proper balance.

**Discussion**

This search for possible roots of assessment, superficial as it of necessity must be, nevertheless shows some significant and maybe quite unsuspected facts. The first observation is that, indeed, before the beginning of the 20th century assessment had already developed into the forms and procedures that still characterize it today. This underscores that our 'assessment culture' is, for the better or the worse, the legacy of societies that long since have gone. Another conclusion from this historical exercise is that the history of 'educational measurement,' going back to Galton and Binet, is surely not the history of assessment. Assessment itself was seen to be a complex concept that could be analyzed in terms of its content, its context of the graded curriculum, its descent from medieval university examinations, its instrumental quality to motivate students, its uses in (societal) selection as an instrument of the state, and as strengthened in its meritocratic character by the example of China's mandarin examinations. Still, some aspects had to be left out, such as how medieval masters and students use their time and what their attitude towards work is (Van den Hoven, 1996), or the period immediately before the rise of the universities (Jaeger, 1994).

This article might give rise to more questions than it answers, in which case it would fulfil its intention to stimulate reflection on assessment. The historical facts in this article concern educational systems primarily serving the upper classes of society, while education in the 20th century is mass education, even extending to mass higher education. Why then would knowledge of the roots of assessment be relevant for understanding current assessment practice? The fascinating observation is that assessment procedures handed down by tradition were in this century rather uncritically adopted in mass education, possibly leading to major inefficiencies in education and for too many students a lack of quality of school life.

---

**Note.** The research for this paper was partly subsidized by the Netherlands Foundation for Educational Research (SVO) in The Hague, grant number 94707.

**References**


Correspondence

Geert Vanpaemel, June 25 1996 (from an email, in Dutch):

I am not aware of any special study on the system of examinations in Leuven. That its competitive system is so unique I did not know. As I have not researched the period before 1650, I do not know where it originates from. It existed already in the beginning of the sixteenth century and has probably been copied from Paris.

There is a paper (licentieverhandeling) on the organisation of the Artes department in the eighteenth century by Cleenewerck de Clayencour, not published, it should be available (typoscript) at Universiteitsarchief, Mgr Ladeuzeplein 21, B-3000 Leuven. The following publications might be relevant:

After being invited to write the article for the StEE I did not have the time to follow up the leads Geert Vanpaemel provided me with, regretfully. May 1997 I obtained van Vocht's (1951) four-volume history of the Collegium Trilingue Lovaniense, that could shed some light on the early sixteenth century situation Geert Vanpaemel hinted at. I have not yet found the opportunity to browse the more than 2000 pages for information on the system of examination used. A relative of a forefather of my grandchildren was Maarten van Dorp, Martinus Dorpius, a professor at Leuven, the college 'De Lelie', who died in 1525. He probably got this position on the basis of excellent examination results. 'De Universiteit te Leuven, 1425-1975' mentions the following on the ranking of students within their college:
"Aan het eind van deze reeks examens werden de studenten gerangschikt—locatio heette dat—volgens de behaalde uitslagen. De rangschikking gebeurde volgens vier categorieën of ordines; eerste rang de rigorosi. tweede de transibilies, derde de gratiosi capaces tamen gratiae, afgewezen werden de gratiosi seu refutabiles. Nummer één in de rangschikking werd tot primus uitgeroepen en in triomf door de stad gevoerd." (p. 90).

[After the close of the series of exams the students were ranked (locatio according to the results obtained. The ranking followed the four categories of ordines; first rank the rigorosi, second the transibilies, third the gratiosi capaces tamen gratiae, while the gratiosi seu refutabiles were failed. The number one in the rank order was acclaimed to be the primus and he was carried through the city in triumph.]
The last sentence suggest that the 'winner' was only one person, not the bunch of four winners from the four colleges. This secondary source is not clear on this point. The relative numbers of students passed and students failed are not known from the sources. So much is known, however, that instruction at the Artes department was directed at the high standards that only the best students were able to comply with. Even with this small amount of information on the examinations in the sixteenth century it is evident that there must have been a fierce competition between students, and many students dropping out of it already at the early stages of their study.


An early, 1555! Brabant student winning the Leuven concours (see de Vocht, mentioned above) is Rogerus (Rutger) Hessels, alias Alardi (Alarts) (Bots, p. 49, 363 #2148):


"De volledige Artes-cursus vergde in de 15e en 16e eeuw te Leuven 2 1/2 jaar, later 2 jaar. De tweejarige cursus was opgebouwd uit 9 maanden logica, 8 maanden fysica, 4 maanden metafysica en ethica, waarna 3 maanden volgden voor het repeteren van de behandelde stof. Viermaanden na het begin van de studie moest de Artiist zijn eerste proeve van bekwaamheid leveren, de actus determinantiae. De disputen voor de titel van baccalaureus artium vonden plaats in het begin van het tweede jaar. De Artes-studie werd afgesloten met het licentiaatsexamen in de vorm van een concours, waaraan alle candidaten deelnamen. Daaraan vooraf ging een examen in elk van de vier pedagogiën, dat gericht was op de voorselectie van de beste studenten, de lineales, die in een tweede ronde met elkaar zouden gaan wedijveren. De drie studenten uit elke pedagogie die bij dit vooronderzoek, calamus geheten, de hoogste ogen gooiden, plaatsten zich voor de eerste linie. Degenen die bij de calamus op de 4e tot 6e plaats eindigden, kwamen terecht in de tweede linie. De nummers 7, 8 en 9 vormden de derde linie. De studenten die zich niet voor een van de linies wisten te klasseren, werden als postlineales aangeduid. De resultaten tijdens het slotconcours bepaalden de plaats binnen de linies alsmede de rangschikking van de postlineales. De beste student uit de eerste linie werd uitgeroepen tot primus universitatis. De primus was het voorwerp van uitbundige festiviteiten, zowel te Leuven als in zijn plaats van herkomst."

Onder de Brabantse studenten zijn de volgende primi:
2148 Rogerus Hessels (details zie box hierboven),
2264 Lambertus Hoex (Houckx) primus 17-11-1726,
3348 Petrus de Louw, Den Bosch, primus nov 1588,
3736 Laurentius Nagelmaeker alias Bacx alias Van Westerhoven primus 18-2-1563,
5546 Lambertus Vincent, Grave, primus 12-11-1648,
5575 Godfriedus van Vlierden, Den Bosch, primus 18-2-1574,
5747 Johannes van den Warck (Waerck), Breda, primus 27-11-1590,
5915 Goswinus Witte, Hilvarenbeek, primus 14-2-1576,
en in Douai: 1485 Simon Fierlands *ca 1602 Den Bosch

Bots p. 46-47
Related issue: if grading really is a form of ranking, then what is a GPA?

This article has articulated how grading really is a form of ranking as it has been developed on the bases of straightforward ranking as practiced in schools throughout Western Europe in modern times. Once in place, a grading system invites one to regard grades as a kind of quantities that may be added, middled, subtracted, etcetera. A GPA seems a natural way to combine grades, no matter how individual grades might be obtained: in what time of year, on what kind of exercise, in which discipline. The next step then is to use GPA in decisions on admissions, tracking, selection. Scores on tests, such as the American Scholastic Achievement Test, may be standardized, yet their fundamental character is that of ranks also. What has changed since the eighteenth or nineteenth century, then, is the complexity of the rankings involved. In former centuries the ranking was based on, for example, the number of errors the students had made in their work. In more recent times, the GPA is a kind of ranking of rankings, possibly quite another kind of procedure than the old way of ranking pupils.

As far as grading may be regarded to be (a form of) ranking, it is possible to precisely analyze its inner workings. This kind of analysis uses the famous Impossibility Theorem by Arrow (1963), a result in social choice theory concerning the impossibility of consistently rankordering a set of choice alternatives. The bite is this: inconsistent ranking of individuals definitely is unfair. The analysis of assessment rankings in terms of Arrow's result is presented in Vassiloglou and French (1982). What is unfair about ranking on the basis of ranks is, for example, that the resulting rank order is dependent on who else is being ranked together with you. The authors present an example from the literature, a ranking of five candidates on the basis of five subrankings, and show that deleting one candidate may change the ranking of the other four relative to each other. Now imagine these candidates applying for admission to Harvard University ..... . The SAT results etcetera might validly indicate merit, yet in the details of the selective process it will be the case that admissions also depend on trivial circumstances.

There are at least two possibilities to escape the unfairness of ranking of rankings: the first is to get rid of the problem itself by replacing simple ranking of performances with assessment of the strengths of performances (French and Vassiloglou, 1986), the second is to inform candidates about the workings of the ranking method that will be used in the coming examination (De Groot, 1970; Van Naerssen, 1970).

- About relative and absolute strength of performance, in an attempt to escape the Impossibility Theorem of Arrow (see Vassiloglou and French, 1982)


- abstract Usually in examinations an overall assessment of a candidate's performance is made by means of a weighted sum of the marks attained on the various components. However, recently it has been suggested that the combination should be based on the candidate's rankings on the components alone, and not on the actual marks. This paper discusses whether such an approach can lead to a fair and consistent system of assessment.

---

**Other, recent, or recently found, publications on the subject**


- o.a. De pedagogiek van de wedijver in historisch perspectief
- Prijsboeken op de Latijnse school, het begin van een traditie
- De gang van zaken bij promotie en prijsuitreiking
- De prijsuitreiking in Europees perspectief en de laatste fase van haar geschiedenis in de Nederlandse context

Heikki Lempa (2006). Patriarchalism and Meritocracy: Evaluating Students in Late Eighteenth©


- "This study probes the ways in which a meritocratic of student evaluation emerged in German educational institutions in the late eighteenth century."
- p. 736: "In fact, the Meritentafel was the final point, the public representation of a complex process of assessment. The system was constantly changed and fine-tuned but, as we know from the first two decades of the institute, it had the following four components. The first step was the collection of data and observations of students. Among the teachers [of the Schnepfenthal Institute], GutsMuths was known for carrying a notebook that he used to record students' achievements in sports. We have more precise information about the second step, the daily assessments that took place during the last hours of the day. For their daily achievements, students received a mark (Marke) or lost one if their performance was lower than expected. For exceptional performance, students could also receive major recognition, a ticket (Billett). The third step was the senate that convened every week after the Sunday service to assess students' weekly performance. Including the whole community the senate awarded a ticket or, as often happened, several tickets if the student passed scrutiny and could present all the marks received over the week. All these tickets
were displayed with a thumbtack placed on the Meritentafel after the name of the student so that the whole community could follow and control the student's development. After collecting fifty thumbtacks, the student was entitled to be named Knight of Diligence (*Orden des Fleisses*).

- Wow, fantastic. Almost the same Meritentafel was used in Dessau some years earlier (see description on p. 737)
- This particular assessment system was highly controversial, however, and was discontinued in Dessau. The controversy was about its intrusion into the traditional domain of the aristocracy.
- This GutsMuths sighhandedly devised a kind of objective assessment system in sports and gymnastics, elaborately described by Lempa. p. 740: "For him ambition was not the privilege of a nobleman but a universal characteristic of a man, any man. Honor had become identical with merit, with the measurability of achievements and outcomes, but it also had become the very definition of masculinity."


- abstract Systems of testing and assessment are shaped in part by personalities and institutions who pursue research insights and technical innovations. Out of these they fashion 'dreams' which drive their efforts to improve these systems. This paper develops this perspective, whilst acknowledging that it overlaps with and complements analyses of assessment systems from social and cultural perspectives. Four different examples are considered. Two from past and current history are the growth, from an origin in IQ testing, of standardised multiple choice tests and the dream of raising standards by external testing. The other two, nascent with their influence yet to be determined, are the dream of improvement by formative assessment and the dream that recent developments in psychology can provide a basis for new and improved assessment practices.
- I do have a copy available, yet?


- Merit ranking as a phenomenon in the medieval university, according to Sullivan, 'is little known and even less considered.' Which is somewhat amazing, because it should be generally known that and how at the University of Leuven the merit order was quite important (see for example Bots, as cited in this webpage)
- In the 'Blanchard affair,' the Parisian chancellor removed from his position in 1386, his abuse of the chancellor's prerogative to depart from the merit order as determined by the masters themselves was one of the grievances held against him. This fact itself indicates that the order of merit was important
- Sullivan uses available archival documents to study the question what typically distinguished the first from the last in each merit order. It is clear that the merit order was not necessarily one according to status or prestige, but was correlated with later academic positions and achievements. Again, no information on whether and how a high merit rank would open doors to particular positions, at least for the secular clerics (for clerics they all were, as the faculty involved is that of theology.
- Sullivan apparently does not have data on typical academic achiements—such as winning disputes—that would result in high placements in the merit order. He ends the article citing Hastings Rashdall (volume 1, p. 481 in the 1936 edition of *The universities in Europe in the Middle Ages*), that:
  "the only questions were whether [the candidate] had duly performed all the residence, exercises, and acts required by the statutes, and whether the reputation he had acquired during his university course for ability, character, and orthodoxy was such as to entitle him to the license."

The Smith's Prize competition was established in Cambridge in 1768 by the will of Robert Smith (1689-1768). By fostering an interest in the study of applied mathematics, the competition contributed towards the success in mathematical physics that was to become the hallmark of Cambridge mathematics during the second half of the nineteenth century. (from the abstract)

The Smith competition was an examination that could be taken shortly after the Mathematical Tripos. The citation above indicates its importance, and its being an instance of an examination having important external effects. While Cambridge is famous for its tripos examination, for mathematicians the Smith prize was more important. This stems from the difference in character between the two examinations, at least after 1883, "the distinction between the capacity for examination work and the capacity to write an original thesis." (p. 301) Before that time the questions set, "unlike Tripos questions, they were often geared towards evincing an original or creative, as opposed to a rote-learning, approach." (p. 282)

Quite remarkable is the number of examinees: typically only a little more than two, sometimes even only two. The number of distinguished positions (prizes) was only two, therefore most of the time only those who would have a reasonable chance to win a prize did participate. Although being much different in character, most of the time the senior wrangler in the Tripos also won the first prize in the Smith competition.


Kleinschmidt solves my problem in defending the particular approach I have chosen in my study.

History and Description. *Measurement and Evaluation in Counseling and Development*, 26, April, 6-10. [I have yet to look this one up, does anybody have a digital copy?]

- **abstract**
  
  ERIC Author/NB Examines proposals to establish national test/testing system as administrative, policy mechanism to foster good teaching and learning, monitor student progress, help in college admission and employment decisions, and monitor progress toward national goals. Notes that history of testing as administrative mechanism suggests that linking of important rewards or sanctions to test performance is key element.


- **abstract**
  
  Over the last 100 years, the ever-increasing demand for testing as a measure of educational reform has created a very profitable market for the US testing industry. We follow the growth of this market since the 1900s in two different, but related, ways. First, we discuss some of the technical developments that have encouraged the use of standardized testing in general and contributed to the growth of the commercial testing industry. Second, we attempt to quantify the expansion of the testing marketplace during the 20th century by tracking several indirect indices of growth over time. We conclude that although technical innovations may have contributed to the growth of the US testing marketplace, they do not necessarily lead to better tests or better outcomes for those who use them. There is a need to more carefully monitor the effects of these tests on teaching and learning in general, particularly when the tests are used in high-stakes contexts.

I have not yet seen the article. Does anyone have a digital copy for me?


- Extracts and annotations in a dedicated page [here](http://www.questia.com/googleScholar.qst?docId=5001256493)
  
  Among other things, Richards writes some important history of the Cambridge Mathematical Tripos, easily the most spectacular European examination in the nineteenth century. An interesting event is the abolition of its extreme ranking procedure, only in 1907. Another point of transgression from ranking to the use of grades.


  Covers the period of development of the competitive tripos examinations, giving many details concerning the examinations and developments over time.


  The early history of a famous Leuven college, not one of the four colleges participating in the competitive Leuven examinations for artes students. Vocht gives many details on the ranking in the Promotion to Master of Arts, of many persons, also in the fifteenth century. He uses E. H. J. Reusens (1869). *Promotions de la Faculté des Arts de l'Université de Louvain, 1428-1797* (1st part, 1428-1568); and the Leuven manuscript *Promotiones in Facultate Artium Universitatis Lovaniensis ab anno 1500 ad annum 1659;* and for example *Extracts from the Sextus Liber Actorum Facultatis Artium* (itself now lost; the first (1427-1441), second (1441-1447) and fifth (1508-1511) have (partly) survived)


- **abstract**
  
Abstract

W. Todd Rogers and Donald A. Klinger (NCME, 2007). Purposes of and Issues with the Provincial Testing Programs in Alberta. pdf

The authors use history to reflect on the purpose et cetera of large scale assessment programs.


De brochure telt 12 bladzijden en is nog te koop. Maak 2 euro over op rekeningnummer 9125297 tnv L.J. Meilink-Hoedemaker"

Inleiding en vraagstelling In de achttiende eeuw was de functie van stadsorganist en klokkenist een begerenswaardige plaats. De musici waren in dienst van de stedelijke en de kerkelijke overheid en het salaris was goed te noemen. Wanneer zo'n plaats vacant kwam nodigden de burgemeesters en de kerkmeesters gezamenlijk de kandidaten uit voor een proefspel. In juli 1741 vonden dergelijke procedures plaats in Delft en Den Haag op twee achtereenvolgende dagen. De examens vonden plaats op drie achtereenvolgende dagen, namelijk op 18 juli 1741 in Delft en op 19 en 20 juli 1741 in Den Haag. Het is niet bekend of bij het vaststellen van deze data overleg is gepleegd.

Van beide gebeurtenissen zijn verslagen bewaard gebleven. Noch in Delft, noch in Den Haag zijn namen bekend van de leden van de sollicitatiecommissie, maar des te meer is bekend over de deelnemers.

In publicaties over deze examens was tot nu toe de aandacht gericht op de procedure en de winnaars van het vergelijkend examen. Op deze webpublicatie komen beide examens opnieuw in de Schijnwerper, maar nu wordt de aandacht gericht op de afgewezen sollicitanten. In totaal dertig kandidaten waren in deze vacatures geïnteresseerd. Ze waren afkomstig uit twintig verschillende woonplaatsen. In Den Haag waren twintig deelnemers en in Delft een en twintig. Elf candidaten traden in beide steden aan. Wat is er van hen terecht gekomen?


abstract Founded in Britain in 1846 to standardize the teaching profession, the College of Preceptors is little known today. The College was closely linked to the Educational Times (hereafter ET), a journal of "Education, Science and Literature " launched in 1847. This paper examines in detail a sample of College examinations, articles on mathematics education, and reviews of mathematics textbooks that appeared in the ET. Key figures in the mathematical discussion were William Whewell, Augustus De Morgan, and Thomas Tate. The paper shows how the discourse on mathematics education led to the introduction of entrance examinations for Oxford and Cambridge Universities.


• Reviewed by Kathryn M. Olesko, Training for the tripos. American Scientist May-June 2004 html
  • "Yet although copies of the Tripos examinations themselves are abundantly available, the answers—the real evidence for knowledge-in-the-making—are less so. Warwick successfully reconstructs the content and more so the experience of working through those answers by triangulating from sources that touch on the learning experience: diaries, correspondence,
autobiographies and textbooks. Likewise, although Warwick has only two sets of coaching notes (Stephen Parkinson's from the 1850s and Robert Webb's from the 1890s), his wide-ranging knowledge of Cambridge culture permits him to extrapolate general observations about the coaching and training experience as a whole from his sources.

- Reviewed by Ivor Grattan-Guinness (2004). The Tripos in a Century of Mathematical Physics at Cambridge University. *SIAM News*, 37, Number 4,
  - "The many illustrations include not only portraits of some Wranglers (famous later or not) but also pertinent lecture notes, research publications, and Tripos examination questions and even candidates' attempts to answer them. This last, particularly welcome, source is one result of the author's extensive use of manuscript collections."

More material on the mathematical tripos in the Wikipedia [html]

In the Elibron Classic series a lot of nineteenth century books on mathematics as well as on the examinations themselves have been republished recently. See [here](#). For example:

  - ISBN 140216131X paperback
  - ISBN 1402128185 hardcover

Lynn Thorndike (1940). Elementary and Secondary Education in the Middle Ages. *Speculum*, 15, 400-408. [pdf][JStor]

  - [abstract](#)
  - possibly the same article, or a revised version, reprinted in *History of Universities*, vol 20; PART 2, 76-130

  - Part of the article treats the subject historically.

---

**links**

Pictura Paedagogica Online, der Digitalen Bildarchiv zur Bildungsgeschichte. "Der Bestand umfasst derzeit mehrere 10.000 Buchillustrationen aus der Zeit vom Mittelalter bis 1850 sowie historische Postkarten des Zeitraums 1870 bis 1933." [site](#)

The Encyclopedia Britannica 1911 on examinations: [http://encyclopedia.jrank.org/EUD_FAT/EXAMINATIONS.html](http://encyclopedia.jrank.org/EUD_FAT/EXAMINATIONS.html)

Francis Galton writes on the mathematical tripos in his *Hereditary genius*, see [html](#) [In fact, the whole book is made available there] Especially watch the fantastic figures in the table on page 19 [here](#).
to be researched

- In a way, the article itself is a small catalogue of theme's deserving further research.

May, 2006. A crucial historical development in assessment is where, next to assessment as an integral part of the instructional process, a kind of functional assessment develops, in connection with the universal licence to teach everywhere. It would be a nice thing to have a thorough understanding of what is happening here, and to connect it with other developments in culture and society, especially the end of feodalism and the rise of more or less autonomous cities. From this moment in history on, assessment is a schizophrenic thing, a beast with two souls, the two souls fighting each other in all possible ways, involving many actors on all levels in society. Functionalism itself is not the problem, remember Charles the Great establishing schools because of the need for clerks able to read and write, to say the least. It is the functionalism of this functionalism—the diploma disease—that introduces stresses in education.


- **first paragraph** "Psychology cannot attain the certainty and exactness of the physical sciences, unless it rests on a foundation of experiment and measurement. A step in this direction could be made by applying a series of mental tests and measurements to a large number of individuals. The results would be of considerable scientific value in discovering the constancy of mental processes, their interdependence, and their variation under different circumstances. Individuals, besides, would find their tests interesting, and, perhaps, useful in regard to training, mode of life or indication of disease. The scientific and practical value of such tests would be much increased should a uniform system be adopted, so that determinations made at different times and places could be compared and combined. With a view to obtaining agreement among those interested, I venture to suggest the following series of tests and measurements, together with methods of making them."

- The site [Classics in the history of psychology](http://www.benwilbrink.nl/publicaties/97AssessmentStEE.html) contains more full texts of key publication in psychological (and educational) measurement. The one by Cattell in a way marks the beginning of psychometrics as an industry.

- Wainer, H. (1987). The first four millennia of mental testing; from ancient China to the computer age. Educational Testing Service Research report 87-34. 6 pages. [also: The Score, 13, 4-5, 11-13, April, 1990.] [ I have seen neither. ETS does not offer an online copy. Does anyone care to send me a pdf copy?]

- May, 2006. It is an intentional omission in the 1997 article to leave out developments in the 20th century, because they are mainly of another kind, connected with the explosion in educational participation, and processes of assessment automation.

An article on these 20th century developments will have to deal with the spectacular development of psychological testing techniques and the way especially American education almost immediately gets infected with these techniques and their accompanying philosophies. Early in the twentieth century the crucial and seemingly irreversible turn taken in assessment is that assessment is a kind of measurement, and it is the better served the less subjective the measurement devices are. This philosophy goes against the grain of extensive experience in education that actors in education are strongly influenced in their behavior by the kind of examinations etcetera that society (politicis) confronts them with.


- Daniel Starch (1916). *Educational measurements*. New York: Macmillan. "If there are any products or by-products of education which are too subtle to be distinguished or judged as existing
in greater or less amounts, or as having higher or lower quality, we may be suspicious of their actual existence. Any quality or ability of human nature that is detectable is also measurable. It remains only to discover more and more accurate means of measurement." This is a spectacularly naive approach to assessment in education, justifying the use of psychometric methods for assessment also. Echo's of this harmful philosophy abound in the educational measurement literature, take for example a linguistically sophisticated one:

- John R. Bormuth (1970). *On the theory of achievement test items*. Chicago: University of Chicago Press. p. 81-2: "It is difficult to overemphasize the importance to instruction of research which attempts to analyze the cognitive processes underlying responses to item types. [maar zie ook het tweede citaat hierbeneden] It is commonly accepted that, in many subject-matter areas, the learning of the knowledge explicitly taught by the instructional programs is less valued as a learning outcome than learning the complex cognitive processes by which that and other knowledge is discovered, evaluated, organized, and applied. Achievement test items which can test these complex processes are useful not only for evaluating the student's achievement and the effectiveness of his instruction, but also for providing the instructional exercises which force him to practice those processes. But few of these benefits can be reaped until we identify exactly what it is that the different classes of items test."

- The point to make now is the following. Early in the 20th century the testing virus escapes from the psychological laboratories and infects the educational establishment. Philosophy and technology of testing are wholeheartedly taken in by the educational community. Large investments, many people and institutions are involved. What is happening is a lock-in on an enormous scale: educational assessment has become educational measurement using standardized achievement tests, or at least teacher-made tests resembling such standardized tests. What is a *lock-in*? The QWERTY-keyboard is an example of a lock-in on a particular keyboard layout, an early pragmatic layout that later on proves to be the final one because all attempts to introduce better layouts fail. The VHS video technology is another example of an inferior technology winning the market against the superior Video 2000 technology developed by Philips (later Philips developed cd-laser-technology, earning huge profits on the patent).

- A corollary of the lock-in hypothesis concerning educational measurement is the following. In the course of the 20th century there has been a flurry of research on educational testing, resulting in more efficient testing formats, more refined standards of (achievement) testing, highly sophisticated statistical techniques used in developing and evaluating tests, online testing technology, etcetera. Progressively better techniques and tests have become available, while the funding philosophies of this educational measurement, if they ever have been there in a form more elaborated than the passages quoted above, have stayed as inappropriate to the educational process as ever. It is highly exceptional for educational measurement specialists to rise to the issue, and propose new approaches recognizing the way assessments in education truly function. Names: Popham (US), Van Naerssen (Netherlands, [html](http://www.benwilbrink.nl)). Themes: backlash, feedforward. Research: H. Becker, B. Geer and E. C. Hughes (1968). *Making the grade: the academic side of college life*. New York: Wiley. James S. Coleman (1990). *Foundations of social theory*. Cambridge, Massachusetts. James S. Coleman (1-6-1994) (concept) What goes on in school: a student's perspective. paper 3/25/94; rev. 6/1/94. See numerous entries, articles etc. on this website [www.benwilbrink.nl](http://www.benwilbrink.nl)