ASSESSMENT IN HISTORICAL PERSPECTIVE

Ben Wilbrink

*SCO-Kohnstamm Institute for Educational Research,
University of Amsterdam, The Netherlands[1]*

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 USA, 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 unexpected 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 as they have in the 20th to the actors involved. 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 for "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 these texts by heart; as medieval manuscripts were difficult to read one had already to know the text by heart in order to be able to read it (Bolgar, 1954, p. 111). Arabic manuscripts were ambiguous because they consisted only of consonants, so that the student of these texts 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 of the learning by heart of famous grammars dating from the Roman Empire, or simpler textbooks aimed at beginners. These grammars were written in the style of questions-and-answers, a familiar style in antiquity, of which God's questioning of Adam and Eve in the Bible is a very early example. Memory could use some support, so many manuscripts had illustrations that had a mnemonic function. 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 exigencies took the form of having students recite, answer questions as posed in the particular grammar used, or question each other. The arts

examinations at the medieval universities consisted mainly of very simple questions and answers (Lewry, 1982, p. 116). Question and answer 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 study and examination, but in the American colonies the examination 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 Oxford and Cambridge 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).

Even today a substantial part of all questioning and assessment in education consists of recitation and giving the "right" answers to known types of questions. Most standardized tests only count the proportion of correct answers. The difference between modern and medieval testing seems to be mainly that not the salvation of one's soul but rather one's career now depends on producing the right answers. Irony aside, however, the question-and-answer paradigm deserves the critical consideration 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 Hanseatic town in the Low Countries, between approximately 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 into eight classes, as well as a curriculum divided into 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 into 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 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 he translated.

Cele's innovations were important because his students introduced these 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*, established this pedagogy in Europe's schools and universities (Codina Mir, 1968). Joan Cele single-handedly created 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, without being aware of the source of the innovations in 14th century Zwolle.

The medieval class contained pupils of different ages, who were 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 came to be a consequence of this bureaucratic approach: earlier students used to be promoted on the basis of their learning potential (Ingenkamp, 1972, pp. 24, 42; Paulsen, 1921, p. 621). The educational philosophy legitimating this organization of the modern school after the model of the 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 and not in comparison to their peers in learning.

Examinations at the Medieval University of Paris

The medieval university of Paris was an organization of masters, in contrast to the university of Bologna, which was an organization of wealthy students. Most of the Parisian masters, however, were masters of arts, who were at the same time students in one of the "superior" faculties of law or theology. In the medieval university students first had to study grammar and philosophy in the four-year arts curriculum since only *masters of arts* were admitted to the superior faculties of medicine, law or theology. Nobody could be a student in Paris without the supervision of 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 studying and not in idleness, he set daily exercises and heard their recitations. The master also made his students compete with each other, explicitly praising 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 preceding years. The level of this arts examination was surpassed by that of many schools, such as the schools of the Brethren of the Common Life (Schwinges, 1986, pp. 336, 356). To help students prepare for their exams, already in the 13th century, *examination compendia* were available (Lewry, 1982), the same type of book with questions and answers and model-poems for 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 when he deemed them ready. Examinations were public and formal events. Failing a candidate was an extremely rare event; 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 at only a few hours' notice, and he had to take part in a public disputation. The candidate had to give proof of what the examination would qualify him to do: to lecture.

What really was innovative and characteristic of the universities, as new institutions, was the examination by a committee of masters acting on behalf of the representative of the Pope, the chancellor of Paris (see Weijers, 1995). "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 licensee to teach anywhere in the Christian world and to get his own 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, in approximately 1200, of the individual master's absolute autonomy. 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 with its examination practices for the first time in Western history defined knowledge, thereby also encouraging the new phenomenon of professionalization (Bullough, 1978).

The university examination was a new institution, without antecedents in the past or elsewhere. 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 Genghis Khan established his realm, there were no direct contacts with the Chinese. Marco Polo, on the other hand, went on his journey to China when examinations were already being conducted in Europe. More importantly, the Chinese examinations in question did not particularly resemble the new medieval university examinations. Another possibility would be that the idea of the university examination was copied from practices in higher education in the Muslim world. There, however, 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 for the student to hear the lecture series on a particular book more than once, before he had a reasonably reliable command 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 social status. Maintaining the social order was extremely important at the universities, and this determined even apparently insignificant things like seating arrangements at daily lectures; rich students could buy themselves a place in the *noble bench*. The order of merit at examinations, the *locatus*, too, was first of all an order of social merit (by birth), and was only in second place determined by criteria such as length of study (the longer the stay, the higher the place) (Rashdall, 1895, p. 459; Schwinges, 1986, p. 355; 1992, p. 234).We may conclude that there was an honours list for every examination but placement 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 was 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. Especially famous are the disputations between Abelard and William of Champeaux; Abelard's autobiography renders 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 genuine problem eagerly awaiting solution; here the disputation served as a method of finding new reliable knowledge. In the Middle Ages the disputation was the *only* method to develop new knowledge, and to critically analyze newly discovered or translated 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 and Stump (1988, p. 6).

There is an extensive body of literature on the disputation. Many reports of disputations have been preserved in the particular literary form of the report as authorized

by the master. McDermott (1993), in his anthology of the works of Thomas Aquinas, presents a number of *quaestiones disputatae*. This anthology also includes a lecture by Thomas in which one can find the same elements as used in the disputation: arguments and counterarguments, conclusions and refutations. A number of disputations in the field of logic and an introduction to the genre can be found in Kretzmann, Kenny, and Pinborg (1982). Lawn (1993), treating 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 disputations were 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 in the early 15th century *trial disputations* were delivered: under very strict and specific rules and of course under the rules of logic. In reference to the 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 used to rank students according to merit: the lower-ranked student could 'win' the rank of his adversary, and vice versa (Compère, 1985, p. 83). Victory or defeat 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 disputations 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 it gradually began giving way to modern forms of examination in the 19th century. In early 17th century Leiden disputations took place once every two weeks, and they frequently ended in a serious scuffle (Schotel, 1875, p. 332). By the 18th century, students and faculty in Oxford and Cambridge no longer took disputations seriously, but they disappeared altogether only in the mid 19th century (Rothblatt, 1974) .

The disputation is the only major type of exercise assessing students' intellectual agility not to survive as such in modern times. The disputation was a public event, and because of this the participants must have been highly motivated to do a good job and make 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 scientific research and all the preparation that goes into it in the form of current secondary and higher education may well be its latter-day equivalent. To be able to do scientific research in the late 20th century demands extensive preparation in statistics, discipline-specific research methods, and in the peculiar stylistic scholastics that has developed around reporting and publishing research (e.g., for 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 the educational tasks at hand. Punishment is traditionally used to this purpose, often taking the form of punishment for undisciplined 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 Cele 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) perceives a connection between the emergence of these new ideas and practices and the innovations of Joan Cele; he also points out that in the Renaissance there was an extraordinary eagerness to learn, in contrast to the periods before and after (Scaglione, 1986, p. 93). The influence of the humanists led to a system of prizes awarded to the best students of the class that dominated Western education far into the 19th century.

In order to be able to reward the best student she or he must be identified by means of some rules designed for this purpose. The prize mechanism led to a bookkeeping system of points or *notae* which was kept 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 was 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 highest ranking student was held (Codina Mir, 1968, p. 173). Haskins (1923, p. 74) quotes from a 15th century student manual an example of the daily disputation held by the master with his own pupils, where a prize as well as a symbolic punishment (*asinus*) were given until the following dispute. The same practice existed in 1559 in Calvin's Academy in Geneva, which already kept 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 soloecismi*, "the mark of bad Latin" (Scaglione, 1986, p. 47). Centuries earlier, in the Muslim world, the same practice existed of ranking of pupils according to merit (Makdisi 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). Lists have survived from the 17th century on which every student was graded according to his achievements and capacities at the end of the school year (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 frequent 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 that arose as the result 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 centuries, 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 lowest with a title as well as a man-sized attribute: the Wooden Spoon. Competitive examinations in Oxford and Cambridge, 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 seems to have escaped the attention of historians.

Exactly why and how ranking systems were replaced in the 19th century 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 of *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 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 (i.e., *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 Oxford and Cambridge 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 between 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 USA not yet participating in this process of state formation. University enrolment in the 17th and 18th century was low and in many countries examinations either did not exist any more, or had become 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" adds Rothblatt (1974, p. 247). In continental Europe the general trend in the 17th and especially the 18th century was for the state to try to get a hold on the universities and its examinations in order to control the numbers and quality of its civil servants (Frijhoff, 1992). Where earlier one's family, wealth and relations were decisive in getting 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). Nor did importance of merit mean that positions

were now in fact open to the talented: the costs involved in reaching competitive positions in education were so high that only the established elites and wealthy merchants could shoulder them, as had been the case in the Middle Ages too (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 as early as 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). How this development came about 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: "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" (1982, p. 15).

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 ten 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 (*Ecole du Génie*) and the government (*Ecole des Ponts et Chaussées*); after the revolution the *Ecole Polytechnique* was established, an institution that set a major example for 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 civil service career; for this purpose it instituted examinations that did not previously exist 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 alongside to the traditional faculties of theology, law, and medicine. To regulate numbers restrictions were introduced, also for the other faculties, in the form of a final examination at the *Gymnasium*: the *Abitur*. In the 19th century students of limited means in government tracks, 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). During these major developments taking place in the 18th and 19th century the form and function of assessment in Germany was definitely set.

The characteristic development of the period is that assessment became a serious matter. No longer was it only a matter 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 the 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 what counted was only what they would eventually be tested on.

Because so much now depended on the outcome of examinations, the pressure was to design the type of questions that would not divide assessors, and on procedures of counting errors or assigning marks that would give the impression of exactness. Assessors now side with state interests or the professional association, no longer with the student, as the medieval master had done. Merit assessment has its price, i.e., an objectifying distance between assessors and assessed. Yet, the same meritocratic procedures, once in place, made it possible in the 20th century to 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).

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 every few years in halls especially 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). The title of Miyazaki's book, *China's Examination Hell*, adequately depicts the character of these examinations. The main characteristics of these examinations are that they were in written form and held one every three years; elaborate measures were taken to ensure objective assessment; their literary content was based on the Confucian classics; participation was open to all, with chances of success only one in a hundred but unlimited opportunities to resit. Examinations were the main way of becoming a civil servant, a well paid and highly esteemed function. With the exception of the reign of the Khans, 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 events, from small numbers of participants to a number of participants many times higher than the number of available places. This resemblance together with the period's intellectual fashion, in Europe, for things Chinese, suggests some influence of the Chinese model. Many other 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. But the availability of the model of Chinese civil service examinations deserves special mention. It was widely known in Europe, and examinations modeled after the 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 a 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 words suggest: a means for the 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, assessment and selection in proper balance.

Discussion

This search for possible roots of assessment, superficial as of necessity it must be, nevertheless shows some significant and maybe quite unsuspected facts. The first 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 better or for worse, the legacy of societies long since 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 have had to be left out, such as how medieval masters and students used their time and what their attitude towards work was (Van den Hoven, 1996), or a closer look at 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? It is fascinating to observe 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 in quality of school life.

Note

References

Ariès, Ph. (1960). *L'enfant et la vie familiale sous l'ancien régime* [Child and family life under the Ancient Regime]. Paris: Plon.

Berkey, J. (1992). *The transmission of knowledge in medieval Cairo. A social history of Islamic education.* Princeton: Princeton University Press.

Bolgar, R.R. (1954). *The classical heritage and its beneficiaries.* Cambridge: Cambridge University Press.

Bot, P.N.M. (1955). *Humanisme en onderwijs in Nederland* [Humanism and education in the Netherlands]. Utrecht: Het Spectrum.

Brokaw, C.J. (1991). *The ledgers of merit and demerit. Social change and moral order in late imperial China.* Princeton: Princeton University Press.

Bullough, V.L. (1978). Achievement, professionalization, and the university. In J. IJsewijn, & J. Paquet (Eds.), *The universities in the late Middle Ages* (pp. 497-510). Leuven: Leuven University Press.

Chervel, A. (1993). *Histoire de l'agrégation. Contribution à l'histoire de la culture scolaire* [History of the 'agrégation' exam]. Paris: INRP Editions Kime.

Codina Mir, G. (1968). *Aux sources de la pédagogie des Jésuites; le 'Modus Parisiensis'* [Origins of Jesuit pedagogy]. Roma: Institutum Historicum S.I.

Compère, M.M. (1985). *Du collège au lycée (1500-1850). Généalogie de l'enseignement secondaire français* [Genealogy of French secondary education]. Paris: Gallimard/Julliard.

Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. *American Psychologist, 34,* 379-401.

Engel, A. (1974). Emerging concepts of the academic profession at Oxford 1800-1854. In L. Stone (Ed.), *The university in society. Vol I: Oxford and Cambridge from the 14th to the early 19th century* (pp. 305-351). Princeton: Princeton University Press.

Fischer, W., & Lundgreen, P. (1975). The recruitment of administrative personnel. In Tilly, C. (Ed.), *The formation of national states in western Europe* (pp. 456-561). Princeton: Princeton University Press.

Foden, F. (1989). *The examiner. James Booth and the origins of common examinations*. Leeds: School of Continuing Education.

Fortgens, H.W. (1956). *Meesters, scholieren en grammatica; Uit het middeleeuwse schoolleven* [Medieval schools]. Zwolle: Tjeenk Willink.

Frederiks, J. (1960). *Ontstaan en ontwikkeling van het Zwolse schoolwezen tot omstreeks 1700. Een historische studie* [Origins and development of schools in Zwolle until 1700]. Zwolle: Tijl.

Frijhoff, W. (1992). Universities: 1500-1900. In B.R. Clark & G.R. Neave (Eds.), *The encyclopedia of higher education* (II, pp. 1251-1259). Oxford: Pergamon.

Gascoigne, J. (1984). Mathematics and meritocracy: The emergence of the Cambridge Mathematical Tripos. *Social Studies of Science, 14*, 547-584.

Guy, B. (1963). The Chinese examination system and France, 1569-1847. In T. Besterman, *Studies on Voltaire and the eighteenth century*, Vol. 25, 741-778. Geneva: Institut et Musée Voltaire.

Hanson, F.A. (1993). *Testing testing. Social consequences of the examined life*. Berkeley: University of California Press.

Haskins, Ch. H. (1923/1957). *The rise of the universities*. London: Cornell University Press.

Ho, P.T. (1962). *The ladder of success in imperial China. Aspects of social mobility, 1368-1911*. New York: Columbia University Press.

Hu, C.T. (1984) The historical background: Examinations and control in pre-modern China. *Comparative Education, 20*, 7-26.

Ingenkamp, K. (1972). *Zur Problematik der Jahrgangsklasse* [Problems of the graded class]. Weinheim: Beltz

Jaeger, C.S. (1994). *The envy of angels. Cathedral schools and social ideals in Medieval Europe, 950-1200*. Philadelphia: University of Pennsylvania Press.

Kretzmann, N., & Stump, E. (Eds.) (1988). *The Cambridge translations of medieval philosophical texts. Volume one: Logic and the philosophy of language.* Cambridge: Cambridge University Press.

Kretzmann, N., Kenny, A., & Pinborg, J. (Eds.) (1982). *The Cambridge later Medieval philosophy*. Cambridge: Cambridge University Press.

Kula, W. (1986). *Measures and men*. Princeton: Princeton University Press.

Lawn, B. (1993). *The rise & decline of the scholastic 'quaestio disputata' with special emphasis on its use in the teaching of medicine and science*. Leiden: Brill.

Lewry, O. (1982) Thirteenth-century examination compendia from the faculty of arts. In *Les genres littéraires dans les sources théologiques et philosophiques médiévales* (pp. 101-116). Louvain-la-Neuve: Institut d'Études Médiévales, Université Catholique de Louvain.

Lim, R. (1995). *Public disputation, power, and social order in late antiquity*. Berkeley: University of California Press.

Madigan, R., Johnson, S., & Linton, P. (1995). The language of psychology: APA style as epistemology. *American Psychologist, 50*, 428-436.

Makdisi, G. (1981). *The rise of colleges: Institutions of learning in Islam and the West*. Edinburgh: Edinburgh University Press.

McClelland, Ch. E. (1980). *State, society, and universities in Germany 1700-1914*. Cambridge: Cambridge University Press.

McDermott, T. (Ed.) (1993). *Thomas Aquinas, Selected philosophical writings*. Oxford: Oxford University Press.

Miyazaki, I. (1976). *China's examination hell*. New York: Weatherhill.

Moraw, P. (1992). Careers of graduates. In H. de Ridder-Symoens (Ed.), *A history of the university of Europe. Volume I, Universities in the Middle Ages* (pp. 244-279). Cambridge: Cambridge University Press.

Palmer, R.R. (1985). *The improvement of humanity. Education and the French revolution*. Princeton: Princeton University Press.

Paulsen, F. (1921/1960). *Geschichte des gelehrten Unterrichts auf den deutschen Schulen und Universitäten vom Ausgang des Mittelalters bis zur Gegenwart* [History of academic teaching in German schools and universities, from end Middle Ages to the present]. Vol. II. Berlin: de Gruyter.

Perreiah, A.R. (1984). Logic examinations in Padua circa 1400. *History of Education, 13*, 85-103.

Post, R.R. (1954). *Scholen en onderwijs in Nederland gedurende de middeleeuwen* [Schools and education in the medieval Netherlands]. Utrecht: Het Spectrum.

Prahl, H.W. (1974). *Abschlussprüfungen und Graden. Sozialhistorische und ideologiekritische Untersuchungen zur akademischen Initiationskultur* [Social history of university examinations]. Unpublished dissertation. Kiel University.

Rashdall, H. (1895). *The universities of Europe in the Middle Ages.* Edited by F.M. Powicke & A.B. Embden (1936) Oxford: Clarendon.

Riché, P. (1989). *Ecoles et enseignement dans le Haut Moyen Age Fin du Ve siècle - milieu du XIe siècle* [Schools and teaching in the High Middle Ages]. Paris: Picard.

Ringer, F. (1979). *Education and society in modern Europe.* Bloomington: Indiana University Press.

Roach, J. (1971). *Public examinations in England 1850-1900.* Cambridge: Cambridge University Press.

Rohlen, T. P. (1983). *Japan's high schools.* Berkeley: University of California Press.

Rothblatt, S. (1974). The student sub-culture and the examination system in early 19th century Oxbridge. In L. Stone (Ed.), *The university in society. Vol. 1 Oxford and Cambridge from the 14th to the early 19th century* (pp. 247-303). Princeton: Princeton University Press.

Rothblatt, S. (1982). Failure in early nineteenth century Oxford and Cambridge. *History of Education, 11,* 1-21.

Rothblatt, S. (1993). The limbs of Osiris: Liberal education in the English-speaking world. In S. Rothblatt & B. Wittrock (Eds.), *The European and American university since 1800. Historical and sociological essays* (pp. 19-73). Cambridge: Cambridge University Press.

Rudolph, F. (1977). *Curriculum. A history of the American undergraduate course of study since 1636.* San Francisco: Jossey Bass.

Scaglione, A. D. (1986) *The liberal arts and the Jesuit college system.* Amsterdam: Benjamins.

Schotel, G.D.J. (1875) *De academie te Leiden in de 16e, 17e en 18e eeuw* [The Leiden Academy in the 16th, 17th and 18th centuries]. Haarlem: Kruseman & Tjeenk Willink.

Schwinges, R.C. (1986). *Deutsche Universitätsbesucher im 14. und 15. Jahrhundert: Studien zur Sozialgeschichte des alten Reiches* [University students in 14th and 15th century Germany]. Stuttgart: Steiner.

Schwinges, R.C. (1992). Student education, student life. In H. De Ridder-Symoens (Ed.), *A history of the university of Europe. Volume 1, Universities in the Middle Ages* (pp. 195-243). Cambridge: Cambridge University Press.

Smallwood, M.L. (1935). *An historical study of examinations and grading systems in early American universities.* Cambridge: Harvard University Press.

Spaulding, R.M. (1967). *Imperial Japan's higher civil service examinations*. Princeton: Princeton University Press.

Stern, W. (1920). *Die Intelligenz der Kinder und Jugendlichen und die Methoden ihrer Untersuchung* [Children's intelligence and ways of testing it]. Leipzig: Barth.

Têng, S. (1943). Chinese influence on the Western examination system. *Harvard Journal of Asiatic Studies, 7,* 267-312.

Thorndike, L. (1944). *University records and life in the Middle Ages*. New York: Columbia University Press.

Van den Hoven, B. (1996). *Work in ancient and medieval thought*. Amsterdam: Gieben.

Van Herwerden, P.J. (1947). *Gedenkboek van het Stedelijk Gymnasium te Groningen* [Memorial publication of the Groningen Gymnasium]. Groningen: Wolters.

Vanpaemel, G. (1986). *Echo's van een wetenschappelijke revolutie. De mechanistische natuurwetenschap aan de Leuvense Artesfaculteit (1650-1797)* [Echoes from a scientific revolution]. Verhandelingen van de Koninklijke Academie voor Wetenschappen, Letteren en Schone Kunsten van Belgie, Klasse der Wetenschappen, Jaargang 48, Nr. 173. Brussel: Paleis der Academien.

Verger, J. (1992). Patterns. In H. De Ridder-Symoens (Ed.), *A history of the university of Europe. Volume I, Universities in the Middle Ages* (pp. 35-74). Cambridge: Cambridge University Press.

Webber, C. (1989). The mandarin mentality: Civil service and university admissions testing in Europe and Asia. In R. Gifford (Ed.), *Test policy and the politics of opportunity allocation: The workplace and the law* (pp. 33-60). Dordrecht: Kluwer.

Weijers, O. (1987). *Terminologie des universités au XIIIe siècle* [Terminology of 13th century universities]. Roma: Edizione dell' Ateneo.

Weijers, O. (1995). Les règles d'examen dans les universités médiévales. In M.J.F.M. Hoenen, J.H.J. Schneider, & G. Wieland, (Eds.), *Philosophy and learning. Universities in the MiddleAges.* (pp. 201-223). Leiden: Brill.

Yates, F.A. (1966). *The art of memory*. London: Routledge & Kegan Paul.

The Author

BEN WILBRINK is a psychologist and a researcher in the Higher Education Research unit of the SCO-Kohnstamm Institute, Amsterdam. He specialises in research on the labour market and in educational measurement and selection, including personnel selection. He is currently working on an alternative to the traditional educational measurement model, an alternative that empowers teachers to build a measurement environment that is more conducive to learning instead of competition/selection.