Educational assessment in the Western world has a long but very irregular history. Two distinct threads are woven together: the first is the variety of settings in which testing itself came to have practical use while the second is the incorporation of increasingly rigorous methods by which to make sense out of the results of that testing. This chapter sets out some of the key developments in each of these two areas, from their origins until the dawn of contemporary psychometrics. For extended periods of time even the simplest improvements in either testing or statistics fought long and hard against tradition and inertia. It took many generations for the two threads to finally merge into a full-fledged science of educational measurement.

Seven centuries ago, one English college was deemed remiss in its responsibilities because its founder had determined that its recent graduates "...expressed themselves very inaccurately in the learned languages..." (Sylvester, 1970, p.19); the method of such determination was not described. A tradition of oral examinations slowly built up over several centuries; though the evidence is ambiguous, the earliest written exam may have been in place around 1510. By the time Isaac Newton attended college about 1660, however, the tradition had already fallen, and fallen hard. Not only were there no examinations but frequently lecturers themselves simply never showed up for classes. Then, in a series of major reforms the faculties of both Oxford and Cambridge, recognizing the deteriorated situation, decided to improve their curriculum and instituted regular examinations in a variety of topics. The first clear indication that written examinations were relied on for purposes of determining
admission and graduation was 1702 (Burt, 1936). The exams of this era were almost exclusively essay questions emphasizing factual recall; one extant example shows eight questions each in history and geography, and six in grammar, primarily Latin and Greek.

In the education of the younger pupils, examinations began to become more prevalent as textbooks for the grammar school came to be formulated into distinct grade levels.

The new sequences of textbooks allowed a more precise grading to be implemented in schools in various parts of Europe...Within the school a further step was the development and application of the principle of a child's regular progression through grades at various intervals of about a year (Bower, 1975, p.419).

The Jesuits, finding that such a procedure fit perfectly into their concept of the systematically ordered body of knowledge took up the idea with vigor, and it rapidly spread across Europe. Examinations were seen as the best way to focus the students' academic efforts, and to set criteria.

Yet, as one might expect, examinations were not necessarily viewed favorably by either faculty or students. Students at Yale rebelled outright in 1762, writing to the trustees that until there was an actual law on record they were unanimous in their refusal to participate (Smallwood, 1935). "Cramming" for exams was recognized as a major deterrent to good scholarship as early as 1786. Various faculty committees, hoping to aid things along, attempted to establish testing systems that would represent a proper balance of content, memory and skills. Balance proved exceedingly elusive to achieve as academic politics also had to be considered, so most of these well-intentioned attempts foundered. Testing continued and grew only because the schools continued to demand some formal way to assign ranks and evaluate student progress.

Meanwhile, in China, civil service examinations were already several millenia old. The earliest proficiency testing on record dates from 2200 B.C., and formal procedures for examinations date from