
1

Prologue

1.1 General Introduction: The Functions of Assessment

Recently I had a cause to enquire of a friend how he was recovering from an operation on his heart. He mailed a reply, which said, “They opened my chest, split my sternum, pried my rib cage apart, turned off my heart and lungs and let a machine do the work, replaced my aortic valve with a device fashioned from a cow’s pericardium, cut out a piece of my ascending aorta and replaced it with a Dacron tube, restarted my refurbished heart and lungs, pulled my sternum back in place, and stapled my chest back together. Miracle 1—I’m still alive after all that. Miracle 2—three weeks later, and I’m almost fully functional unaided and what mild aches and pains I have are managed well with gabapentin (a nerve pain pill) and Tylenol.”

“These docs are magicians [...]”

I am sure I would have felt the same. I am not sure that I would have considered it magic but I would certainly have thought it incredible even though hospital “soaps” lead me to believe that such operations are normal every day activity, much more exciting than operations on the brain! Be that as it may, the decision I would have to make, that is, to have or not to have the operation as my friend had to make would have been made on the basis of trust in the surgeons. Such trust is acquired from the understanding that the surgeons have considerable experience at doing such operations and have a not inconsiderable training that enables them to understand that experience as

enabling learning so as to better utilize that experience in the future. That understanding is reinforced by the knowledge that at all the stages in that training the surgeons have been examined or assessed (as some prefer) in formal situations to ensure they can do the job. Moreover, we expect those examinations to be psychometrically reliable and valid so that we can safely assume that the candidate will perform like that in the future. When we go to the surgeon's clinic, we expect to see his credentials, for that is what the accumulated certificates are, hanging on a wall. Should we not expect that from engineering educators?

Fortunately, we do not often have to trust surgeons but there are others in whom we have to place continuing trust, for example, the members of our family. Like them are the teachers to whom we trust our children. In the United Kingdom and the United States, that trust expects the teachers to act *in loco parentis* in activities that go well beyond the classroom although this is not the case in some European countries like France and Germany, where the teaching role is a teaching role without any social attachments. A great deal more is expected of teachers in boarding schools. Just like the surgeons, the trust extended to teachers is helped by the knowledge that they have had a similar training although not as long. They have acquired the knowledge that will enable them to teach a specialism; we expect a person who teaches mathematics to have a qualification in mathematics. But just as we expect surgeons to have gained a high level of craft skill, so many of us expect teachers to have developed the craft of teaching, or as it is more properly called pedagogy. I say many rather than all because there are individuals, politicians among them who think teaching is an intuitive activity that anyone can do. Their expectations do not stretch much beyond the experience of the teaching in their own school which they took to be easy. They find it difficult to believe that there is a serious activity of pedagogical reasoning that requires training on which experience can be built. As Shulman (1987) wrote, any explanation of pedagogical reasoning and action requires a substantial number of categories (i.e., Comprehension; Transformation [preparation, representation, selections, adaptation and tailoring to student characteristics]; Instruction; Evaluation [including testing]; Reflection, and New Comprehensions). Fortunately, the "many" do expect teachers to have credentials that document they have been trained in the theory and practice of teaching, and that to include assessed practice in real classrooms. There is a creeping realization that teachers exert very powerful influences over our children like no other they will experience, and these experiences can be for good or ill.

One of the primary functions of examinations is to aid the credentialing process. Thus, before a person can become a consultant, they have to perform junior roles and be mentored by senior doctors who all the time are monitoring their performance. There may even be performance tests to be taken. All of these tests are to judge their competency both of knowledge and performance. Knowing that they have had years of training is the first step in establishing trust. Much less is required of teachers although some countries require a period of probation and in some countries they are regularly evaluated by government inspectors in their classrooms.

Examinations and tests—assessments—perform many interrelated functions. For example, while an important function of assessment is to ensure that the goals of the

program are being met, the certification of that achievement provides an individual with a credential.

Credentials are also summative: they bring together all that has been learnt in training and they are gained only if a person demonstrates mastery of both skill and knowledge in some way or another. Examinations and tests (assessments) also function as motivational agents: they make some students very competitive, but all students benefit from the role of examinations and tests as formative agents, that is from the feedback they get about their performance the intention of which is to highlight their strengths and weaknesses. Related to the concept of credentials is the idea of a profession and belonging to a profession. In Britain and the United States, not so much in Europe, value is attached to belonging to a profession. Professions give prestige, status, and esteem (Hoyle, 2001) and in these countries, credentials initiate a candidate into the “tribe” and in some circumstances, they enable the “tribe” to regulate entry into itself. In the United Kingdom, groups seek professional status by increasing the level of qualifications required; for example, nurses are now required to possess a university degree in nursing. To be a professional is a valued goal, notwithstanding the sociological view that the term *profession* has lost its meaning (Runté, 1995).

There has been a long-standing debate about whether or not teaching is a profession. Heywood and Chevillat and Heywood (2015) have been bold enough to ask if “engineering educators are professional.” One outcome of the debate about the teaching profession has been a distinction originally drawn by Hoyle (1975, Exhibit 1.1) between restricted and extended professionalism that irrespective of whether teaching is profession or not indicates what we might expect from good and poor teachers. Logically, it would extend to teaching in higher and engineering education in particular.

In the United Kingdom, the issues of status, esteem, and power have continued to bother the engineering profession since the end of the Second World War. They were upset by a finding of Hutchings (1963) that entrants to engineering schools had lower A level grades (see Section 3.1) than those in the sciences and they have bothered about such differences ever since. Similarly, they believe and continue to believe that there is a shortage of qualified engineers. Currently, in the United States it is supposed that there is a shortage of candidates for STEM (Science, Technology, Engineering and Maths) courses.

My friend who is very distinguished in his field of activity went on to say, “I’m ashamed to call myself doctor—I can’t do anything that even comes close.” While I do not happen to believe that is the case I was rather facetious in my reply for I said “on this side of the Atlantic you would not have that problem because we call surgeons ‘Mr’ or ‘Miss’ not doctor which is reserved for physicians!” This is said to point out that there are considerable differences between the educational systems of Europe, the United States, and United Kingdom and therefore with many countries of the old British Empire, where Britain established the systems that they developed. This is particularly true of Australia and New Zealand (Yeung, 2014) and countries in Asia. Canada, in contrast, mirror the system in the United States. Because education structures vary from country to country, establishing data that is transferable is exceptionally difficult. Although everyone is concerned with the basic parameters of examining and testing namely, achievement, validity, and reliability, exogenous variables that are unaccountable influences on the

Restricted Professionalism in Engineering Education	Extended Professionalism in Engineering Education
Instructional skills derived from experience	Instructional skills derived from mediation between experience and theory
Perspective limited to immediate time and place	Perspective embracing broader social context of education
Lecture room and laboratory events perceived in isolation	Lecture room and laboratory events perceived in relation to institution policies and goals
Introspective with regard to methods of instruction	Instructional methods compared with those of colleagues and with reports of practice
Value placed on autonomy in research and teaching	Value placed on professional collaboration in research and teaching
Limited involvement in nonteaching professional and collegial activities	High involvement in nonteaching professional and collegial activities
Infrequent reading of professional literature in educational theory and practice	Regular reading of professional literature in educational theory and practice
Involvement in continuing professional development limited and confined to practical courses mainly of a short duration	Involvement in continuing professional development work that includes substantial courses of a theoretical nature
Instruction (teaching) seen as an intuitive activity	Instruction (teaching) seen as a rational activity
Instruction (teaching) considered less important than research	Instruction (teaching) considered as important as research
Assessment is a routine matter. The responsibility for achievement lies with the student	Assessment is designed for learning. Achievement is the coresponsibility of the institution, instructor (teacher), and student

EXHIBIT 1.1. Eric Hoyle's characteristics of extended and restricted professionals among schoolteachers adapted for teachers in higher education (Hoyle, E. (1975). Professionalism, professionalism and control in teaching, in V. Houghton et al. (eds.), *Management in Education: The Management of Organizations and Individuals*. London: Ward lock in Association with Open University Press. See also Hoyle, 2001).

data often make it difficult to ascertain what is actually happening within the system, its teachers, and its students (Berliner, 2014).

Apart from the basic functions discussed earlier in this section and the difficulties of making comparisons, there are, I think, two issues that are common to most assessment systems. The first is illustrated by the text in Exhibit 1.2. It is the opening paragraph of a book that I published on *Assessment in Higher Education* in 1977 (Heywood, 1977). I did not put the last sentence in bold as it is here. In spite of changing structures in the United Kingdom and Ireland, I find that colleagues have an affinity with the picture in that Exhibit. I had titled the chapter after a weekly political satire televised by the BBC and hosted by the late David Frost called "Not so Much a Programme, More a Way of

“Senate approved the syllabus for our new degree course at its last meeting. X will act as Director of Studies and co-ordinate the timetable for the new degree (Head of Department, January, 1975)”

“Please let me have your questions for the Part I examination (of the new degree course) by the end of next week at the latest (Director of Studies February, 1975)”

“Hell, I’d forgotten about the exam. Let’s see what I set in this course in the other degree programme last year... (Teacher, February, 1975).”

The Director of Studies got his questions several weeks later and the students were sensible enough to look up the questions set by the lecturer in previous examinations in similar courses, and with a bit more luck, few will fail. **Examinations are the great afterthought of the educational process.**

EXHIBIT 1.2. Extract from Chapter 1 of Heywood, J. (1977). *Assessment in Higher Education*. Chichester, UK: John Wiley & Sons.

Life.” Another distinguished colleague in the United States told me the other day that he rushes around at the end of the semester putting together his tests for the end of term. And several colleagues apologized for delay in replying to my e-mails before Christmas because they had to mark tests. Examinations (tests) are a nuisance, and they remain a nuisance!

When I published that book, there was no text on assessment in higher education available in the United Kingdom. However, in the United States, Paul Dressel had edited a useful text on *Evaluation in Higher Education* in 1960. The chapter on “learning” by J. Saupé occupies an important place in my memory system for the very reason that he defines learning. It is that definition that informs this text: “Learning is that process by which experience develops new and reorganizes old responses” (conceptual frameworks). His views of the principles by which learning is achieved are shown in Exhibit 1.3 and still resonate, although McKeachie (1974) pointed out that learning is an interaction between a variety of complex variables that Saupé’s list does not seem to convey.

In the 1960s, *Universities Quarterly* published an important issue on the topic as well as another on the topic of wastage (then termed “dropout(s)” in the United States), that is the proportion of students withdrawing before they had completed their courses. I believed that there was a need for a theory of assessment in higher education that looked at all the factors that contributed to achievement. There are many. By 1987, it was apparent that the book needed to be substantially revised and a second edition was published in 1989 (Heywood, 1989): 10 years later the story repeated itself and a completely new book was published in 2000 (Heywood, 2000). Whereas I used to think I was providing a theoretical basis for discussions about assessment, I think in today’s world it might have been better called a philosophy. I argued that assessment should not be the afterthought of the curriculum process but integral to it and it is that philosophy that underpins this study. Philosophy has a major role to play in determining the aims

1. *“Without appropriate readiness a learning experience will be inefficient. Learning will not occur”* (relates to the entering characteristics of students and the assessment of prior knowledge).
2. *“Learning will more effective when motivation is”* (as opposed to extrinsic: relates to formative assessment and self-assessment).
3. *“Learning proceeds much more rapidly and is retained much longer when that which has to be learned possesses meaning, organization and structure”* (relates in the first instance to the learning of concepts; also relates to the planning of tuition).
4. *“The learner learns only what he himself does”* (a reason for the use of inquiry/discovery-based learning)
5. *“Only those responses that are learned are confirmed.”*
6. *“Transfer can only occur to the extent that students expect it to.”*
7. *“Transfer will occur to the extent that students expect it to.”* (5 and 6 relate to both the design of instruction and the design of assessment).
8. *“Knowledge of a model(s) of problem finding and solving or aspects of critical thinking can contribute to its improvement.”*

EXHIBIT 1.3. Saupé’s conditions that lead to successful learning. Adapted from Saupé, J. (1961). Learning. In: Dressel, P. L. (ed.), *Evaluation in Higher Education*. Boston, MA: Houghton Mifflin.

of a program since our values and beliefs determine those aims, and those values and beliefs are supported by our personal philosophy.

At the time of writing, engineering education is, in some degree, in turmoil. A number of leading engineering educators are looking for a new vision, a new curriculum that would emerge in different ways in different parts of the world. This book seeks to establish what we have learnt about assessment, its design, implementation, and validation so that it can contribute to the understanding and making of that vision. Necessarily, it surveys developments but it is not a textbook as such, yet may serve as a handbook. The other point about which I think there can be universal agreement is the total inability of people in the system to agree a common set of terms, the most critical of which is assessment.

1.2 Health Warning: Ambiguities in the Use of the Term “Assessment”

“Assessment” seems to have become a term that means what you want it to mean. For example, it has more or less replaced the term “evaluation” that was commonly used in the 1960s and 70s, not it would seem for any significant reason. As is seen in section 1.8, in the late 1960s the Committee of Vice-Chancellors and Principals (CVCP) in the United Kingdom began to use the term *assessment* as a cover for examinations and coursework (CVCP, 1969). By the end of the 1980s, Hartle (1986) had distinguished between six usages of the term “assessment” in the United States. Sadler (2007), an Australian, observed “that many of the terms we use in discussion on assessment and

grading are used loosely. By this I mean we do not always clarify the several meanings a given term may take even in a given context, neither do we necessarily distinguish various terms from one another when they occur in different contexts. For example, the terms 'criteria' and 'standard' are often used interchangeably." Yokomoto and Bostwick (1999) chose "criteria" and "outcome" to illustrate problems posed by the statements in ABET EC 2000 (now called ABET Engineering Criteria). In this text, there is room for plenty of confusion between terms such as "ability," "appraisal," "assessment," "capability," "competency," "competences," "competencies," "criteria," "criterion-referenced," "evaluation," "objectives," "outcomes," "objectives," and "performance-based," all of which will be used in the text that follows. Semantic confusion abounds and there is disagreement about plurals, for example, "competences" or "competencies." While some just represent changes in usage, as for example "assessment" for "evaluation" others hide different theories, for example the "inside" and "outside" theories of competence development discussed section 1.3. Larkin and Uscinski (2013), writing of the evolution of curriculum assessment in a particular program, felt it necessary to define their terms in a separate section in their paper (see Appendix A).

Writing of performance assessment, the concept that was the driving force behind this text, McGuire (1993), a distinguished medical educator, argued that the use of the term *performance* is a naïve disregard of plain English. We are not concerned with performance *per se*, "rather we are concerned about conclusions we can draw and the predictions we can make on the basis of that performance. And this is a high inference activity." One of the reviewers of this text suggested that the original title that was "*Performance Based Assessment in Engineering: Retrospect and Prospect*" should be changed because he or she feared that engineering educators would think it had been written for industrialists and not for them. To be clear, this book is addressed to both.

Our emotions are expressed through language and the terms used can cause us to be very emotive, hence the care with which they have to be chosen. In the British Isles, if university lecturers were called "*instructors*" and they would be very offended. Instruction is what training officers do when they teach apprentices craft skills! Similarly in collegiate universities "tutor" has a different meaning to its use in the United States, so I have used "teacher" where I would sometimes have used "tutor."

Related to the concept of "performance" is the concept of "competency." But even this term has its difficulties. Griffin (2012) wrote that "whilst the competence vocabulary may be 'jargon' to some, it offers clarity to the notion of competency and competence as is evident in management literature (Young, 2002). In contrast, the medical and nursing literature offers the reverse, for example, some authors separate performance from competence, whilst others argue that there is direct relationship and others, still use 'competence' when they mean 'competency' and vice versa (Talbot, 2004). Defining the terms is more than semantics as it has implications for teaching, learning, assessing and the granting of licence to practice to medical and nursing practitioners" as we shall see. The same applies to engineering and engineers, as well as to other professions and their practitioners. According to Griffin, the prevailing trend in medical and nursing literature is that performance is what the practitioner does and competence is what the practitioner is capable of doing.

One of the functions of assessment is the measurement of achievement that *ipso facto* is a measure of a specific performance. That seems to be the way it is taken by the engineering community for a crude count of papers presented at the *Frontiers in Engineering Conferences* between 2006 and 2012 yielded some 80 papers that had some grammatical form of the term “assessment” in their title that might be relevant to this text although only one of them included the term “performance.” Eleven included the term “competence,” which is clearly related to performance. But most of them were European in origin. Two years later, the 2014 proceedings of the American Society for Engineering Education contained a few papers from the United States that used competence synonymously with outcomes. It is evident small changes in language usage are taking place all the time.

Therefore, while this text clearly relates to the testing of academic achievement of student learning, that is only relevant in so far as it predicts the performance of individuals in the practice of engineering, and many attempts have been made to do this (e.g., Carter, 1992; Nghe, Janecek and Haddaway, 2007; Schalk et al., 2011). Much of the recent concern with the assessment of performance in the United Kingdom and elsewhere has been driven by politicians seeking accountability in order to try and demonstrate that value is added to the students and the economy by the education they have received (Carter and Heywood, 1992). This is nothing new; moreover, it is terribly difficult to do (Berliner, 2014).

While the term “assessment” is now commonly used to describe any method of testing or examining that leads to credentialing it embraces “formative” assessment that while not contributing directly to credentials helps students improve their “learning,” and “self-assessment” that may or may not contribute to those credentials (George and Cowan 1999). It may be carried out in real time (Kowalski, Kowalski, and Gardner, 2009). An American-centered view on the techniques of assessment written for engineers is found in the *Cambridge Handbook of Engineering Education Research* (Pellegrino, DiBello, and Brophy, 2014). In Appendix A of this text, I have tried to provide an every person’s guide to the changing terminology in the field of assessment with the exception of the psychometric terms of reliability and validity. The latter is considered below.

1.3 The Assessment of Persons for the Professions

Given that the purpose of professional study is preparation to undertake work in a specific occupation, a major question is the extent to which a curriculum prepares or does not prepare students to function satisfactorily in that occupation. Assessments that make this judgment may be made at both the level of the program and the level of student learning. In recent years, engineering education in the United States has had a significant focus on the program level for the purpose of accreditation. In the Health Care professions, both doctors and nurses receive clinical training and are assessed on their performance during this training. Assessments are made on the basis of judgments they make about “real” patients and rubrics for their assessment date back to the 1960s (Freeman and Byrne, 1973; see Section 3.5). The problems they faced continue to be relevant. Similarly, when teachers are trained, they have to conduct live classes of pupils in the age range that they

(the students) want to teach (e.g., elementary [primary], High School [secondary] that have to be assessed) (e.g., Dwyer, 1994; Educational Testing Service., 1995). In both cases, “clinical” work is integrated with the curriculum in one way or another. Over the years, schedules for the assessment and self-assessment of the performance of teachers have been developed (Roberts, 1982; Reeves, 2009).

There is no equivalent “clinic” for engineering students except in cooperative programs (“sandwich” in the United Kingdom) where the students interweave, say 6 months of academic study with 6 months of industrial training throughout a period of 4 years. In that case, given that they are given jobs that are appropriate to the professional function, the equivalent of a “clinical” experience is provided and their professional performance can be assessed. This is what this writer has always taken performance-based assessment to mean, a view that has some similarity with that of Norman (1985). Miller (1990), who has had a great influence on medical education, put forward a pyramid model for the assessment of medical students. At its base is “KNOWS” (knowledge) that supports “KNOWS HOW” (competence) and that in turn supports “SHOWS HOW” (performance), and “DOES” (action) at the apex. He argued that competence was the measurement of an examinee’s ability to use his or her knowledge, this to include such things as the acquisition of information, the analysis and interpretation of data, and the management of patient problems (Mast and Davis, 1994; see also Norman, 1985). Thus, while competence embraces the whole of the pyramid, university examinations at the time, only assessed competence from the “KNOWS HOW” perspective. Miller took the view of Freeman and Byrne (see Section 3.5) and others that no single assessment method could provide all the data required to judge if a person was able to deliver the professional services required of a physician. This same principle informed the designers of the engineering science examination described in Section 3.1. Rethans et al. (2002) took a different view. They thought that Miller’s pyramid was no longer useful and suggested that the top two sections should be inverted thereby classifying performance and competence into the “DOES” section and the “SHOWS HOW” section. This approach, known as “The Cambridge Model,” defines all assessments taken under examination like settings as competence-based assessments, and all assessments of actual practice as performance based. In this case, Freeman and Byrne’s (1973) scale of overall competence in the practice-based situation is performance based (see Section 3.5). Griffin (2012) considered that the examples given by Rethans et al. “bear no resemblance to a cohesive model for assessment of competence or performance. The validity and reliability of some of the methods would also need to be established to avoid instances where doctors are wrongly classified.” Unfortunately, Griffin makes no reference to Freeman and Byrne’s assessment practices (see Section 3.5) that would seem to provide a cohesive model for postgraduate education that could be used as an exemplar. But she does take up an issue that is avoided in discussions of this kind namely, incompetency.

It is not a function of this text to consider the social psychology of incompetence that is, with the way that incompetence is dealt or not dealt with by professions. It is, however, the task of competency-based measures to state what incompetency is. In my experience, there is in Ireland, at least there was in my day, a desire not to want to fail intending teachers except in very severe circumstances. Griffin (2012) reports some evidence that some student nurses passed their clinical assessment without demonstrating competence,

and she notes that it seems that clinical assessments may not always be recognized by third-level institutions as being important within the nursing program. A key purpose of self-assessment is that students learn what they are incompetent in (see Section 8.9). This has implications for the design of questionnaires that require respondents to make judgments about their capabilities.

More generally, it seems that the “inside” view of competency has been taken by the engineering profession. Competency in this view is something within the person, the quantity of which can be measured. It depends on habits of mind, including attentiveness, critical curiosity, self-awareness, presence, and a willingness to recognize and correct errors that is, to (self)-evaluate (Griffin citing Epstein and Hundert, 2002). The corollary is that if it can be measured it can be taught. Thus, many academics and industrialists believe that the competences industry requires can be taught without any assistance from industry. This position suggests that the acquisition of competence is not a developmental process, but this is rather contrary to observation as is rather well explained by Sternberg’s (2005) model of the development of abilities into competencies, and competencies into expertise (see Sections 3.2 and 5.8). Students are experts at many levels. They may have a hobby such as photography in which they have the highest level of expertise. Young children often have more expertise with computers than their parents. It is evident that even in academic studies, students develop expertise as many papers on concept learning in engineering show (see Fordyce, 1991, in particular, and Cowan, 2006). There should be an increase in expertise with each additional year in the learning institution.

Griffin (2012) argues that the idea that competence is located in the individual is deeply entrenched in Western conceptions of intelligence and the mind (see Section 10.1). It is this theory that governs the thinking of engineering educators, more especially organizations like ABET (see Chapter 9 and also Section 6.8, which are based mainly on this premise). Even within this framework, there are differences of opinion about what a competence is (see Section 5.1). But there is also an “outside” view that suggests that the self has to be considered within the context that a competence is acquired. Sandberg (2000) has demonstrated that the acquisition of specific competences within an industrial organization is very much dependent on the context in which they are acquired (see Section 10.2). Taken together, these have a bearing on the structure and design of courses in engineering as well as their assessment (see Chapter 12).

It is not surprising that among the professions especially medicine that there should have been several studies on what it means to be professionally competent. From the literature, we are led to the view that competence is a complex, if not elusive, concept that is necessary for understanding the purposes of the education of professionals. Its assessment is, therefore, likely to be equally difficult. Given the closeness if not equivalence of “outcomes” with competence, the same finding applies.

1.4 The Engineering Profession

To become a professional engineer in the United States or a Chartered Engineer in the United Kingdom, the graduate is required to have completed a number of years of approved experience. In many other countries, there are similar requirements. Many of

those in the old British Empire follow the pattern established in the United Kingdom, for example, Australia and Ireland. But there is no requirement for engineering educators to have professional status. The changing patterns in the workforce described in Chapter 12 suggest that new approaches to credentialing will be required as well as assessment of the work undertaken throughout a person's career.

Industrialists in the United Kingdom hoped that the Colleges of Advanced Technology (CATs) that were created in 1956 would produce engineers more suited to industry than the products of the engineering science courses offered by the universities that they thought prepared students for research. The electrical industry, in particular, had made a very heavy investment in them but it was disappointed with the results (Heywood, 2011). G. S. Bosworth, one of their leaders, published a paper in *Universities Quarterly* (Bosworth, 1963), in which he called for engineering courses to be more creative (see Section 2.3). Subsequently, he gave up the idea of changing the undergraduate curriculum, and in a report of an official committee that he chaired, rested his hopes that the needs of industry would be met by postgraduate programs (Bosworth, 1966; Heywood, 2013). Since the majority of engineering students do not pursue such programs, it is pertinent to ask how, if at all, engineering educators can predict the capability that a person will bring to an engineering task. Or, for that matter the occupation they have entered. To be fair, in colleges, students do a lot of laboratory practice and project work that is relevant. However, such predictability presupposes that something is known about where students go and what they do when they graduate.

Until this decade, engineering educators have paid very little attention to professional practice although this is now being rectified (Williams and Figueredo, 2013; Trevelyan, 2014). Throughout the second half of the twentieth century, there has been a strong presumption among academics that university examinations predict subsequent behavior. Equally, at regular intervals, industrialists and the organizations that represent industry have complained that graduates, not just engineering graduates but all graduates, are unprepared for work (performance) in industry. To put it in another way, the examinations being offered did not have predictive validity. What is remarkable, as will be demonstrated in the sections that follow, is the persistence of these claims over a very long period of time (50 years), in spite of a lack of knowledge by all the partners of what it is that engineers actually "do." Nevertheless, from time to time during the last 60 years, attempts to find out what they do either directly or indirectly have been made. Each has contributed to the picture we have, which is not inconsequential (see Chapters 4 and 11).

If education is associated with what is learnt in college and training is associated with what is learnt in industry in order to become an engineer who can practice, it is clear that education cannot be divorced from training. It is also self-evident that the "complete" preparation of an engineer for practice cannot be accomplished only in a course of university study. Industry has a vital role to play in that training. One aim of this text is to demonstrate that that is the case.

Assessment has become a preoccupation of many engineering educators. At the time of writing the 2013 ASEE Conference proceedings, search engine yielded 273 entries for the term *assessment*, and most had the term in the title of their paper (<http://www.asee.org/>). There were 272 in 2014. There were also sessions on assessment

at the Frontiers in Education Conferences and there are contributions in the growing number of journals in engineering education. The material is overwhelming.

Challenging the view that more rigorous research is the key to achieving educational reform, Felder and Hadgraft (2013) write: “we believe that if engineering education research were stopped completely right now (which we are in no way advocating) and engineering faculties could be induced to put into practice everything we know about teaching and learning from past research, cognitive science and experience, then we would achieve innovation with impact to an extent beyond the wildest dreams of the most idealistic reformers. The question then becomes, how can we do that?”

I believe that this is as true of assessment as it is of teaching and learning. My contribution to the answer to their question is to show the truth of their proposition through this survey of the development of assessment during the last 60 years with a view to suggesting directions it might take in the future as an integrated part of the curriculum process. Can we learn from the past ideas, philosophies if you will, that can enable us to judge where we are in the present? It is the contention of this study that we can.

1.5 The Development of Higher and Engineering Education as Areas of Academic Study in the 1960s

The early 1960s saw the beginnings of research in higher education in the United Kingdom, and in particular in technical and technological education (Heywood and Ann Abel, 1965). In the United States, Nevitt Sanford (1962) had edited *The American College* and Alexander Astin had begun his mighty studies of the impact of college on students (Astin, 1968). In the United Kingdom, there were rumblings that university teaching was not all that it should be and by 1964 the Universities of Essex, Lancaster, and Manchester had research units in various aspects of higher education. At Lancaster, a university that opened its doors to students in 1964, the Vice Chancellor (President) created a two-person department of higher education, the first in the country. This writer was one of those two persons and was appointed to do research in the area of university examinations (the term *assessment* was only just beginning to be used). The other person was appointed to research university teaching methods. A researcher also in the area of examinations was appointed at the University Essex. He (Roy Cox) became widely known for a much-cited review of research in examinations that was published among other papers on the same topic in *Universities Quarterly* that had been brought together by a working party of the newly founded (1964) Society for Research into Higher Education (SRHE) (Cox, 1967). When Bradford College of Advanced Technology was given the status of a technological university, it created an education department that undertook numerous studies of the university at work, including teaching and learning in technology (Cohen, 1968; Smithers, 1965).

1.6 Assumptions About Examinations: Reliability

It should be appreciated that at the time there were no textbooks published in the United Kingdom on teaching and learning in higher education let alone assessment and

examinations which, to a limited extent, was made good by a monograph on group teaching (Abercrombie, 1964). It showed that there was much more to teaching than lectures, and in this respect by a much-quoted Penguin publication on the use and abuse of lectures (Bligh, 1971). However, two books on assessment related to higher education had been published in the 1930s. They reported investigations into the reliability of examinations. They were written by Sir Philip Hartog (Principal of the University of London) and E. C. Rhodes (a distinguished statistician) and published in 1935 and 1936, respectively (Hartog and Rhodes, 1935, 1936). Their results suggested that university examinations were not very reliable in the sense used by psychometricians. Reliability in this case, or “consistency” as Steven Wiseman, a distinguished psychologist preferred to call it, is the measure of the extent to which a test or examination gives consistent results with repeated applications. This is the reliability of a test in time. In the world of psychometric testing, much attention had focused on “consistency,” but a test that is “consistent” is not necessarily valid. So when I was interviewed for the job at Lancaster, I suggested that we knew a lot about reliability (or unreliability!) and that we should now begin to look at questions of validity, and it was the issue of validity that guided my subsequent interest in examinations research.

The degree of consistency is determined by the method of correlation and the particular inconsistency that Hartog and Rhodes considered related to the fact that when two examiners are given the same essays to mark, not only do they assign different marks for the same performance but they are also likely to produce different distributions for the same group of candidates. I found that there could be different distributions as between the subjects of the curriculum (Heywood, 1977, p. 31). It could be that some subjects were more difficult than others or that the marking was more objective in some subjects than others. Either way, in some subjects the probability of getting a high grade was better than that in other subjects, a fact that was of particular concern to the National Union of Students (1967, 1968). Most recently, Lorimer and Davis (2014) have reported on a 10-year longitudinal study of mathematics, and (over 6 years) engineering assessments administered to preengineering students who show a remarkable degree of consistency when responding to the Force Concept Inventory.

The Hartog and Rhodes study led some to condemn essay examinations, although Philip Vernon (1965), another distinguished psychologist, rose to their defense. He noted, as had other reviewers, weaknesses in some of the experiments that were conducted, but in contrast to those reviewers he was impressed by the smallness of the discrepancies rather than their largeness. He found that in the best conducted examinations the median disagreement between any two examiners was not more than 3%. However, Vernon argued that Hartog and Rhodes had to be taken seriously because less thorough examinations are shown to be deplorably unreliable: “that in the absence of a scheme of instructions drawn up and applied by experienced examiners much worse discrepancies may arise; that when the average and dispersion marks are not standardised, gross differences may appear in the proportions of credits, passes and fails etc. which are awarded and that even a percentage error may make all the difference between a pass and a fail, or a first and a second class.”

These concepts were not part of the language of examining used by teachers. In the many examiners meetings I attended, I found that across the spectrum of subjects there

Number of Items in Objective Test	Standard Error
<24	2
24–47	3
48–89	4
90–109	5
110–129	6
130–150	7

EXHIBIT 1.4. The standard error for items (questions) that can be scored 1 or 0 that are intended to measure performance within a reasonably generous time limit due to P. B. Diedrich (1960). *Short Cut Statistics for Teacher Made Tests*. Princeton, NJ: Educational Testing Service. McVey (1976) attempted to calculate the standard error of a problem-solving examinations in electronics and arrived at a standard error of 8 for a typical 3-hour paper.

was little if any understanding of the concept of “standard error.” Suppose a candidate scores a mark at a borderline grade level such as 59 instead of 60, or 69 instead of 70, assessors would stand by their mark, irrespective of the fact that in any mark there is an inbuilt error, and this even applies to objective tests (see Exhibit 1.4). Moreover, in any group of examiners, there are those prepared to mark “up” and others who will always mark “down.”

In defense of reliability, it was also pointed out that the final year classes in many subjects might be rather small and the teacher may know the students very well. Double marking that some universities now insist on is fraught with difficulty. McVey (1975), who examined scripts in electronic engineering, showed that marks left on a script by the first examiner could influence those of the second examiner. In any case who is right? McVey’s investigations also showed that examinations in electronic engineering that might be thought to demonstrate a high level of reliability were not as reliable as expected. McVey also showed that the standard error was an important factor in marking (McVey, 1976a, 1976b). Engineering educators continue to be interested in reliability (Allen et al., 2008).

1.7 Myths Surrounding Examinations

University examinations were shrouded in myth. Oppenheim, Jahoda, and James (1967) listed 20 assumptions that were made about university examinations that for most part could be tested empirically for their validity. These are shown in Exhibit 1.5. They were, of course, specific to the system of examining in Britain, although some are universal and some have changed or are being pursued. The fourth and fifth assumptions are one of the reasons for this text and the pursuit of the engineering education to relate assessment to performance in the real world beyond academia. Nowadays in the United Kingdom, much more use is made of coursework assessment (seventh assumption). The 10th assumption was never true of engineering because the professional institutions have Royal Charters that enable them to examine at this level. Attempts have been made

1. The assumption that university examinations can include some so-called imponderables such as “quality of mind,” “independent critical thinking,” breadth, etc., in their assessment.
2. The assumption that “quality” of academic performance is rateable on a single continuum for first-class honors to failure.
3. Whereas many courses include a good deal of practical work, and a few approach some kind of apprenticeship training scheme or sandwich (cooperative) course, we usually pay less attention to those aspects of the course when examining.
4. To some extent, it is assumed that examination performance is a mock–real-life performance.
5. The assumption that each examinee should have individual responsibility for his own performance; we do not expect collaboration or teamwork, no matter how common this may be in real-life performance.
6. The assumption that a student who fails has only himself to blame for not working hard enough, or for being stupid, or in some other way.
7. The assumption that the proper place for examinations is at the end of certain courses—not later or sooner.
8. The assumption that university teachers should also be university examiners and university selectors.
9. The assumptions about the impartiality of examination.
10. The assumption that the university should have the sole right to examine at this level.
11. The assumption that the use of external examiners prevents bias.
12. The assumption that forced regurgitation of knowledge under stress is predictive of future performance.
13. Assumptions concerning mental growth and development and the acquisition of an “educated mind.”
14. The assumption that pressure is required.
15. The assumption that anxiety is necessary.
16. The assumption that examination results should be distributed in a certain way.
17. We are forced to make, and then retract, all kinds of assumptions about the comparability of degree from university to university and from country to country.
18. The assumption of the need for uniformity in undergraduate examinations: all students in a given year group must pass the same examination paper, and we do not allow examinations to be tailored to individual needs.
19. The assumption that “learning” is to be valued for its own sake and not merely as a preparation for career and financial gain.
20. The assumption that the outside world wants the results of university examinations or takes much notice of them.

EXHIBIT 1.5. Twenty assumptions underlying the use of university examinations in the United Kingdom listed and discussed by Oppenheim, A. N., Jahoda, M., and James, R. L. (1967). Assumptions underlying the use of university of examinations. *Universities Quarterly*, 21(3), 341 (also cited in Heywood, J. (1977). *Assessment in Higher Education*. Chichester, UK: John Wiley & Sons).

to improve external examination but questions remain (Warren-Piper, 1994), and there have been worldwide attempts to ensure the comparability of degrees, for example, the ABET Engineering Criteria, the Bologna Agreement, and the Washington Accord (17th assumption) (Bucciarelli, Coyle, and McGrath, 2009). What industry, of for that matter society, wants from examinations remains an “open” question. At the time of writing, the UK division of Ernst & Young (EY) the professional services organization announced a transformation of their recruiting policy because their own research had shown that screening graduates on academic performance alone was too blunt an approach to recruitment. They had found that success in higher education was not correlated with performance in subsequent professional qualifications. Therefore, they proposed to evaluate candidates for their potential through the use of online tests. Only at the final stage would the academic performance be made known to the recruiters. While they would still value academic achievement and maintain high intellectual standards, they hoped that this would enable the organization to become more socially inclusive. It will be of no small interest to see if EY achieves this goal. PwC is also selecting graduate trainees by aptitude and behavioral tests (The Times, 2015).

The reader should be reminded that one of the difficulties when talking or writing about the assessment of student learning is that the assessment systems to be found in countries such as the United Kingdom and the United States are so different that they are difficult to understand if you do not work in those systems. Moreover, they are subject to change. In the early 1960s, the stereotype of assessment in the United States was of objective tests (known to the public as multiple-choice questions) set during and at the end of courses and its equivalent in the United Kingdom of 2- to 3-hour written examinations set at the end of the year in each of the subjects studied, grades being distributed in the former at the top end of the mark spectrum while the latter were neo-norm referenced.¹ Objective tests can meet the criteria specified by Vernon. The two tests commonly used for selection in the United States (Scholastic Aptitude Test [SAT] and the American College Testing Program) are standardized tests and have very high reliability. An objective test almost inherently has a higher reliability than an essay test but teacher-designed tests that have not been designed and piloted properly may not be as good as those that are. Ager and Weltman (1967), who were participants in the SRHE group, concluded that “no single examination technique is completely satisfactory in terms of both reliability and validity,” and for this reason they thought a variety of techniques should be used. For the quite different reason that anxiety could be reduced, university medical officers also recommended the use of a variety of techniques (Malleon, 1964; Ryle and Lungi, 1968; Ryle, 1969). It was not appreciated that if a distributed model of assessment was used that the redistribution of stress in the system might have an equally negative effect on those who preferred terminal examinations. The practice of examining is very much the art of compromise. But these studies did not relate their findings to validity and the purposes of university examinations in any detailed or considered way. However, support for multitechnique approach is also to be found in a detailed analysis of recent literature together with changes in the techniques he uses for teaching and assessment by Parsons (2008). Apart from encouraging a variety of styles of assessment, he encouraged the use of open-book examinations because of the memory support they gave students, opportunities for students to comment on the

assessment at or around the time of assessment, and the relaxation of time constraints on assessment activities.

Currently in the United States, a mixed techniques approach is often cited that makes use of direct and indirect methods of assessment. By indirect methods are meant such things as exit and other interviews, archival data, focus groups, and written surveys and questionnaires. In addition to all kinds of examinations, Sundarajan (2014) includes simulations, behavioral observations, and performance appraisal as direct measures. The problem with mixed approaches, as it is with continuous assessment, is that they can easily overload students to the extent they are unable to cope with the work they are given. This seems to be an aspect of testing that has received little consideration (Myllymaki, 2013).

1.8 The Introduction of Coursework Assessment

However, concerns about university examinations in the United Kingdom not least the view that students could have an off-day on the day they took their examination led to changes in the 1960s notably to the introduction of what was commonly called “continuous assessment” that in effect was the periodic assessment of some form of work done during the course (Coursework Assessment). There is a correspondence with what happens currently in American programs where assessments are made during each course that include tests, quizzes, home examinations, and homework. It may incorporate self-assessment.

In the United Kingdom, it was introduced in a haphazard way such that natural justice could be compromised. For example, a student could be required to undertake a lot of coursework for which he or she would receive no marks in the final collation of marks, whereas in other courses the students might be told that 10% of their final mark would contribute to their final score (grade). The variations in practice in one university are shown in Exhibit 1.6. The study summarized in this exhibit did not take into account the procedures adopted by the Open University. Nowadays, courses are a mix of the traditional 2 or 3-hour written papers and continuous assessment. Homework is not set. The weighting between the two varies considerably across the British Isles. It could be as many as 80% for coursework or equally for the examinations. One program in Design Technology (for teachers) relies solely on coursework assessment for its grade. There have been some radical innovations in engineering courses and there is an affinity between some practices of continuous assessment and mastery learning (e.g., Cole and Spence, 2012; see Section 6.4).

Today continuous assessment is used in a number of institutions across the world and is seen as a means of improving performance. At the University of Oulu in Finland, its purpose “is to help the students themselves to become more effective self-assessing, self-directed learners and it is based on cognitive, constructivist, and socio-cultural theories” (Myllymaki, 2013).

Continuous assessment is also used by universities seeking accreditation from ABET that are located outside of the United States, where a range of instruments is used to evaluate the level of achievement of the program’s educational objectives

Cumulative assessment

A proportion of the mark in a 3 (or 4)-year course is arrived at from scores achieved during the first and second years. The proportion of marks allowed for the first year is usually smaller than that for the second.

For example, 1st year = 12% of the years marks; 2nd year = 24% of the years marks; 3rd year = 64% of the years marks.

The marks may be achieved by combinations of coursework assessment and examinations, or examinations alone. For purposes of certification, they could be equated to 100%.

Diagnostic coursework

An early assessment of student coursework to determine those in difficulty and the nature of the difficulties.

Informal coursework assessment

Most tutors make a judgment about the qualities of their students while a course is in progress. Such judgments are often used to moderate the marks given to candidates at the final meeting of examiners with the external assessor. Students are not informed about such procedures.

Formal Coursework Assessments

The characteristics of such systems are that the students know how coursework will contribute to their final mark. There are several systems (the terminology is the writers):

- (i) *Fixed percentage schemes*: In such schemes, coursework is formally assessed and contributes a fixed percentage to the total degree mark. In England, some universities operating such procedures do not impose restrictions on departments. The proportions awarded depend on the value ascribed by individual departments to coursework assessment.
- (ii) *Positive moderation schemes*: These are schemes in which coursework is formally assessed but the result is used only to raise a candidate's mark.
- (iii) *Formal requirement schemes*: are those that require satisfactory performance in coursework, before a person can either obtain a final degree or sit at a final examination.

The scheme adopted depends very much on the department (or faculty) objectives in initiating coursework assessment procedures. Broadly speaking, there are two categories that can be described as *supplementary* and *complementary* in terms of the information sought by examiners about a candidate.

- (1) In schemes that provide *supplementary* information, the examiners hold that the coursework assessment is measuring the same abilities (qualities) as the written examination. It is, therefore, a check on the written examination and, as such, is used as a means of moderating the final mark.
- (2) Coursework work assessment that provides *complementary* information to the examination is thought to measure different qualities to those measured in written papers.

EXHIBIT 1.6. Types of coursework in operation in English Universities in the late 1960s. Heywood, J. (1969) in Committee of Vice Chancellors and Principals and Association of University Teachers report of the 1969 Spring Conference. *The Assessment of Academic Performance*. London, Committee of Vice Chancellors and Principals (also cited in Heywood, J. (1977). *Assessment in Higher Education*. Chichester, UK: John Wiley & Sons).

and outcomes. They embrace course assessment (Abu-Idayil and Al-Attar, 2010; Al-Nashash et al., 2009; Christofourou and Yigit, 2008). It seems to me that they are more like Continuous Quality Improvement than the continuous assessment of learning.

Because the basic parameters of reliability and validity apply in every system of assessment, the propensity for one system to learn from another and vice versa is considerable.

1.9 Rethinking Validity

The application of the coursework assessment mark to the final mark gave rise to an important debate about the value of the single mark (assumption 2 of Exhibit 1.5) and some academics suggested that a candidate's performance should be recorded in a profile form. For example, practical abilities in science and technology should be separated from the theoretical. There began a move to find ways of measuring these complementary abilities. The lack of clarity about what the purposes of coursework were was a factor that encouraged the objectives movement to promote its wares in the United Kingdom. At the same time it raised serious questions about the validity of what was being done. Much of what went on seemed to rely on face validity—"if it looks right, it is right" but as Exhibit 1.7 shows validity is a much more complex concept than that (Wigdor and Garner, 1982) and measures are made of content, predictive, criterion, and construct validity. Content validity is the extent to which a test measures the content or skill that it is supposed to measure. Criterion (or concurrent) validity is closely related to predictive

Face validity: The extent to which an assessment appears to be measuring the variable it is intended to test, for example, visual inspection of the items or questions in comparison with declared objectives; do they, for instance, measure analysis?)
Content validity: The extent to which a test measures the content (or skill) which it is supposed to measure.
Predictive validity: The extent to which an assessment predicts future performance (e.g., degree grades as a predictor of work performance). Criterion validity is similar.
Criterion validity: Comparison of an assessment designed to evaluate performance in a task with an alternative evaluation (e.g., a test designed to predict driver performance compared with actual observations by a skilled judge). It predicts what an individual or group with particular scores on a test will perform on the criterion measure that will have been chosen to be close to the issue of interest. Cronbach (1971) calls this concurrent validity (see also Kline, 1993).
Construct validity: The extent to which an assessment measures the content (aptitude, attitude, and skill) it is intended to assess, and predicts results on other measures of content, aptitude, attitude, and skill as hypothesized. Wigdor and Garner (1982) write that it is a "scientific dialogue about the degree to which an inference that a test measures an underlying trait or hypothesized construct is supported by logical and scientific analysis" (see also Kline, 1993). "Intelligence" is one such construct.

EXHIBIT 1.7. Traditional measures of validity.

<i>Ideological validity</i> (Ridgeway and Passey, 1992). Refers to the educational moral, philosophical, and political values that are implied by the use of any particular assessment scheme. Related to this is <i>Stick and Carrot validity</i> , which assesses the extent to which an assessment system can be used to control the education system.
<i>Generative validity</i> . Refers to changes in behavior that occur because of a particular set of measures. The way in which particular measures influence the direction of the curriculum and teaching (Ridgeway and Passey, 1992).
<i>Tentative generative validity</i> . Identifies likely directions of change and their inherent value (Ridgeway and Passey, 1992).
<i>The Corruption coefficient</i> . Measures the extent to which scores can be manipulated without changing what is actually being measured (e.g., raising coursework scores without benefiting a student's understanding) (Ridgeway and Passey, 1992).
<i>Experiential validity of the curriculum</i> (Mentkowski, 2000). The student's observation that he or she has grown in some valued ability (or abilities) through his or her learning experiences.

EXHIBIT 1.8. Other concepts of validity resulting from the rethinking of validity in the 1990s.

validity, since it is the comparison of an assessment designed to evaluate performance in a task with an alternative evaluation (see Section 3.2; Cronbach, 1971; Kline, 1993). Construct validity is the extent to which an assessment measures the content (aptitude, aptitude, and skill) it is intended to assess and predicts results on other measures of content, aptitude, and skill as hypothesized. Wigdor and Garner (1982) write that it "is scientific dialogue about the degree to which an inference that a test measures an underlying trait or hypothesised construct is supported by logical and scientific analysis."

Gipps (1994) has pointed out that the definition of validity has been taken beyond that of measuring what a test measures to become an assessment of the evidence available to support test interpretation, and the potential consequences of test use. This indicates that validity is not a value neutral concept as is demonstrated by Ridgeway and Passey (1992; see Exhibit 1.8) which takes us into the realm of personal, institutional, and societal goals, and in terms of the quality assessment of higher education, the process itself.

The rethinking of the concept of validity in the 1990s challenges some of the assumptions listed in Exhibit 1.6 not least the view that student abilities can be assessed and reported unidimensionally (Rogers, 1994). It has an important bearing on the development of competency-based assessment.

1.10 Wastage (Dropout): The Predictive Value of School Examinations for Satisfactory Performance in Higher Education

As previously indicated, in the 1960s in both the United Kingdom and the United States, there was much concern with the predictive reliability and validity of the entrance tests and examinations used for the selection of students to university programs. This

continues to be the case (e.g., Howell, Sorensen, and Jones, 2014). Interest in selection led the CVCP to develop a British equivalent of the SAT called the Test of Academic Aptitude with the intention of its result being combined with the results of General Certificate of Education (A Level) subjects to give a better prediction of university performance (see Chapter 3, Note 1). In spite of it being relatively successful, little was heard of it in subsequent years. One reason for this was that irrespective of the correlations between the grades of some A level subjects and the final grade awarded by the university being quite small some subjects, particularly the science subjects depended on school subjects to provide the content base for further study at university. This is why degree programs could be completed in 3 years because sixth-form studies, as the last 2 years of schooling were called, were highly specialized and the equivalent of the first year of study in many other university systems, including the United States.² The Engineering Science subject that is one of the subjects of discussion in Chapter 3 was of this kind. At the same time the GCE Advanced level examinations could be regarded as relatively successful when measured against an annual wastage rate from the universities that was constant at between 13 and 14% (University Grants Committee [UGC], 1968). Looked at in this way 85% of the intake completed and passed their examinations, most of them within a 3-year time limit. Nevertheless, looked at either from the perspective of the system or that of the “failed” individual, the costs could be considered to be high, but as Vaizey (1971) showed that there is no easy way of calculating such costs.

It is not surprising, therefore, that much interest should focus on wastage (dropout) (now more commonly considered from the perspective of retention). Investigators sought to understand the reasons for withdrawal from courses and examination failure. They found a complexity of reasons that ranged from getting married to psychological difficulties related to examinations. Numerous studies related to wastage were completed in the 1960s in the United Kingdom and a number of them related to technological studies and by 1971 there was a substantial literature on the topic (UGC, 1968; Miller, 1970; Heywood, 1971).

Considerable variations were found between the subjects. The failure rate in engineering and technology was approximately twice that of any other subject grouping except that of architecture. Between the different institutions, the highest failure rates for academic reasons in engineering and technology were in the ex-Colleges of Advanced Technology that had become universities in 1966.

These ranged from 24 to 40%. The lowest failure rates for academic reasons were at Oxford (2.6%), Cambridge (4.1%), and Birmingham (7%). Of those who persisted to their final examinations, only 1.7% of the candidates failed. Simple inspection of the data suggests that those institutions with the lowest rates were able to select candidates with higher “A” level grades than those institutions with the highest rates. Moreover, many students in those institutions were in 4-year programs.

Important methodological issues were discussed by American (Lavin, 1967) and British investigators (Entwistle, 1970; Kelsall, 1963). Lavin (1967) issued the reminder that a “significant relationship between predictor and criterion does not necessarily establish that the predictor is a causal determinant of the criterion.” Kelsall (cited in Heywood, 1971) pointed out that for “any population of applicants in one year, we can never, in the nature of things, compare the academic performance of those admitted

by the selectors somewhere with that of those not admitted by the selectors anywhere. For all we know, therefore, existing methods may be highly efficient in weeding out those with the poorest chance of success in university studies. We can only judge the overall efficiency of selection at most within the population of those admitted to some university somewhere in Britain.” This is why, at the time, there was interest in cross-institutional studies initiated by Noel Entwistle that focused on persons with similar if not identical entry qualifications to courses designed to be of similar standard in Colleges of Education. Entwistle also noted elsewhere that correlation analyses can be misleading because they produce descriptions of the average successful student. They do not provide a method for comparing every real student with every other real student for the purpose of establishing the differences in test score profiles as between successful and unsuccessful students. At the time the new technique of cluster analysis held out some hope that this might be achievable; Brennan using this technique on Entwistle’s data (Entwistle, 1970) identified different types of science and arts students and different types of successful student. Numerous UK and US studies looked at factors that might contribute to success or failure such as personality, mental health, motivation, and study habits.

Today in the United States, much interest in engineering education has focused on the reasons for transfer to other programs of study. In Seymour and Hewitt’s (1997) seminal work, the focus is on why students do or do not persist. They found a sizeable problem. Among higher than average ability students, there was a loss of between 40 and 60% of science, maths, and engineering students. Moreover, they were not surprised to find that faculty wanted “to marginalize the issue of wastage” given the size of the problem. They argue that to improve “the retention among women and students of colour, and to build their numbers over the longer term is to improve the quality of the learning experience for all students” [...] Though faculty sometimes like to begin a program of reform with discussions of curriculum content and structure, this is unlikely to improve retention unless it is part of a parallel discussion of how to secure maximum student comprehension, application and knowledge transfer, and give students meaningful feedback on their academic performance” (Seymour and Hewitt, p. 394). The design of assessment, formative and summative, is central to such a reappraisal as are the aims and objectives they are supposed to obtain. Investigations into retention continue (e.g., Bernold, Spurlin, and Anson, 2007; Walsden and Foor, 2008).

1.11 Factors Influencing Performance in College Courses

It seems that many factors influence performance in college. In so far as engineering was concerned, Furneaux (1962) described the results of a personality test given to first-year mechanical engineering students at Imperial College. The tests were designed to differentiate between extraversion (unstable and stable) and introversion (unstable and stable). He found that there were significant differences in academic performance between the groups and that those with tendencies toward neurotic introversion tended to do better in examinations than those tending to extraversion (see Section 2.5). By the end of the 1960s, other studies had been completed that arrived at a similar conclusion. Child (1969) investigated the 1966 intake at Bradford University (503 men and 103 women) to

compare personality intelligence and social class. While introversion was a characteristic of these students, they tended to extraversion when compared with norms for university students obtained from the Eysenck Personality Inventory. In what must be one of the earliest references to women scientists in the UK literature, Child found that they were the most extravert of all the groups. He suggested that this might be because extraverts are less susceptible to social conditioning and in consequence were less concerned about what others might think of irregular or unfeminine career choices. His findings about social class did not differ from other studies of technological courses. A large number of students in these courses, particularly in the ex-Colleges of Advanced Technology that had become universities, came from working-class homes. Relating this to Furneaux's study that suggested that neuroticism is a measure of drive level, Child suggested that "if this is the case, students from the homes of semi-skilled and unskilled workers tend to have higher drive levels than other students specialising in the sciences."

Entwistle and Wilson (1970), in a survey of the literature, concluded that while unstable students in mechanical engineering examinations were more successful they were atypical of university students as a whole. As a result of a large-scale study among students in universities, colleges of education and polytechnics they were led to conclude "the possibility of their being distinct differences in the relationship between neuroticism and academic performance that may explain some of the contradictions in previous findings." Some years later, Kline and Lapham (1992) found no differences between engineering students and other students in five British universities in respect of extraversion and emotional stability.

In the United States, Elton and Rose (1974) studied engineering students, using the *Omnibus Personality Inventory*. They found a significant difference between engineering students on the dimension of intellectual disposition. Strangely, an absence of high intellectual interests was found among those who persisted. They suggested that the faculty might consider a second experiment with the objective of developing new avenues of professional competence for the 25% of engineering students who withdrew. In another study of students in residence, they were led to conclude that personality differences among students sharing accommodation could enhance or impede achievement. Notwithstanding criticisms of its psychometric qualities, the Myers Briggs Type Indicator (MBTI) was used by Smith, Irely, and McCauley (1973) to investigate the personal qualities that might influence performance. In a later article, McCauley (1990) argued that people skills were undervalued by engineering educators as measured by the Feeling dimension of the MBTI. The MBTI became a test that was favored by personnel selectors in industry, and there are still references to its use by engineering educators.

Studies continue to report that Feeling is important, for example, among Chinese students (Zhang et al., 2014; see also Heywood, 2005). Recently, a study of 103 Dutch engineers (mean age = 48.4 years) using the Five Factor Personality Inventory³ (Hendricks et al., 1999) found that this group was somewhat more extraverted than the population as a whole, yet, almost paradoxically they were more autonomous and less friendly than ordinary people, which might be problematic in interpersonal relations (van der Molen, Schmidt, and Kruisman, 2007). They would need to learn to be more "agreeable" (agreeable/quarrelsome being one of the five factors). The authors pointed

out that their findings had some similarity with Chinese engineers who were found to be more emotionally stable and conscientious than a comparison group (Dai, 2003).

On both sides of the Atlantic, there was interest in study habit inventories. While there was no reported study of their use with engineering students, marked differences were found between preclinical and clinical students in medicine. Preclinical students tried to memorize work recently covered but clinical students did not. While there were other differences, the investigators did not find any clear relation between these differences and subsequent academic performance (Malleon, Penfold, and Sawiris, 1967). Entwistle and Wilson developed a study habit inventory and administered it together with the Eysenck Personality Inventory and found that among graduates of a Diploma in Education course that the study methods scale distinguished clearly between the worst students and the remainder. The extraversion scale sorted out the successful students from the rest. In another study, Entwistle and Entwistle (1970) concluded that good study habits made a partial contribution to an introvert's success.

The more important work on assessment was to be conducted later in Sweden and Britain, the vocabulary of which has entered into higher education discourse. Marton and Säljö (1984) distinguished between deep and surface approaches to understanding. They concluded that the strategies found, apart from anything else, were indicative of different perceptions of what is wanted from learning. From the perspective of assessment design, there is no guarantee that a student will perceive the demands of assessment in the same way as the tutor. Moreover, the design of assessment and the method of instruction can cause either deep or surface learning.

Entwistle and Ramsden (1983) developed an *Approaches to Study Inventory* that yielded four factors. *Meaning orientation* had high loadings on the deep approach associated with comprehension learning and intrinsic motivation: whereas the *reproducing orientation* was highly loaded on the surface approach, operation learning and improvisation were associated with fear of failure and extrinsic motivation. A nonacademic factor *nonacademic orientation* related to disorganized approaches to study, and *achieving orientation* to strategic approaches to study. Previously Entwistle, Hanley, and Ratcliffe (1979) had identified a *strategic approach* to learning in which students try to "manipulate the assessment procedures to their own advantage by careful marrying of their efforts to the reward system as they perceive it."

Twenty years earlier, Nevitt Sanford (1962), in the United States, brought together a number of authorities to give psychological and social interpretations of higher learning. There were sufficient authors to create a substantial volume. Among them was Christian Bey (1962), who, in a chapter titled "A Social Theory of Intellectual Development," suggested three types of student relationships with assessment. First, there were those who were academically oriented and who worked hard to obtain academic rewards. They would seem to have some affinity with the "strategic learners" identified by Entwistle. Presumably they would vary their learning from deep to surface as a function of their perception of what was required. Second, there were those who were intellectually oriented: it would seem that they would be likely to be deep learners: and third, there were those who were socially oriented. Twenty or more years later, Taylor distinguished between four orientations that she called, academic, vocational, social, and personal that is somewhat different. She distinguished between intrinsic and extrinsic persons

in each category. Thus, an intrinsic person with a personal orientation is concerned with broadening or self-improvement, whereas an extrinsically motivated person in this category would look for proof of capability (Gibbs, Morgan, and Taylor, 1984). The point here is not the orientations of students but the fact that these orientations pointed toward the need for a social theory of intellectual development. In 1967, Sanford published a study with the title *Where College fails. A Study of the Student as Person*. These studies pointed to the need for assessment to take into account the “person.”

These limited remarks are intended to show that consideration of assessment is very much more than setting examinations and tests. Satisfactory performance is dependent on a whole mix of factors that, if adjusted, may not necessarily solve problems of performance. The factors influencing student performance extend to the organization of the institution and its values as expressed through its faculty, as well as myriad aspects of a student’s life [style] (Berliner, 2014). Any discussion of assessment has to take them into account, as is done, for example, in Chapter 7.

1.12 Assessment: Results and Accountability

One of the few parameters that can be used to measure the performance of an academic institution is examination (test) results. Politicians love league tables that are based on results and they often grievously misinterpret the data. In both the United States and the United Kingdom, there have been concerns about grade inflation. That is, substantial increases in the number of top and near top grades awarded. Very often politicians take this to mean that standards have fallen. But, in all probability the same politicians have said that standards need to be raised 3 or 4 years earlier. Universities are damned if they do increase the number of better grades and damned if they do not. This contradiction highlights the need for data that are collected, such as examination scripts, to be retained for a number of years so that past and present standards can be evaluated.

The point here is not to debate the accountability issue but rather to point out that in the 1960s a positivist philosophy came into play and higher education came to be viewed as an instrument for producing people to do particular jobs—engineers to do engineering. In Britain, there was, for example, a National Board for Prices and Incomes that took a particular interest in the productivity of universities or at least how to measure it. The Board wrote “We have received very little evidence of experimentation with new methods of teaching which would economise on the number of staff required. It is possible that the nature of control exercised from the outside has made it difficult for a university to experiment with different combinations of capital or labour, or with different combinations of staff which might lead to an improvement in teaching methods, helping it to determine the nature of the teaching staff required” (cited by Heywood, 1971).

It will be noticed that for the Board improvement in teaching is equated to reducing the number of teachers not with possible improvements in teaching or student learning *per se*. It will also be noticed that there is an expectation that universities should prepare students for their careers and select teaching staff to that end. Although the Board had a short life, subsequent governments listened to industry and influenced the agencies responsible for higher education to cause them to create corporate cultures

within universities. At the same time it reflects a utilitarian view of what the aims of education should be. Not surprisingly, no official report on higher education that followed contained an adequate discussion of the aims of higher education. In such a climate, performance (competency)-based assessment would have an appeal. It is not surprising, therefore, that during this period, all across the Western world, without any specific announcement the universities began to reverse their aim to one of the primacy of research over the primacy of teaching. The performance of faculty came to be measured by the number of papers published in elite journals, and a conflict between research and teaching emerged to the detriment of teaching. This change also marked a change in the complex relationships between academia and industry.

1.13 Assessing the Learner

Assessment was and remains for many the afterthought of the educational process—a necessary evil. In the formative years of the study of higher education, the power of assessment to influence student learning and teaching was shown. Assessment was shown to be a major component of the curriculum process. Studies of assessment could not be confined simply to the measurement of performance in examinations since many factors, including several in the affective domain, contributed to that performance. Present-day thinking about assessment has to be judged by the extent it has become part of the curriculum process and taken into account all these other factors especially those from the affective domain that contribute to performance. The way in which engineering schools take into account the “person” is as important to the understanding of the impact of assessment on learning as anything else.

The curriculum process should examine the aims of education on which that curriculum is based in detail. The evidence collected suggests that very little has changed.

So far I have failed to define assessment. Georgine Loacker to whom this book is dedicated and I disagreed about many things but we did not disagree about the principles of assessment. It is appropriate, therefore, to take the definition of assessment from a book written under her direction in 1985 for the students of Alverno College. It reads “assessment: a multidimensional attempt to observe and, on the basis of criteria, to judge the individual learner in action” (Alverno College, 1985). In the case of students studying for a profession (e.g., engineering) that means, judging them in both academic and workplace (industrial training) situations. “To undertake assessment means sampling, observing, and judging according to criteria. It means developing a whole array of techniques to take into account the fullest possible range of human talents. And it means an ongoing commitment to dealing with these kinds of questions” (Alverno College, 1985). This book is about how the engineering community has and is answering these questions and how it has and is determining the criteria on which such judgments are made.

E. J. Furst (1951), one of the authors of *The Taxonomy of Educational Objectives* (see Section 2.5) and an expert in evaluation, wrote that every teacher should have a defensible theory of learning that in this writer’s view would derive from a defensible philosophy of education. The intention of this book is to provide its readers with information about theory and practice that will enable them to develop their own theory or philosophy of

assessment. This chapter has begun that process by describing the different meanings that have been given to the terms associated with assessment and in particular those of competence, outcomes, and performance. By definition, a professional education implies that much of it is about preparing the student to be a professional, which in turn implies that the curriculum is to some extent a reflection of what engineers do. Surprisingly, engineering education has been regularly criticized during the last 70 years for not knowing what it is that engineers do. The chapter that follows opens a discussion on the relationship between academia and industry.

Notes

1. Forms of norm-referenced grading are in use in American engineering courses and Seymour and Hewitt's (1997) seminal study on retention in STEM courses thought that norm-referenced grading fostered a competitive atmosphere that alienated otherwise capable students. Wolf and Powell (2014) reported that *left-of-center* grading was more likely to discourage females than peers. The students reported that it caused stress and frustration and focused their attention on the system than on their own learning. In the United Kingdom, it has been the system and students were brought up to it from elementary school onward. But the introduction of competency-based systems of assessment has meant that school examining authorities give marks up to 100%. At school level, politicians to the right have taken against continuous assessment and modular courses and there is a move back to more traditional examining.
2. In England, there is no such thing as a liberal arts degree. Students will take mostly a single subject like history or engineering for 3 years; this would include any requisite studies. For example, engineering would require mathematics. Today engineering has been extended to a fourth year at the end of which a master's degree is awarded. Professional recognition as a chartered engineer now requires applicants to have a master's degree.
3. The Five-Factor Personality Inventory is also known as the Big Five and is proving attractive to engineering educators. The five domains are extraversion, agreeableness, conscientiousness, emotional stability, and openness. It is considered to be a better test than the MBTI. See Rhee, Parent, and Oyamont (1992) for its use in engineering. They cite a meta-analysis by Ozer and Bet-Martinez (2008) in support of its use.

References

- Abercrombie, M. L. J. (1964). *Aims and Techniques of Group Teaching*. London: Society for Research into Higher Education.
- Abu-Idayil, B., and Al-Attar, H. (2010). Curriculum assessment as a direct tool in ABET outcomes assessment in chemical engineering. *European Journal of Engineering Education*, 35(5), 489–505.
- Ager, M., and Weltman, J. (1967). The present structure of examinations. *Universities Quarterly*, 21(3), 272.
- Allen, T., Reed-Rhoads, T., Murphy, T. J., and Stone, A. D. (2008). Coefficient Alpha: an engineers interpretation of reliability. *Journal of Engineering Education*, 97(1), 87–94.
- Al-Nashash, H., Khaliq, A., Qaddoumi, N., Al-Assaf, Y., Assalah, K., Dhaouadi, R., and El-Tarhuni, M. (2009). Improving electrical engineering education at the American University of Sharjah through continuous assessment. *European Journal of Engineering Education*, 34(1), 15–28.

- Alverno College (1985). *Assessment at Alverno College*. Milwaukee: Alverno Productions.
- Astin, A. (1968). *The College Environment* Washington, DC: American Council on Education.
- Berliner, D. C. (2014). Exogenous variables and value-added assessment: a fatal flaw. *Teachers' College Record*, 116(10). <http://www.terecord.org>. Accessed March 24, 2014.
- Bernold, L. E., Spurlin, J. E., and Anson, C. M. (2007). Understanding our students. A longitudinal study of success and failure in engineering with implications for increased retention. *Journal of Engineering Education*, 96(3), 263–279.
- Bey, C. (1962). A social theory of intellectual development. In: N. Sanford (ed.), *The American College*. New York: John Wiley & Sons.
- Bligh, D. G. (1971). *What's the Use of Lectures*. London: Penguin.
- Bosworth, G. S. (1963). Toward creative activity in engineering. *Universities Quarterly*, 17, 286.
- Bosworth, G. S. (Chairman of a Committee) (1966). *The Education and Training Requirements for the Electrical and Mechanical Manufacturing Industries*. Committee on Manpower Resources for Science and Technology. London: HMSO.
- Bucciarelli, L. L., Coyle, E., and McGrath, D. (2009). Engineering education in the US and UK. In: S. H. Christensen, et al. (eds.), *Engineering in Context*. Aarhus: Academica.
- Carter, G. (1992). The fall and rise of university engineering education in England in the 1980s. *International Journal of Technology and Design Education*, 2(3), 2–21.
- Carter, G., and Heywood, J. (1992). The value-added performance of electrical engineering students in a British university. *International Journal of Technology and Design Education*, 2(1), 4–15.
- Child, D. (1969). A comparative study of personality, intelligence and social class in a technological university. *British Journal of Educational Psychology*, 39, 40
- Christofourou, A. F., and Yigit, A. S. (2008). Improving teaching and learning in engineering education through a continuous assessment process. *European Journal of Engineering Education*, 33(1), 105–116.
- Cohen, L. (1968). Attitude deterioration and the experience of industry. *The Vocational Aspect of Education*, 20, 181.
- Cole, J. S., and Spence, S. W. T. (2012). Using continuous assessment to promote student engagement in a large class. *European Journal of Engineering Education*, 37(5), 508–525.
- Cox, R. (1967). Examinations and higher education. A review of the literature. *Universities Quarterly*, 21(3), 292.
- Cowan, J. (2006). *On Becoming an Innovative University Teacher*, 2nd edition. Buckingham: Open University Press.
- Cronbach, L. J. (1971) Test validation In: R. L. Thorndike (ed.), *Educational Measurement*, 2nd edition. Washington, DC: American Council on Education.
- Committee of Vice-Chancellors and Principals. (1969). *The Assessment of Academic Performance*. London: Committee of Vice-Chancellors and Principals.
- Dai, X. (2003). A study of the occupational characters of different occupational groups. *Chinese Journal of Clinical Psychology*, 10, 252–255.
- Dressel, P. L. (ed.) (1961). *Evaluation in Higher Education*. Boston, MA: Houghton Mifflin.
- Dwyer, C. A (1994). *Development of the Knowledge Base for the PRAXIS III: Classroom Performance Assessments. Assessment Criteria*. Princeton, NJ: Educational Testing Service.
- Educational Testing Service. (1995). *Teacher Performance Assessments. Assessment Criteria*. Princeton, NJ: Educational Testing Service.

REFERENCES

29

- Elton, C. F., and Rose, H. A. (1966). Within university transfer. Its relation to personality characteristics. *Journal of Applied Psychology*, 50(6), 539.
- Entwistle, N. J. (1970). Students and their academic performance in different types of institution. Part of a report to the Joseph Rowntree memorial trust cited in Heywood (1971).
- Entwistle, N. J., and Entwistle, D. (1970). The relationship between personality, study methods and academic performance. *British Journal of Educational Psychology*, 40, 132–143.
- Entwistle, N. J., Hanley, M., and Ratcliffe, G. (1979). Approaches to learning and levels of understanding. *British Educational Research Journal*, 5, 99–114.
- Entwistle, N. J., and Wilson, J. D. (1970). Personality, study methods, and academic performance. *Universities Quarterly*, 24(2), 147–156.
- Epstein, R. M., and Hundert, E. M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association*, 287(2), 226–235.
- Felder, R. M., and Hadgraft, R. G. (2013). Educational practice and educational research in engineering: partners, antagonists, or ships passing in the Night. *Journal of Engineering Education*, 102(3), 339–345.
- Fordyce, D. (1992). The nature of student learning in engineering education. *International Journal of Technology and Design Education*, 2(3), 23–40.
- Freeman, J., and Byrne, P. S. (1973). *The Assessment of Post-Graduate Training in General Practice*. Guildford, UK: Society for Research into Higher Education.
- Furneaux, W. D. (1962). The psychologist and the university. *Universities Quarterly*, 17, 33–47.
- Furst, E. J. (1958). *The Construction of Evaluation Instruments*. New York: David MacKay.
- George, J., and Cowan, J. (1999). *A Handbook of Techniques for Formative Evaluation: Mapping the Students Learning Experience*. London: Kogan Page.
- Gibbs, G., Morgan, A., and Taylor, E. (1984). The world of the learner. In: F. Marton, D. Hounsell, and D. Entwistle (eds.), *The Experience of Learning*. Edinburgh: Scottish Academic Press.
- Gipps, C. V. (1994). *Beyond Testing. Toward a Theory of Educational Assessment*. London: Falmer.
- Griffin, C. (2012). A longitudinal study of portfolio assessment to assess competence of undergraduate student nurses. Doctoral Dissertation, University of Dublin, Dublin.
- Hartle, T. W. (1986). The growing interest in measuring the education achievement of college students. In: C. Adelman (ed.), *Assessment in Higher Education*. Washington, DC: US Department of Education.
- Hartog, P., and Rhodes, E. C. (1935). *An Examination of Examinations*. London: Macmillan.
- Hartog, P., and Rhodes, E. C. (1936). *The Marks of Examiners*. London: Macmillan.
- Hendricks, A. A. J., Hofstee, W. K. B., and de Raad, B. (1999). The Five-Factor Personality Inventory (FFPI). *Personality and Individual Differences*, 27, 307–325.
- Heywood, J. (1971). A report on wastage. *Universities Quarterly*, 25(2), 189–237.
- Heywood, J. (1977). *Assessment in Higher Education*. Chichester, UK: John Wiley & Sons.
- Heywood, J. (1989). *Assessment in Higher Education*, 2nd edition. Chichester, UK: John Wiley & Sons.
- Heywood, J. (2000). *Assessment in Higher Education. Student Learning, Teaching, Programmes and Institutions*. London: Jessica Kingsley.
- Heywood, J. (2005). *Engineering Education. Research and Development in Curriculum and Instruction*. Hoboken, NJ: IEEE/John Wiley & Sons.

- Heywood, J. (2011). Higher technological education in England and Wales 1955–1966. Compulsory liberal studies. In: Proceedings of Annual Conference of the American Society for Engineering Education. Paper 635.
- Heywood, J. (2013). Higher technological education and British policy making: a lost opportunity for curriculum change in engineering education. In: Proceedings of Annual Conference of the American Society for Engineering, June 2013. Paper 6331.
- Heywood, J., and Cheville, A. (2015). Is engineering education a professional activity? In: Proceedings of Annual Conference of the American Society for Engineering Education, June 2015. Paper 12907.
- Heywood, J., and Ann Abel, R. (1965). *Technical Education and Training. Research in Progress. 1962–1964*. NFER Occasional Publication, No 8. Slough, National Foundation for Educational Research.
- Howell, L. L., Sorensen, C. D., and Jones, M. R. (2014). Are undergraduate GPA and General GRE percentiles valid predictors of student performance in an engineering graduate program? *International Journal of Engineering Education*, 30(5), 1145–1165.
- Hoyle, E. (2001). Teaching, prestige, status and esteem. *Educational Management and Administration* 29(2), 139–152.
- Hutchings, D. G. (1963). *Technology and the Sixth Form Boy*. Oxford, UK: Oxford University Department of Education.
- Kelsall, R. K. (1963). University student selection in relation to subsequent academic performance. *The Sociological Review*. Monograph No 7.
- Kline, P. (1993). *The Handbook of Psychological Testing*. London: Routledge.
- Kline, P., and Lapham, S. (1992). Personality and faculty in British universities. *Personality and Individual Differences*, 13, 855–857.
- Kowalski, S. E., Kowalski, F. V., and Gardner, T. Q. (2009). Lessons learned when gathering real-time formative assessment in the university classroom using Tablet PC's. In: ASEE/IEEE Proceedings of Frontiers in Education Conference, T3F-1 to 5.
- Larkin, T. L., and Usinski, J. (2013). The evolution of curriculum assessment within the physics program at American University. In: Proceedings of Annual Conference of the American Society for Engineering, June 2013. Education Paper 6739.
- Lavin, D. E. (1967). *The Prediction of Academic Performance* New York: John Wiley & Sons.
- Lorimer, S., and Davis, J. A. (2014). Consistency in assessment of pre-engineering skills. In: Proceedings of Annual Conference of the American Society for Engineering, June 2014. Paper 9596.
- Marton, F., and Saljö, R. (1984). Approaches to learning. In: F. Marton, D. Hounsell, and N. J. Entwistle (eds.), *The Experience of Learning*. Edinburgh: Scottish Academic Press.
- Malleson, N. (1964). *British Student Health Services*. London: Pitman.
- Malleson, N., Penfold, D. M., and Sawiris, Y. (1967) Medical students' study. Time and place. *British Journal of Medical Education*, 1, 169.
- Mast, T., and Davis, D. A. (1994). Concepts of competence. In: D. A. Davis and R. D. Fox (eds.), *The Physician as Learner. Linking Research to Practice*. Chicago, IL: American Medical Association.
- McCaughey, M. M. (1990). The MBTI and individual pathways in engineering design. *Engineering Education*, 66(7), 729–716.
- McGuire, C. (1993). Perspectives in assessment. In: J. S. Gonnella et al. (eds.), *Assessment Measures in Medical School Residency and Practice*. New York: Springer.

REFERENCES

31

- McKeachie, W. J. (1974). Instructional psychology. *Annual Review of Psychology*, 25, 163–193.
- McVey, J. (1975). The errors in marking examination scripts in electronic engineering. *International Journal of Electrical Engineering Education*, 12(3), 203.
- McVey (1976a). Standard error of a mark for an examination paper in electronic engineering. *Proceedings Institution of Electrical Engineers*, 123(8) 843–844.
- McVey, P. (1976b). The “paper error” of two examinations in electronic engineering. *Physics Education*, 11, 58–60.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65, 563–567.
- Miller, G. W. (1970). *Success, Failure and Wastage in Higher Education*. London: Harrap.
- Myllymaki, S. (2013). Incorporation of continuous student assessment into lectures in engineering education. *European Journal of Engineering Education*, 38(4), 385–393.
- National Union of Students. (1967). *Briefing on Wastage*. London: National Union of Students.
- National Union of Students. (1968). *Executive Report on Examinations*. London: National Union of Students.
- Nghe, N. T., Janacek, P., and Haddaway, P. (2007). A comparative analysis of techniques predicting academic performance. In: ASEE/IEEE Proceedings of Frontiers in Education Conference T2G-7 to 12.
- Norman, G. R. (1985). Defining competence. A methodological review. In: V. R. Neufeld and G. R. Norman (eds.), *Assessing Clinical Competence*. New York: Springer.
- Oppenheim, A. M., Jahoda, M., and James, R. L. (1967). Assumptions underlying the use of university examinations. *Universities Quarterly*, 1(2), 89.
- Ozer, D. J., and Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, 57, 401–402.
- Parsons, D. (2008). Is there an alternative to exams? Examination stress in engineering. *International Journal of Engineering Education*, 24(6), 1111–1118.
- Pellegrino, J. W., DiBello, L. V., and Brophy, S. P. (2014). The science and design of assessment in engineering education In: A. Johri and B. M. Olds (eds.), *Cambridge Handbook of Engineering Education Research*. New York, NY: Cambridge University Press.
- Reeves, D. B. (2009). *Assessing Educational Leaders. Evaluating Performance for Improved Individual and Organizational Results* Thousand Oaks, CA: Corwin.
- Rethans, J. J., Norcini, J. J., Barón-Maldonado, M., Blackmore, D., Jolly, B. C., LaDuca, T., Lew, S., Page, G. G., and Southgate, L. H. (2002). The relationship between competence and performance: implications for assessing practice performance. *Medical Education*, 36(10), 901–909.
- Rhee, J., Parent, D., and Oyamoto, C. (2012). Influence of personality on senior project combining innovation and entrepreneurship. *International Journal of Engineering Education*, 28(2), 302–309.
- Ridgeway, J., and Passey, D. (1992). An international view of mathematics assessment: through the glass darkly. In: M. Niss (ed.), *Investigations into Assessment in Mathematics Education*. Dordrecht: Kluwer.
- Roberts, R. F. (1982). *Preparation for Teaching Practice*. Liverpool, City of Liverpool College of Higher Education cited in Heywood, J. (1982). *Pitfalls and Planning in Student Teaching*. London: Kogan Page.
- Rogers, G. (1994). Measurement and judgment in curriculum assessment systems. *Assessment Update*, 6(1), 6–7.

- Ryle, A. (1969). *Student Casualties* London: Allen Lane.
- Ryle, A., and Lunghi, M. (1968). A psychometric study of academic difficulty and psychiatric stress in students. *British Journal of Psychiatry*, 114(506), 57.
- Runté, R. (1995). Is teaching a profession. In: G. Taylor and R. Runté (eds.), *Thinking about Teaching. An Introduction* Toronto: Harcourt Brace.
- Sandberg, J. (2000). Understanding human competence at work. An interpretive approach. *Academy of Management Journal*, 43(3), 9–25.
- Sadler, D. R. (2007). Perils in the meticulous specification of goals and assessment criteria. *Assessment in Education: Principles, Policy and Practice*, 14(3), 387–392.
- Sanford, N. (ed.) (1962). *The American College. A Psychological and Social Interpretation of the Higher Learning*. New York: John Wiley & Sons.
- Saupé, J. (1961). Learning. In: P. Dressel (ed.), *Evaluation in Higher Education*. Boston, MA: Houghton Mifflin.
- Schalk, P. D., Wick, D. P., Turner, P. R., and Ramsdell, M. W. (2011). Predictive assessment of student performance for early strategic guidance. In: ASEE/IEEE Proceedings of Frontiers in Education Conference, S2H-1 to 5.
- Seymour, E., and Hewitt, N. M. (1997). *Talking about Leaving. Why Undergraduates Leave the Sciences*. Boulder, CO: Westview Press.
- Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.
- Smith, A. B., Irely, R. K., and McCauley, M. H. (1973). Self-paced instruction and college students' personalities. *Engineering Education*, 63(6), 435–450.
- Smithers, A. G. (1965). Occupational aspirations and expectations of engineering students. *British Journal of Industrial Relations*, 7, 415.
- Sternberg, R. J. (2005). Intelligence, competence and expertise In: A. J. Elliot and C. S. Dweck (eds.), *Handbook of Competence and Motivation*. New York: Guilford Press.
- Sundarajan, S. (2014). A strategy for sustainable student outcomes assessment for mechanical engineering program that maximizes faculty engagement. In: Proceedings of Annual Conference of the American Society for Engineering Education, June 2014. Paper 9251.
- Talbot, N. (2004). "Monkey see, Monkey do." A critique of the competency model in graduate medical education. *Medical Education*, 36(6), 587–592.
- The Times*. (2015). Exam results just don't add up, says accountancy firm; August 3. Authored by Andrew Ellson.
- Trevelyan, J. (2014). *The Making of an Expert Engineer*. Leiden, CRC Press.
- University Grants Committee (1968). *An Enquiry into Student Progress*. London: University Grants Committee.
- van der Molen, H. T., Schmidt, H. G., and Kruisman, G. (2007). Personality characteristics of engineers. *European Journal of Engineering Education*, 32(5), 495–501.
- Vaizey, J. (1970). The costs of Wastage. *Universities Quarterly*, 25(2), 139–145.
- Vernon, P. (1965). *The Measurement of Human Abilities*. London: University of London Press.
- Walsden, S. E., and Foor, C. E. (2008). What's to keep you from dropping out? Student immigration into and within engineering. *Journal of Engineering Education*, 97(2), 191–206.
- Warren-Piper, D. (1994). *Are Professors Professional. The Organization of University Examinations*. London: Jessica Kingsley.

REFERENCES

33

- Wigdor, A. K., and Garner, W. R. (eds.) (1982). *Ability Testing: Uses, Consequences and Controversies* Washington, DC: National Academies Press.
- Williams, B., and Figueiredo, J. (2013). Finding workable solutions: Portuguese engineering experience. In: B. Williams, J. Figueiredo, and J. P. Trevelyan (eds.), *Engineering Practice in a Global Context. Understanding the Technical and the Social*. London: CRC Press/Taylor and Francis.
- Yeung, A. T. (2014). Education and training requirement for a practicing civil engineer to qualify in Australia and the United Kingdom. In: Proceedings of Annual Conference of the American Society for Engineering Education, June 29-14. Paper 8227.
- Yokomoto, C. F., and Bostwick, W. D. (1999). Modelling the process of writing measurable outcomes for EC 2000. In: ASEE/IEEE Proceedings of Frontiers in Education Conference, 2, 11b-18-22.
- Young, M. (2002). *Clarifying Competency and Competence*. Henley On Thames. Henley Management College.
- Zhang, D., Yao, D., Cuthbert, N., L., and Ketteridge, S. (2014) A suggested strategy for team-work teaching in undergraduate engineering programs particularly in China. In: ASEE/IEEE Proceedings of Frontiers in Education Conference, pp. 537–544.

