D. ROYCE SADLER

MAKING COMPETENT JUDGMENTS OF COMPETENCE

INTRODUCTION

Comprehensive English dictionaries list multiple meanings for the words "competence" and "competency". Although the variety of meanings may not matter in ordinary conversations, in rigorous thinking about the measurement and development of competence or competencies, clarity is indispensable. For the purpose of developing the theme in this chapter, a distinction is made between what may be conceptualized as an integrated and large-scale characteristic, capability or attribute, and smaller-scale identifiable elements that contribute to such an attribute, in particular demonstrable skills in performing a task. The first of these, the envelope term, is referred to as *competence*; a contributing element is referred to as a *skill* or *competency*, the latter two being used more or less interchangeably. (Elsewhere, competencies may be called *competences*, and *skill* may be restricted to physical or psychomotor activity.)

The distinction in principle between competence and a skill/competency is convenient but at least partly a matter of degree. Thus mastery of a sufficiently large or complex "skill" may be referred to as "competence in (a particular field)." The nature of the distinction depends on the context and the communicative purpose to be served, and to that extent is arbitrary. Notwithstanding those differences, a competent professional (such as an engineer, dentist or accountant) is characterized by competence in the corresponding field; when professional competence is put into practice, numerous skills or competencies are ordinarily involved. An underlying question is whether competence can be exhaustively decomposed into identifiable constitutive skills, or whether it involves something more than applying a set of separate skills which have been acquired or mastered.

Higher education is heavily involved in the development of both particular competencies and overall competence. Interest in these concepts has increased dramatically in Western countries over recent decades. Many employers along with academics who teach advanced programs have expressed disquiet (or even dismay) about the perceived shortcomings of new graduates' general competencies. Whereas previously it could have been taken for granted that these competencies were developed during degree studies regardless of discipline, field or profession, it is currently alleged that this is no longer the case. Responses to these concerns by higher education institutions and quality assurance agencies have included: the identification of general attributes and skills that are important in contexts after graduation, being potentially transferable from academic degree studies to

S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Eds.), Modeling and Measuring Competencies in Higher Education: Tasks and Challenges, 13–27.

workplaces, to advanced studies, across career sequences and to life in general; the development of sound ways to assess such "graduate competencies"; and the design of strategies to improve student performance on them.

Taking as a given that the concerns are justified, what changes in higher education over the same time period may account for them? The many factors are no doubt interrelated, but only three are identified here, the third being elaborated later in the chapter. First, access to higher education has been progressively opened up from an academically elite segment of the population to a significant proportion of the population (the so-called massification of higher education). A result of this has been that at the point of entry many students are now regarded as being inadequately prepared for academic study. Second, the costs of higher education have generally risen and public financial support has generally either fallen or not kept pace in real terms, forcing institutions to economize (one way of cutting teaching costs being to rely progressively and more heavily on part-time academic teachers). The third has to do with changes in teaching and assessment, the aspect of specific relevance to this chapter.

Not surprisingly, institutional lists of intended graduate capabilities show significant overlap. Common elements include student proficiency in: analytical and critical analysis; problem-solving; locating, evaluating and using relevant information; originality, initiative and creativity; and effective communication. This particular selection has a strong emphasis on cognitive outcomes and these are the ones focused on in this chapter, but institutional lists are typically more expansive. Although interest in these types of competencies has been international, the broad movement does not share a standard terminology. Most lists have been framed under headings which are either "graduate" or "generic" and paired with one of the following: attributes, competencies, capabilities, outcomes or skills.

That said, some institutions have two lists, one labeled "generic skills" for specific competencies of the type listed above; and those labeled "graduate attributes" for large-scale student characteristics related to professional outlook and orientation such as: interdisciplinarity; collaboration and teamwork; high ethical standards; a globalized or internationalist perspective; cultural and linguistic sensitivity; social and civic responsibility; lifelong learning; and commitment to sustainability.

In recent years, significant support has been given to the principle of modeling and measuring competencies by broad-spectrum testing of all graduates in a given jurisdiction, preferably by standardized means. The collection of competency measurements is intended to represent levels of graduate competence. In some contexts, differentiation in the content of tests has been proposed as a means of achieving a satisfactory fit for various areas of specialization, something more difficult to achieve with a single omnibus test for all students. Despite those initiatives, the broad interest remains in measuring competencies which characterize graduates irrespective of the particular courses, programs or institutions in which students enroll. Mass testing of graduate competencies is proposed as a way of enabling trends in teaching effectiveness to be identified,

comparisons across institutions or systems to be made, and quality assurance procedures to be more objective and driven by results.

An additional line of thinking is that if performances in tests of graduate competencies are publicized in the form of institutional rankings, this could incentivize poorly ranking academic programs or entire institutions to redirect some of their effort and resources towards improving the performance and employability of their graduates and thus improve their relative standing among similar institutions. A further possibility is that if mass testing were carried out early in an academic program and then again after graduation, gain scores could provide a measure of the value added by participation in higher education as part of the social return on investment. All in all, this initiative has been widely advocated as a logical, direct, efficient and politically feasible approach to the open scrutiny of institutional attainments, the discovery of shortfalls, the implementation of remedial strategies, and the accountability of higher education institutions in terms of playing their full part in national growth, development and prosperity.

Although the importance of the types of cognitive competencies in the sample list above is widely recognized, it does not automatically follow that the most appropriate way forward is to spell out what is to comprise each competency and then implement mass testing programs. In this chapter, the outline of an alternative view is presented. It does not pretend to be a fully argued case or to set out a comprehensive plan for action. The development flows from a number of reservations held by a disparate group of researchers and commentators about: the philosophical legitimacy of decomposing competence as a complex concept into constituent skills-competencies; the uncoupling of various competencies properly expected of study in higher education from regular academic programs and courses; and the prospect that mass testing and its flow-on effects could divert attention and resources away from the primary sites at which competencies should be developed, practiced and refined, these sites being normal academic studies.

In this alternative view, the generic competencies would remain firmly situated within the various disciplinary or professional educational contexts. The final step would be the assessment of these competencies. This would be integrated into holistic judgments of the quality of student work against recognized academic achievement standards which are comparable across courses and academic programs (and, where appropriate, across institutions). Both this goal statement and tentative principles for achieving the goal through systematic peer consensus processes are developed in more detail in four of the author's articles (Sadler, 2009a, 2009b, 2010b, 2011).

DECOMPOSITION

Conceptualizing competence as made up of a number of underlying competencies is an example of a general approach to tackling complex problems and phenomena. Decomposition into constituent parts has proved a powerful tool for probing and developing understanding in many areas of thought and practice. If a complex entity is to be put to practical use, decomposition often makes it possible to devise

methods for testing all the parts separately and then checking that they all function together as they are supposed to. This is well exemplified in mass manufacturing processes. It has also played a significant part in the way technology and the physical and biological sciences have advanced. Parts have been identified, relationships and dependencies explored, theorizing and hypothesis testing carried out, and predictive models developed so that theorizations can be tested. Where appropriate, processes have been modeled with a view to monitoring and controlling them so they can serve human needs.

At this point, a short digression shifts the focus to an adjacent field of education. Decomposition has been a common feature in post-compulsory education, particularly in the vocational and training sectors of countries such as Australia and the United Kingdom. Complex outcomes have been broken down into smaller and smaller skills or competencies, which have then been taught, practiced, tested and checked off a master list when "achieved." The competencies themselves are typically identified through consultation with representatives of trades, crafts, arts, industry and labor unions in a bid to insure they are empirically based and the full set is as complete as possible. Under this model, attainment of all competencies leads to accreditation as a qualified tradesperson or practitioner. One of the claimed instructional advantages of describing multiple competencies in detail is that the competency descriptors provide highly visible targets for instructors and students alike, increasing the likelihood they will be reached and then counted towards a qualification when achieved. This system therefore sounds rational and straightforward. Furthermore, it can produce competent practitioners provided it is accompanied by overt attention to the development of strategies for choosing the most appropriate skills, materials or actions in order to achieve the solution to a given problem.

The case of vocational education and training is instructive for two reasons. The first is that, in practice, the decomposition of vocational capability has been applied to a particular class of skills or procedures which are distinctively different from the higher education skill-competencies focused on in this chapter (critical analysis and so on). Many of the skills common in vocational and technical education and training are of the physical, practical, concrete kind. Similar types of skills are often applied in a range of settings, each "skill" being identified both by the "object" to which it is applied and the intrinsic nature of the skill itself. Not uncommonly, skills are described in terms of both criteria, making them distinguishable in concept and in practice. The contexts in which they are learned have a certain degree of routinization or repetitiveness about them, allowing the skills to be rehearsed and mastered separately. For these reasons, it makes sense to treat them, at least in the context of initial training, as distinct skills.

The vocational education context is instructive for another reason. Decomposition into constituent skills can lend itself to seriously deficient implementation, as has become evident in the United Kingdom. (The November 2007 issue of the journal *Assessment in Education: Principles, Policy and Practice* contains a number of reports of research into the UK experience.) The troublesome aspect has been that, for some qualifications, the competencies have been so finely

grained and the assessments so compartmentalized that teachers have moved towards deliberately coaching their students over the pass line for each competency, one by one, in order to enable them to gain a marketable qualification. In extreme instances, this practice has resulted in a particular competency exercise being completed by the student just once, with constant prompting by the instructor. This in turn has been openly defended as both appropriate and necessary for scaffolding student learning. With scaffolding as the rationale, the skill has been checked off and elevated to the status of an acquired competency. No doubt this practice is not what the curriculum developers intended but it does illustrate how component-based assessment practices can undermine progress towards the goal of overall competence. The collection of discrete competencies "passed" does not necessarily translate into a coordinated ability to complete a complex task with proficiency (Sadler, 2007).

Although decomposition of a complex entity may be carried out in order to achieve some gain, this gain is accompanied by loss of a different kind: it becomes more difficult to see the whole as a unified competence. The logic of this phenomenon is obvious. If something is divided into pieces, whatever originally held it together and accounted for its integrity has to be either supplied or satisfactorily substituted if the sense of the whole is to be restored. In the context of higher education competencies, the "whole" is the graduate who can operate competently, intelligently and flexibly, in contexts that are known now and in those that have not yet been faced or even envisaged.

HIGHER EDUCATION COMPETENCIES

Compared with many of the technical and vocational competencies, the cognitive attributes previously listed (critical analysis, problem-solving, information literacy, originality, and effective communication) are not as easily defined in concrete terms. It is difficult to describe exactly what "critical analysis" consists of, and in particular whether an actual analysis contains enough of the right kind of stuff for it to warrant the label "critical." Assuming that this property is not an all or nothing affair, it is difficult to describe in words where the threshold for an acceptable amount should be set. The same sorts of difficulties arise with "effective" communication and others in the list. To address this issue, it is not uncommon for higher education institutions to develop extended descriptions of what is covered by each of the attributes, elaborations sometimes running to many pages.

As a limited example, consider the following expansion for information literacy, which has been adapted and condensed from an actual discipline description.

The graduate should be able to:

- Access archives, libraries, the web and other written, oral and electronic sources of data and information;
- Effectively employ appropriate technologies in searching out such information;
- Apply research principles and methods to gather and scrutinize information;
- Manage, analyze, evaluate and use information efficiently and effectively in a range of contexts; and

– Respect economic, legal, social, ethical and cultural norms and protocols in gathering and using information.

In interpreting each of these sub-competencies, contextualized judgments are necessary. What each of them means in the abstract and implies in practice is therefore open to interpretation and debate. When institutional descriptions differ significantly, as they typically do, which should be taken as definitive, if any? How much does it matter if different interpretations exist and are used? Social and contextual dependence is signaled by the fact that different meanings of the key terms and concepts are obviously satisfactory to the institutions in which the statements have been formulated. A further aspect is that much the same wording for competencies can be found in formal lists of desired educational outcomes for various levels of mainstream school education. That is, the intrinsic content of the competencies is not definitively characteristic of any particular level of education. Some (such as problem-solving) appear across the educational range, from kindergarten upwards, presumably because they form part of what is normally expected of education broadly interpreted – that is, what education as a collective enterprise is all about.

The above sample of higher education competencies also serves to illustrate the essential fuzziness of the relationships among them. Although they may appear in the abstract to be conceptually distinct, those distinctions are not simple to sustain in practice. The attributes fuse into one another. For instance, problem-solving as an intellectual and practical activity is difficult to conceptualize without involving analysis, seeking out relevant information, creative development (of possible solutions), and effective communication of the solution. Where one competency or skill finishes and another starts is a fine line to draw. Furthermore, the attainment of a particular subset of competencies may, when applied, have "covered" the territory normally associated with one or more other competencies and thereby made the separate assessment of the latter redundant. Potential overlap, nesting, and partial or full interdependencies are common. This raises the issue of the extent to which it is feasible, as an exercise in the abstract, to differentiate the competencies at all. Separating and clarifying "competencies" for the express purpose of constructing tests to measure them is at best a partial exercise because separate reporting of the competencies cannot capture typical (and inevitable) in-context entanglements.

On the other hand, it is important for some purposes to be able to embrace and use the concepts as concepts. They have meaning, they have labels, and they provide both the vocabulary and the tools necessary for making systematic, functional progress. Where they most appropriately fit into the scheme of things could well be as retrospective explanatory devices – after particular judgments have been made. This would be consistent with the philosophical position that valuing, or making evaluative judgments, is a primary act of situational recognition, the justification for which necessarily invokes relevant criteria that are extracted as needed from a larger pool of potential criteria (Sadler, 2009a).

By way of concrete example, suppose an assessor composes a rationale for a holistic judgment of the quality of a student's written response to an assessment

task. Suppose, too, that the rationale refers to a lack of critical analysis in the work. The main purpose served by the statement that the work lacks the necessary critical edge is to draw attention to a desired feature which is inadequately expressed in the work. The assessor chooses this quality for explicit mention from the pool of properties that potentially matter. This act connects the particular work with the judgment made about its quality. A person interpreting the rationale in the absence of access to the work in question has no option but to guess what the work was like. Interpreting the rationale in the presence of the work, however, means that the text and its referent combine together in a message. The soundness of the reason for specifically emphasizing critical analysis can then be explored. The dynamic of the way in which critical analysis as a concept is used with and without access to the work makes a difference. Without the work, the temptation is to commodify the concept rather than communicate a judgmental framework.

In practice, only a small number of aspects or characteristics may be worthy of specific mention. What makes these salient to the appraisal is that they provide insights into the evaluative reasoning behind the judgment. In the process of operating in this mode, some properties will turn out to be pre-emptive. For instance, if a written document is so poorly expressed that it is fundamentally incoherent, it is technically unlikely it will be able to provide evidence of originality or critical analysis – or even of whether the work addresses the nominated issue at all. If the end user, whether professor, peer reviewer or employer, attempts to read between the lines of the text to figure out what the author was possibly trying to express, the high-order inferences involved in that process come at the risk of a poor judgment of actual mastery of the competency of interest and the credit that should be given to it. (This observation, of course, is not intended to imply that all text must be held to literal interpretation; linguistic convention may clearly signal other interpretations, such as irony or humor.)

Real contexts are in some ways simpler and in other ways more complex than is implied by thinking about separated competencies. They are simpler in that competent practitioners or producers normally go about whole tasks without much conscious thought as to the cognitive processes or competencies they are using. They switch effortlessly from figure to ground, and back again. Real contexts are more complex in that when producers do actually reflect on their processes and have reason to describe them, their descriptions are not necessarily framed in accord with pre-existing typologies, but adverse consequences that arise from doing this are rare. Those concerns aside, there is no denying the critical importance of a shared vocabulary with which to engage in discourse about quality and qualities, competence and competencies.

Further questions arise in relation to the legitimacy of treating competencies carrying the same label as somehow similar in essence, structure and cognitive demand across a range of disciplines, fields and professions. The similarity of labels is presumably the reason for treating them as "generic," but whether the apparent common ground accords with reality is questionable. Research on this topic has revealed wide differences in interpretation of specified competencies or attributes in different fields, and even within different sub-domains of a single field

(Jones, 2009). Critical analysis expresses itself differently in different content areas, and at different academic levels within the same content area. Locating, evaluating and using information is carried out differently in degrees in music, information technology and construction engineering. Within construction engineering, it may depend on the purpose for which the information is required and the time available for obtaining and processing it. A broad-spectrum test that purports to tap into critical analysis and information literacy as graduate competencies may not produce test results that can be interpreted appropriately, that is, matching the label for that competency as used in various curriculum specialties. If de-situated test tasks signal different nuances of academic competencies from the mainstream courses in which students enroll, and if the test results are to be used for high-stake decision-making (both of which seem to be likely propositions), to what extent would teaching time, resources and energies be diverted from the mainstream studies in order to provide space for explicit coaching to the tests?

ASSESSMENT OF COMPETENCE AND COMPETENCIES

A significant challenge facing policy-makers is finding appropriate paths through the assessment of overall competence or of individual competencies. One approach to the measurement task is to first formulate and define the competencies as psychological constructs and then to apply psychometric methods. An influential contribution to the substantial and growing literature on this approach is the review and analysis by Weinert (2001). Tested graduate competencies may be considered to stand each in its own right; alternatively, the collection may be interpreted as an assessment of graduate competence. Suppose it is accepted that a person is judged competent if they perform well over a variety of relevant contexts and challenges time after time, with little likelihood of getting things wrong. In the latter case, the collection should be examined to ascertain the extent to which the competencies tested comprise a necessary and sufficient set. Such examination would have two branches.

The first branch would be a test for necessity: do people who are already recognized as (holistically) competent in the workplace or in professional practice demonstrate all the tested competencies? The second branch would be a test for sufficiency: if graduates were clearly able to demonstrate achievement of all the tested competencies, would they subsequently function as (holistically) competent in the context of work or professional practice? Quite apart from the workplace, would they demonstrate respect for evidence, rigor, people, problems and society in their thinking, communications, and general approach to knowledge and knowing? Are these not the types of "educated" characteristics one should be able to expect of higher education graduates? These are important questions which are in principle open to empirical investigation.

An alternative to the decomposition and measurement approach is to start with the common notion of competence and seek out responsible ways to make judgments about a student's level of competence directly and holistically, rather

than by building up the judgment from components. The motivation for proceeding in this direction is the premise that the whole (competence) does not necessarily equate to the sum of the parts (the competencies). ("Sum" here is intended to include all methods of compounding or combining, as well as simple addition in the case of measurements.) This view implies that judgments of competence can properly take place only within complex situations, and not componentially. Generally, the perception is that if the whole differs from the sum of the parts, it does so in the direction of being more than – not less than – the sum of the parts, but differences in the opposite direction are not uncommon either. As Ford (1992) explained it:

> Organization exists when various components are combined in such a way that the whole is different than the sum of the parts. This "difference" involves both gains and losses. In one sense, the whole is *greater* than the sum of the parts because new qualities or capabilities emerge from the relationships among the parts that none of the parts could accomplish on their own…In another sense, however, the whole is *less* than the sum of the parts because the functioning of each of the parts has been restricted by virtue of being "locked in" to a particular organizational form (p. 22).

Reviewers of films, computer games and other creative works sometimes remark that a work under review satisfies all of the generally accepted criteria of excellence (that is, all of the components appear to be technically perfect) but the work as a whole nevertheless fails to "come together" in a way that sets it apart as outstanding, or even just satisfactory. In some cases, several reviewers independently remark that they have difficulty "putting their finger" on the residual problem or weakness, although they clearly sense it. Conversely, when the whole is judged to be more than the sum of its parts, the "something more" that makes up competence includes any extra qualities or properties, of whatever kind, that were not initially identified as attributes or competencies, and maybe others that cannot be clearly identified and named at all. It also includes the ability to "'read" a particular complex situation which is not exactly like any seen before, and know how to call on the various competencies (assuming they can be identified) productively, adaptively, confidently, safely and wisely.

Put somewhat differently, competence could be conceptualized as selecting and orchestrating a set of acquired competencies to serve a particular purpose or goal. In Ford's (1992) terms, organization makes a difference. The ability to orchestrate competencies, by definition, lies outside (and at a higher level than) the given or specified set of basic competencies. If higher-level competencies were also included in the model, the question would then arise as to how and when these also should be invoked, and the same would apply at even higher levels. In the other direction, as decomposition progresses downwards potentially to the atomistic level, it typically becomes harder and harder to conceptualize the components working together, partly because the number of possible interactions of all orders among competencies escalates rapidly.

With this interpretation of competence, sound judgments of competence require qualitative appraisals of how well a person can get it all together in a given situation. Such judgments are integrative and holistic, and are commonly made subjectively. The term "subjective" is frequently used to denigrate holistic appraisals as being little more than mere opinion or personal taste, in some cases with one opinion being more or less as satisfactory or legitimate as any other. Equally problematic are the terms "impression" and "gut feeling." That line of thinking does subjective judgments a grave disservice. Many professionals constantly rely on so-called subjective judgments that cannot be verified by independent objective means such as a standard laboratory test. Subjective judgments can be soundly based, consistently trustworthy, and similar to those made by comparably qualified and experienced professionals. They can also be poorly based, erratic and unreliable. Furthermore, in some circumstances quite different judgments may be equally appropriate for different purposes.

The goal to aim for is this: when presented with the same phenomena or objects which cover a diverse range, members operating within a guild of like-purposed professionals would make the same judgments within a tolerable margin of error. The judgments hold (that is, are accepted as proper) beyond each judge's personally constructed decision space (that is, the space available only to a particular judge), and the parameters for that shared decision space are set and accepted collegially. For a given set of phenomena or objects, the meaning and significance of evidence are shared, as is what is deemed to count as evidence. In short, given the same stimuli, the people making the judgments would react or respond similarly and judge similarly. The existing term that is probably closest in meaning to this state of affairs is "intersubjectivity," a term used with appropriately nuanced interpretations in phenomenology, psychology, philosophy and several other fields. Intersubjectivity is distinct from interscorer reliability or consistency in that not only are similar judgments made but the grounds for the judgments are shared as well. Consistency on its own can be potentially achieved without that. It is also distinct from objectivity in the sense that it is an *objective* fact that one water molecule contains two hydrogen atoms and one oxygen atom.

As Scriven (1972) has pointed out, the quality of a judgment made by a single assessor is not automatically suspect and deserving of being dismissed merely because it has been made without collaboration and without the help of instrumentation. The two latter conditions do not make all such judgments worthless. Professionals who consistently arrive at sound judgments are effectively "calibrated" against their competent peers and also, in professional contexts, against any relevant socially constructed external norms. This points to the direction in which the development of an appraiser's ability to make high-quality holistic judgments can conceivably take place – by providing them not only with experience in making multiple judgments for objects or phenomena in a given class in a wide variety of settings but also with experience in verbalizing their reasons and discussing them with appropriate colleagues who at least initially have

access to the same objects or phenomena so that the shared decision space which is crucial to the enterprise can be constructed.

In the context of assessment where judgments are holistic and integrated, the characterization above is suggested as the appropriate goal statement. The starting-point for making progress towards acceptable levels of intersubjectivity is daunting, given the well-established research finding that assessors who make judgments by operating within their personal decision spaces generally exhibit low interscorer reliability. Furthermore, in some higher education contexts, the right of academic teachers to make grading decisions that way (that is, as they individually see fit) is strongly defended. The challenge ahead is to find ways to create and value shared rather than individuated meanings and knowledge as a primary resource for making competent professional judgments. What might that involve?

COMPLEX JUDGMENTS – THE IMPORTANCE OF NOTICING

In his characterization of knowledge types, Ryle (1949) made a distinction between "knowing how," which is being able to do something whenever required, and "knowing that," which is knowing something such as a fact, a theorem or a classification. Know-that knowledge is commonly memorized, and tested by using language-based items or tasks (words, symbols or other material representations). Know-how knowledge is commonly learned through practice, and tested by setting up various skill-based tasks. Largely overlooked in Ryle's dichotomy is another form of knowing: "knowing to," in which an appraiser notices, detects or "senses" as salient-in-the-circumstances some aspect that contributes to or detracts from the overall quality or effectiveness of a work. In knowing-to, high-level judgments are critically important. This type of knowledge cannot necessarily be made explicit, that is, expressed in words. It nevertheless exists and is widely used, even when a person cannot define it in concrete terms or otherwise explain it. Such know-to accounts for part of what chemist-philosopher Polanyi (1962) called "tacit knowing," captured in his remark that one can know more than one can tell. A decade earlier, Wittgenstein (1953) had expressed much the same idea in his observation:

> I contemplate a face, and then suddenly notice its likeness to another. I see that it has not changed; and yet I see it differently. I call this experience 'noticing an aspect' [XI, p. 93].

Similarly, Abercrombie (1969) in her classic work on judgment discussed the intricacies of perception and prior expectations and how they influence what is noticed and deemed to count as data in a particular context. Consistent with the work of Polanyi, Wittgenstein and Abercrombie is that of Dreyfus and Dreyfus who argued that experts regularly use their "intuitive rationality," on occasion engage in "'deliberative rationality" (when time permits and this provides a workable way forward), and much less often employ formal "calculative rationality." In their 1984 article, they put it this way:

> [E]xperience-based similarity recognition produces the deep situational understanding of the proficient performer. No new insight is needed to explain the mental processes of the expert. With enough experience with a variety of situations, all seen from the same perspective or with the same goal in mind, but requiring different tactical decisions, the mind of the proficient performer seems gradually to decompose this class of situation into subclasses, each member of which shares not only the same goal or perspective, but also the same decision, action, or tactic. At this point, a situation, when seen as similar to members of this class, is not only thereby understood but simultaneously the associated decision, action or tactic presents itself. [p. 225].

The substantial literature on the nature of expertise and how it is developed is an important resource for further thinking. A great deal of what experts do has to be learned through extended experience, but not necessarily through experience alone, a particularly important contribution to that aspect being the seminal volume of Bereiter and Scardamalia (1993). As well as the authors listed, the literature includes research relating to the development of competence in appraisal by medical and health practitioners, airline pilots and many other professionals who are involved in complex decision contexts.

DEVELOPING HIGHER EDUCATION COMPETENCIES

In this section, the third possible cause of concerns about current levels of higher education competencies is picked up again. The dual agenda consists of two questions: what current aspects of teaching and assessment inhibit the development of higher education competencies? how might improvement be brought about? The proposal outlined below is based on the notion that the responsibility needs to be shared between academics as educator-assessors and higher education institutions as controllers of the parameters within which academics work. An approach followed by some institutions is to make it mandatory for course designers and directors to embed at least some of the higher education competencies in each course. The hope is that over an entire degree program all competencies would be developed. This assumes, of course, that competencies are conceptually separable, something which goes against the grain of the theme in this chapter, but on the positive side it allows competencies to be expressed in ways relevant to individual courses. A second approach is to focus on developing the assessment competence of higher education teachers, strengthen their resolve to award course grades according to appropriate academic standards, and concurrently reset the system and the institutional parameters to facilitate both of these.

With the second approach in mind, academics would need to develop high-level capability in: designing assessment tasks that are clearly specified and outline the higher-order cognitive outcomes required, including the specification of a particular product type (such as a critique or a design) if appropriate; holding students to the specifications (task compliance); and becoming calibrated in their appraisal practice so that the standards they employ are not peculiar to themselves.

Task compliance (Sadler, 2010a) implies not awarding credit at pass level or higher for works that do not deliver on the specifications. In particular, merely compiling and reproducing written material without serious intellectual engagement with it may not qualify as evidence of academic achievement, nor would purely the effort that may have been put into producing a response.

That might seem an obvious way forward except for the fact that many academics assert that if they were to apply the standards in their heart of hearts they know they ideally should, the result would be failure rates that are unacceptable to their institution or to external higher education authorities. The policy settings of many institutions and systems work against the realization of the goal. To illustrate, consider an academic faced with a high level of difficulty in deciphering a student's work. In theory, almost incoherent or poorly communicated work should disqualify a student from passing, that is, from gaining credit and progressing to the next stage of the academic program. In practice, non-achievement variables can and do influence many grading decisions. One such variable is related to maintaining high student retention rates in contexts where recruitment of students is competitive and institutional income from public or government sources is, by legislation or policy, inversely related to attrition rates. That external constraint is mirrored by an internal dispositional aspect on the part of an academic: the assessor may wish to award a realistic course grade to a student but in reality is faced with the likelihood of adverse consequences (poor student evaluations or an institutional inquiry into too high a failure rate) and so leans towards giving the student the benefit of the doubt.

An additional practice that detracts from the integrity of course grades is the awarding of marks or points for what on the surface may appear to be sound educational reasons (for encouragement, for demonstrating improvement, or as a reward for engagement or participation) but in reality amount to giving false credit for elements that are not strictly achievements at all. Furthermore, if the course grade is intended to represent the level of achievement reached by the end of the course, something usually implied or stated in the intended course learning outcomes, accumulating points throughout a unit of study is unsound. These and related topics are dealt with in detail in Sadler (2010b). Such practices can create a near-pass score by dubious means. The consequence is that students then have to attain relatively few of the most highly valued educational outcomes in order to pass, gain course credit and progress to the next course. Regardless of whether these factors have their origins in overt institutional policy or are simply practices that have been accepted incrementally into the assessment culture, they reduce the likelihood of attaining the desirable higher-order academic outcomes in graduates. Turning things around would not be fast or simple but the payoff might well be worth the effort.

CONCLUSION

The point of view reflected in this paper follows a different line from that of most contemporary developments. The focus is not on large-scale modeling of

competence or competencies and the measurement of their attained levels in higher education institutions. Instead, the focus is on the concept of competence as the capability to orchestrate knowledge and skill independently, in a range of contexts, on demand and to a high level of proficiency. The complementary focus is on competence as it is acquired and developed by students within their regular academic programs, and how that competence might be enhanced and assessed.

Underlying this orientation is the premise that each academic program and each academic course provides the most appropriate site for learning higher-order cognitive and other skills. This defines a key role for academics as educators and an aspiration for higher education as an enterprise which is central to attainment of academic aims and objectives. What are currently being labeled graduate attributes need to revert to being integral elements of academic learning, with performance in them ultimately being reflected in the course grades recorded on academic transcripts. The success of moving in this direction would depend directly on having not only competent academics but also an institutional commitment to sophisticated outcomes and high academic achievement standards.

REFERENCES

Abercrombie, M. L. J. (1969). *The anatomy of judgement: An investigation into the processes of perception and reasoning*. Harmondsworth, UK: Penguin.

Bereiter, C., & Scardamalia, M. (1993). *Surpassing ourselves: An inquiry into the nature and implications of expertise*. Chicago, IL: Open Court.

Dreyfus, H. L., & Dreyfus, S. E. (1984). From Socrates to expert systems: The limits of calculative rationality. *Technology in Society*, *6*, 217–233.

Ford, M. E. (1992). *Motivating humans: Goals, emotions, and personal agency beliefs*. Newbury Park, CA: SAGE.

Jones, A. (2009). Redisciplining generic attributes: The disciplinary context in focus. *Studies in Higher Education*, *34*, 85–100.

Polanyi, M. (1962). *Personal knowledge: Towards a post-critical philosophy*. London: Routledge & Kegan Paul.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

Sadler, D. R. (2007). Perils in the meticulous specification of goals and assessment criteria. *Assessment in Education: Principles, Policy & Practice*, *14*, 387–392.

Sadler, D. R. (2009a). Indeterminacy in the use of preset criteria for assessment and grading in higher education. *Assessment and Evaluation in Higher Education***,** *34*, 159–179.

Sadler, D. R. (2009b). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, *34*, 807–826.

Sadler, D. R. (2010a). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, *35*, 535–550.

Sadler, D. R. (2010b). Fidelity as a precondition for integrity in grading academic achievement. *Assessment and Evaluation in Higher Education*, *35*, 727–743.

Sadler, D. R. (2011). Academic freedom, achievement standards and professional identity. *Quality in Higher Education, 17*, 103–118.

Scriven, M. (1972). Objectivity and subjectivity in educational research. In L. G. Thomas (Ed.), *Philosophical redirection of educational research* (71st NSSE Yearbook, pp. 94–142). Chicago, IL: National Society for the Study of Education.

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies: Theoretical and conceptual foundations* (pp. 45–65). Seattle: Hogrefe & Huber.

Wittgenstein, L. (1953). *Philosophical investigations* (Anscombe, G. E. M., trans.). Oxford: Basil Blackwell.

D. Royce Sadler
Teaching and Educational Development Institute
The University of Queensland