Assessment of student outcomes from work-integrated learning: Validity and reliability

CALVIN SMITH[1]
Griffith University, Brisbane, Australia

In this article the issues of validity and reliability in the assessment of the learning outcomes of WIL curricula are considered. An argument is made that there are three classes of learning outcome for WIL curricula: experience of the work-world; the development or refinement of skills; and the application of disciplinary knowledge in work contexts. It is noted that the protocols for assessing the first two of these outcome classes have a longer history and are better developed, in terms of established validity and reliability, than those for assessing the third outcome class. An elaboration is then made of the scope and meaning this third class of outcome, and it is argued that the naïve notion of "application" can be developed fruitfully in terms of three requirements that force the explicit integration of theory and practice: (1) interpretations of the setting; (2) predictions about the impacts of possible future action scenarios therein; and (3) action choices. When all three are justified by reference to disciplinary canonical knowledge and theory a strong case can be made that students have successfully integrated theory and practice. This idea is developed and a generic rubric is proffered that may be of use to WIL curriculum designers and assessors. *(Asia-Pacific Journal of Cooperative Education, Special Issue, 2014, 15(3), 209-223)*

Keywords: work-integrated learning; assessment; validity; integration; application

ASSESSMENT: PURPOSES AND FORMATS

A sophisticated perspective on assessment has developed over the past several years as our conceptualizations have become clearer and our practice more focused. This view is encapsulated in Biggs' concise summaries (for example, Biggs, 2003, p. 140-169). Assessment has two fundamental purposes: learning and testing. When used to support learning, results are fed back to students so they may understand better their strengths and weaknesses (whether in knowledge, understanding or skills). When used for testing, results are part of a broader social system of accreditation, assurance, and certification; they are used as the evidence base for the attestation of the institution or teacher that the student has achieved a certain standard (whether in knowledge, understanding or skill).

A further consideration is that which demarcates norm-referencing assessment systems from criterion-referencing systems. With norm-referencing there is a focus on comparing people with each other, student with student. The results in such a system are expressed in ways that allow this to be done meaningfully – for instance using percentile-based ranking to specify which students are in the 'top x percent' of the cohort. This can be important for providing a convenient way of allocating limited high-reward opportunities such as entry into honors or the granting of scholarships and bursaries. Protocols for norm-referenced assessment often claim high reliability and high validity, though in practice, for some purposes validity is mis-targeted (i.e., it validly measures the 'wrong' things).

[1] Corresponding author: Calvin Smith, Calvin.smith@griffith.edu.au

Criterion-referencing on the other hand focuses on comparing students' performances with a set of graded descriptions of performance, in different dimensions. Thus, it relies on being able to (a) specify in advance what typical performances 'look like' at different achievement 'levels', within the dimensions of a task, and (b) give a description or exemplar (or both) of performances at each level within each dimension. For instance, an essay may be broken into dimensions (clarity of expression, grammar, logic, relevance of evidence drawn upon) and for each one of these *dimensions*, students' performances in each of a series of achievement *levels* (from 'fail' through to 'excellent') are described and/or exemplified. Student works are then compared with this scheme and marks allocated accordingly

In practice, regardless of the purpose, a variety of assessment protocols is relied upon (tools, tests, formats, methods, activities, etc.). Essays, multi-choice exams, OSCEs (Objective Structured Clinical Exams), reflective pieces, problem-based learning, action learning sets and so forth, are used to achieve various assessment ends. Not all protocols are equal in terms of validity and reliability, and some serve testing purposes better than they serve learning purposes, whilst for others the opposite is true.

VALIDITY AND RELIABILITY

Validity in assessment is a matter of whether and to what degree a protocol (i.e., a test activity) or an element within a protocol (e.g., an item in a test-bank, a dimension in a standards rubric) measures the thing it purports to measure. For validity to be established, it is necessary to be clear about what is to be measured. For this to be so in educational settings, it is necessary to be clear, in advance, about what the intended learning outcomes are. This clarity must drill right down to sub-dimensional levels of performance – demonstrations of knowledge, understanding, or skill. Assessment devices or protocols can only be evaluated against these known-in-advance conceptualizations of the thing-to-be-measured; they are valid to the extent that they can be shown to measure the thing they are designed to measure.

The concept of "holistic" assessment comes in here (Biggs, 2003, p. 157). An assessment protocol may validly measure only some aspects of performance, but not the whole performance (or at least not all relevant aspects). If the aspects it does not measure are crucial to professional practice, then the assessment protocol is valid but not *complete*; if it then is used to certify a student for practice, there can be negative, even dire, consequences.

Reliability is the degree to which the same result will be obtained when the assessment is done either by different assessors or by the same assessor on different occasions. Similarly, it is the degree to which the same result will be obtained when the assessment is done by the same student on different occasions (assuming no practice effect and no change in the student's abilities).

The question of validity and reliability will be discussed in sections below that deal with specific protocols. Validity and reliability relate to the assessment of intended learning

outcomes, so before examining protocols it is necessary to consider what the intended outcomes of WIL are.

WIL: PROCESSES, INTENDED OUTCOMES AND ASSESSMENT

The creation of WIL curricula is motivated by a desire to make certain kinds of opportunities available to students and to develop in them particular kinds of *intended learning outcomes*.

In brief, WIL curricula are designed to give students opportunities to:

1. Experience the work-world before they graduate;
2. Develop or acquire relevant skills, professional abilities, professional attitudes, and attributes; and
3. Apply their knowledge and skills *in situ*.

These outcomes can be achieved (to differing degrees) by two means: high-authenticity simulated environments and placements in real work/professional settings. Both simulation and placement are forms of WIL because they bring the students into contact with real (or near-real) fields of application – that is, contexts in which they will conduct their work after graduation.

Simulated environments have two characteristics that make them worthy of closer inspection in regard to assessment. First, because they are simulated, they afford their designers much greater control of the reliability of the experiences that students are exposed to than is the case with real workplace experiences. This means that they are easier environments from which to extract standardized, assurable learning outcomes. Second, again because they are simulated, their validity as representative of real-world work environments can be limited; the real world is richer, deeper, and unpredictable. This creates a challenge to the *validity* of assessment where the outcomes assume exposure to rich, unpredictable, real-world experiences, but the simulation provides narrow, controlled experiences. A flight simulator teaches pilots how to fly, but not how to navigate their way in a car though Bangkok traffic and arrive at the airport on time.

In a similar vein, although real-world placement may provide rich experiences, they may or may not expose students to just the right mix of the right kinds of experiences for them to achieve the learning intended for them. It is a contingent and uncontrollable matter what exactly the environment 'throws up' for the student to experience, unless this aspect is controlled in some manner. That is, placements are *unreliable* in terms of the consistency of the experiences a cohort of students gets across a widely disparate set of workplaces. Further, the issue of the *validity* of placement assessments is thrown into the relief: if placement cannot assure the right mix of the right kinds of experiences, then the assessment of placement students against a pre-conceived set of criteria and standards cannot be facilitated fairly, evenly and consistently, and validity cannot be assured.

WIL curricula may fulfil the desire to create learning environments for students that are different from classroom (lecture/tutorial/laboratory) environments in that they either are

situated in the real work-world, or they are situated in plausible, high-authenticity, simulated work-worlds. However, the learning outcomes achieved (by students) and desired (by curriculum designers) are not necessarily convergent. Students may learn things outside of the scope of the intended learning outcomes, or they may not have sufficient of the right kinds of experiences to ensure all the objectives have been met, in a standardized way, at an individual level. Assessment for WIL must address these issues.

WIL ASSESSMENT: PURPOSES, FORMATS, VALIDITY AND RELIABILITY

Experiencing the work-world

One set of outcomes that has received considerable attention in the literature is that relating to *experiencing* the world of work (Patrick et al., 2009). There are two ways of conceptualizing this broad outcome and, therefore, its assessment:

1. Mere exposure to the experience, sometimes accompanied by reflection on whatever reactions, interpretations, observations and so on, that occurred to the student during the experience; and
2. Use of the experience as either
    a. a source of learning about the practice, or
    b. as data for theorizing about the experience or context.

For the first of these outcome sub-classes, reflection-on-practice is commonly set as the main focus of learning where students typically are asked to produce reflective diaries that chronicle their reactions to their experience of work or workplace. Reflective diaries are a common WIL assessment method for this class of outcomes (for a review, see Bates, 2005). These protocols may structure, through trigger questions for instance, what students focus on, or may be free-ranging and open-ended, allowing students to reflect on whatever was of importance to them. This approach is often made more legitimate-seeming by the use of theories of reflective practice to underpin the pedagogical activity (Boud, Keogh, & Walker, 1985; Kolb, 1984; Schon, 1983).

These protocols are quite open-ended in terms of what might be reflected upon. This is both a strength and a weakness. Such protocols go some way to addressing the problem of 'holism' or completeness in the focus of the assessment, but at a likely cost to validity and reliability (especially for the purposes of attesting and accreditation) since each student's account will necessarily be idiosyncratic. These idiosyncratic outputs may have a high degree of personal validity for each student, but with no pre-set external criteria by which to judge them, their use as a valid measure of intended learning outcomes is limited, and the contingent nature of the circumstances in which they are produced, also renders their reliability questionable.

To address the shortcomings of assessment protocols relying on reflection, they should be marked against criteria in a criterion-referencing framework. This approach stabilizes the validity and reliability concerns by using criteria and standards descriptions to guide

marking whilst maintaining freedom in the scope of what different students address (Bates, 2003, 2005, 2008; Boud et al., 1985). Some authors have argued that reflection on work experience may be more valid for post-graduate continuing professional development, than for initial development of abilities (Cheetham & Chivers, 2000).

The second of these outcomes (sub-class 2a) deal with the development of working or 'practical knowledge', generated through reflected-upon practice, observations or experience, and is commensurate with the approaches of Boud (2001; 1991) and Billett (2001, 2004, 2008). The third of these outcomes (sub-class 2b) is commensurate with Kolb's (1984) experiential learning model.

Developing Skills, Professional Abilities, Professional Attitudes and Attributes

This is the most commonly and thoroughly-addressed collection of outcomes that are both espoused and assessed in WIL contexts. It is for the assessment of these kinds of outcomes that has prompted the development of protocols that feature high validity and reliability.

Going back even before Girot's (1993) seminal treatment, there has been much written about the assessment of *competence*. In the several decades since this has been a topic of interest, many instruments and approaches have been developed and tested for assessing practice competency in a variety of disciplines and some good examples of protocols are emerging for this purpose.

However, in a review article in a medical context, Epstein and Hundert (2014, p. 230) note that although multi-choice exams are highly reliable for assessing knowledge, OSCEs, which are used for clinical assessments, require a large number of standardized patients if assessment of complex reasoning skills is to be reliably evaluated. Alternatives, such as assessment of video-recordings of clinical work, are of unknown reliability. They argue for developments focusing more on clinical reasoning.

In recent times, some disciplines have been improved by the development of assessment protocols featuring high validity and reliability. An example is the Assessment of Physiotherapy Practice (APP) tool developed by Dalton in Australia (Dalton, Davidson, & Keating, 2011). This tool has been evaluated with state-of-the-art empirical evaluation methods and rigorous filed testing, and has been shown to have high levels of inter-rater reliability and has passed a variety of validity tests.

In spite of its rigorous provenance, the APP focuses mainly on practice. Performance criteria include descriptors the vast majority of which are descriptions of actions that *imply* understanding, or awareness of standards of performance and so on. The action descriptors *embody* what the assessors *take to be* evidence of the student having this or that theoretical understanding, tacitly, but do not require students to explicate this knowledge. It is not sufficient for validity in the assessment of *integration* of underlying knowledge and practice, to rely upon the assessment of *performance*.

Though it should not stand as a criticism of a practice-focused assessment, that it assesses *practice*, even Girot (1993, p. 84) observed the need for integration, without naming it as such, over two decades ago:

> If competence is concerned with the ability to *coordinate* cognitive, affective and psychomotor skills, in the carrying out of nursing activities, all three elements of learning needs to be addressed in the process of assessment.

Yet, in spite of this remonstration, relatively little has been done to focus on this *integrating* idea. Making this coordination itself *the object of our appraisals* will be the focus of the last sections of this essay.

Another example of recent high-quality developments is COMPASS for assessment of speech pathology students (McAllister, Lincoln, Ferguson, & McAllister, 2010, 2011). Like the APP, this instrument has undergone rigorous development and testing, and it has several dimensions dedicated to practice and some dedicated to clinical reasoning in context.

Validity and reliability in the case of instruments that measure clinical performances and abilities is a complex matter involving multiple kinds of validity and numerical methods for the assessment of both validity and reliability. Advice on validity tends towards multiple dimensions of practice and sometimes multiple sources of evidence (Wass, Vleuten, Shatzer, & Jones, 2001).

Appling Knowledge and Skills in situ: Integration and the Theory-Practice Nexus

The notion of the *application* of knowledge *in situ* becomes increasingly abstruse when attesting that it has happened. Attesting that application-of-knowledge has occurred is crucial for the accreditation purpose of assessment.

Even if it is assumed that (a) students embarking on placement have in their minds the appropriate disciplinary knowledge, and (b) they have been presented with problems during their WIL experiences, the solutions to which make that knowledge relevant, there are no ready-to-hand ways of ascertaining that the students have *drawn on* that knowledge in order to solve those problems, let alone drawn on that knowledge *correctly*.

Instead, it is possible that the students have copied practice routines that were observable (or maybe even taught and reinforced) in the workplace. At worst, this can result in the privileging of corrupt and indefensible practices (the Rodney King case is a case in point (see, Cannon, 1999)). This is the exact opposite of the goal of application of canonical knowledge, which is to show that state-of-the-art thinking can be integrated into students' practice. Nevertheless the idea of the application-of-knowledge *in situ* is *de rigueur* in many discourses about WIL.

The emergence of an interest in the theory-practice nexus (Smith & Worsfold, 2013a) arises from dissatisfaction with what might be termed the curriculum of canonical knowledge. As disciplinary research progresses, the abstract knowledge created by the research process becomes the canonical knowledge of the discipline and the curriculum tends increasingly to

focus on this knowledge. This leaves a gap in the capability-set of graduates, which in turn has motivated increasing interest in the practice-focused curriculum which is seen as a supplement or complement to the knowledge curriculum. Taking this idea a step further, there is a movement from parallel supplemental curriculum strategies to those which attempt to provide a ground in which knowledge and practice can become fused, integrated.

This step invokes the notion of the *integration* (as represented by the 'I' in 'WIL'), whereas the idea of *application* falls way short of the more complex notion of *integration* of theory and practice, or the theory-practice nexus (Smith & Worsfold, 2013a, 2013b).

In fields in which the practice-canon is sufficient for successful practice, it can form the basis of a curriculum – a practice-curriculum – in which novices can be gradually trained into the practice canon and develop into independent, competent practitioners. Some writers, particularly in the field of work-*based* learning, seem to hold the view that practice *is* the curriculum (Billett, 2001, 2004). Clearly, this can only be the case in a limited number of circumstances, and never where application or integration of knowledge and practice is the goal.

In the same manner in which canons about *practice* can be developed, it is a commonplace that the disciplinary knowledge regimes familiar in the higher education sector are canons of *theory*. Neither of these canons alone, nor both together, will allow any student to become a competent professional practitioner, since both are discourses (about practice and knowledge respectively). WIL curricula create the opportunity for students to integrate these two discourses with authentic practice. Figure 1 is a representation of the relationships between practice canon, discipline canon and naïve or untheorized practice.

It is the presence, and integration, of disciplinary theory (not just practice theory) that makes WIL an integrating curriculum strategy – not merely work *experience*, not on-the-job *training*; and potentially much more than an opportunity to develop *practice competency*. The distinction here mirrors the difference between work in which trained and routinized practice is the norm, versus work in which theoretical underpinnings play a part.

Notwithstanding this claim of the centrality of integration to WIL, in their comprehensive review, Patrick and colleagues made no reference to assessment practices that focused on integration of theory and practice (though there were many mentions of reflection in the assessment review) and the closest thing to this might be the "technical defense" listed in their Table 1 as occurring in engineering (Patrick et al., 2009, p. 41). Similarly Orrell (2011: 9) observed that although there is recognition of the importance and significance of the task of developing models of assessment in WIL, and that there are examples of practice- and competency-focused approaches, there remains an important question regarding not just *how* to assess but *what* to assess.

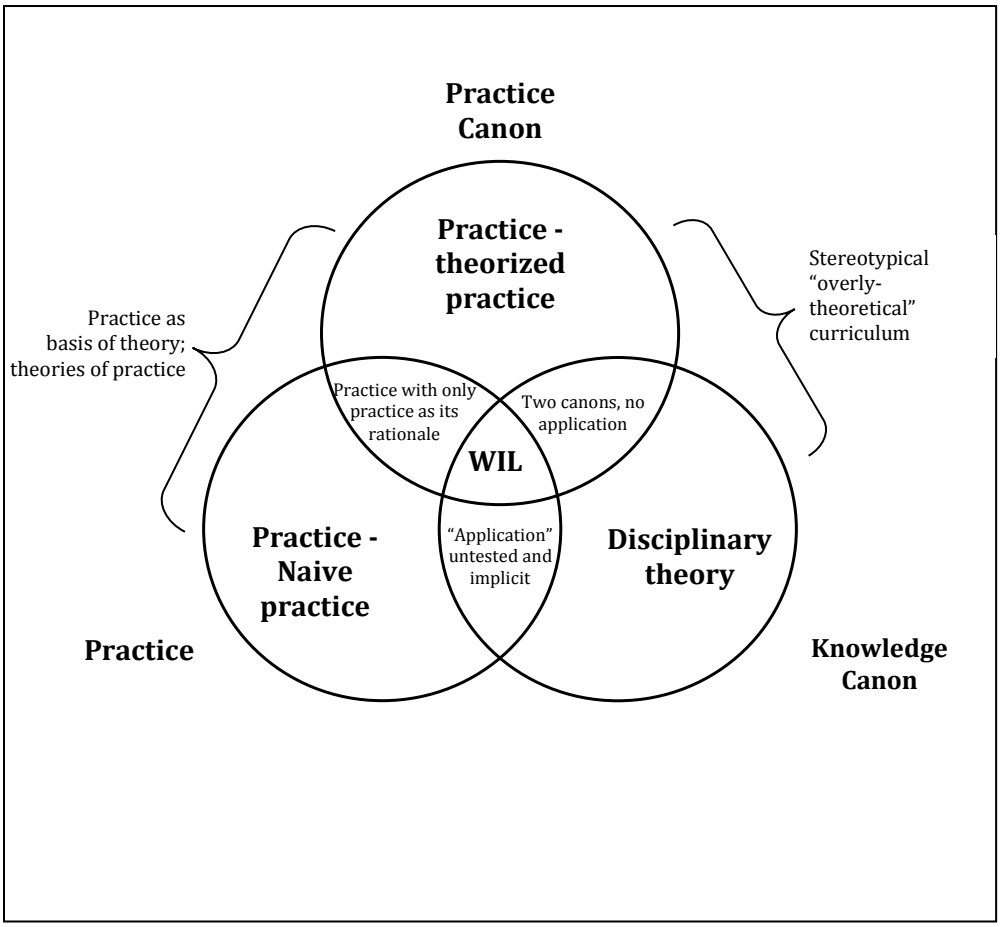

FIGURE 1. Place of practice and theory in the curriculum

To achieve and certify integrative learning, it must be deliberately designed into curriculum activities and assessments. To quote Dewey:

> We never educate directly, but indirectly by means of the environment. Whether we permit chance environments to do the work, or whether we design environments for the purpose makes a great difference. Any environment is a chance environment so far as its educative influence is concerned unless it has been deliberately regulated with reference to its educative effect (Dewey, 1939, p. 19).

Thus integration, if it is a worthwhile outcome, needs to be deliberately designed into WIL curricula. The next section will consider the purposes and benefits of focusing on integration as an outcome of WIL.

The Goal of Integration: Novice-Experts

It would be a degradation of educational aspiration to assert that WIL merely produces an entry-level professional, where that term implies a reduced level of responsibility-taking and

autonomous practice by new graduates in any field; once 'out there' the entry-level practitioner is playing a role that has consequences and impact on people. Rather, what professional preparatory programs produce at first may seem paradoxical: they certify and accredit novice-experts.

Although students and graduates do not have the supposed 10,000 hours of practice to make them experts, they are 'on their way' to that outcome; in that sense the ultimate aim in education is to produce experts. Immediately upon departure though, they are, of course, novices at being expert. Expertise implies the ability not just to practice, as though by rote, but to *be able to say why* (and to be able to say *why* correctly) the practice is to be done a particular way; justification of action in context is the true expression of expertise, even though in practice such justifications are typically tacit. Whilst tacit expert knowledge may be correct, the fact that it is tacit implies that "we can know more than we can tell" (Polanyi, 1966, p. 4). In contrast, a good test of a novice-expert is calling them to make an *explicit account* of their reasoning, their theorization, and their justified action selection in any appropriate professional context.

Shanteau and colleagues developed the Cochran-Weiss-Shanteau (CWS) method for appraising expert performance focusing on two defining abilities or qualities of expert performance (Shanteau, Weiss, Thomas, & Pounds, 2003, p. 4):

- *Discrimination* ("[t]he ability to differentiate between similar, but not identical, cases"); and
- Within-expert *consistency* (the ability consistently to "repeat his/her judgment of a case").

They have successfully applied their approach to novices and experts in a comparative framework indicating the applicability of these ideas to novice practitioners. Thus the ideas that constitute expertise can be present at a novice "level of development", in novices (Weiss & Shanteau, 2003, p. 9).

Here is what Weiss and colleagues say this about expertise:

> We argue that evaluative skill is the basic cognitive ability that characterizes all these areas of expertise. Whatever the task, therefore, the expert must attend to relevant aspects of the situation and decide what needs to be done. It is this common element, evaluation, that our index is designed to capture. What distinguishes the categories is what the expert must do after the evaluation has been carried out:
>
> - Evaluation + qualitative or quantitative expression = expert judgment.
> - Evaluation + projection = expert prediction.
> - Evaluation + communication = expert instruction.
> - Evaluation + execution = expert performance. (Weiss & Shanteau, 2003, p. 3)

The Venn diagram at Figure 2 shows the relationships between actions, the world ('immanence' in the figure), and theory. Justified action in the world is the ultimate goal. A

prior step is that the student has interpreted the world (correctly) in accord with theory and then made predictions about the likely course of events given different alternative action scenarios. A choice between these action scenarios needs to be made, which is itself justified, and then the action can be implemented as a justified action. This sequential package of theoretically informed *interpretation*, alternative action scenario generation and analysis by *prediction*, followed by an action choice, with *justification*, forms a viable package of explicated reasoning that could be solicited at key times during placement and would form the artefactual record of student performance in 'integration'.

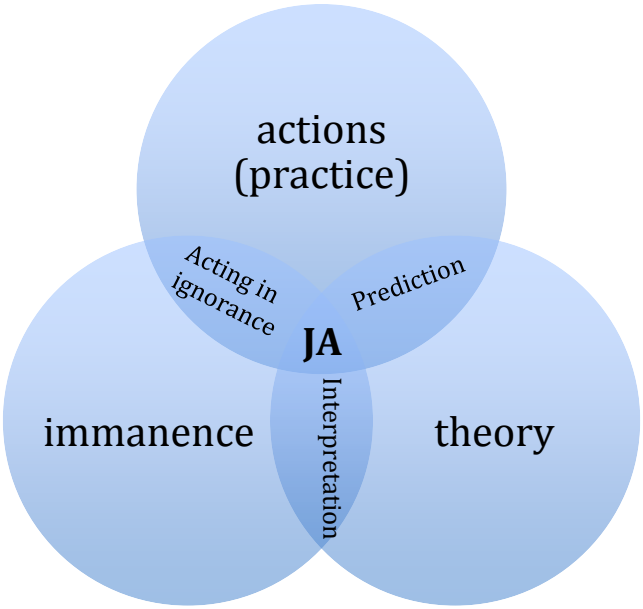

FIGURE 2. Justified actions in context (JA= justified actions)

Validity and Reliability: Implications for WIL Assessment

In assessing integration of theory and practice, to ensure validity, any assessment protocol must capture the notion of a correct *interpretation* of reality, *predications* of future states based on theory, and *justifications* for actions, based on *theorized predictions* (i.e., that refer to theory not just physical or experience/practice-based prediction). *Communication* (see Weiss & Shanteau, 2003: 3), then, hovers above this Venn diagram as the means by which *interpretations*, *predictions,* and *justifications*, are to be conveyed to the assessor, and the *enactment* (execution) can be assessed with existing approaches embodied for example in OSCEs.

Assessment for integrative thinking then should focus on justifications of actions in situ. Justifications come in two stages. When students are applying theory to the immanent

details of the real world they first must interpret that world. That interpretation must be steeped in theory, framed by theory, but in a critical manner. In some disciplines where the science is fairly robust, interpretations can be framed by theory, but in other disciplines especially the interpretive social sciences where theory is contested there should be a mature dialectic between the world and the theory so that one criticizes the other and vice versa. An example would be criterion-referenced assessment rubric for judging interpretations, predictions, and action-justifications in context.

For both formative (feedback for learning) and summative (accreditation) purposes, an example of qualitative assessment (e.g., criterion-based assessment) would be a set of criteria (dimensions of the task) and standards of performance against them. An example is in Table 1

TABLE 1. Example qualitative (criterion-referenced assessment) rubric.

	Standards		
Dimensions	Fail	Acceptable	Excellent
Interprets reality through a lens of theory	Only describes what is visible; un-theorized account	Either is uni-dimensional in interpretation, or theoretical accounts are sometime not complete; has an acceptable level of incorrect theoretical interpretation (the level would be set by discipline)	Has multi-layered account of how reality is interpretable from either different theoretical perspectives or with different dimensional emphases / aspects of reality being brought to the fore; interpretations are complete and correct with respect to how they incorporate theory
Makes predictions about different possible action scenarios referring to theory	Conceives only one possible action scenario; only describes action scenarios and no reference to the theories that generate predictions	Makes sufficient but incomplete set of predictions with acceptable (as set by discipline) levels of correct, appropriate theoretical links	Makes a complete set of correct theory-based predictions about possible action scenarios
Can justify action decisions	Has no justification for choice or makes no choice or action	Successfully justifies an acceptable (set by discipline) range of action choices	Successfully justifies all reasonable action choices

As with all criterion-referenced systems, reliability is determined by the clarity and interpretive stability of the criteria and standards descriptors. An important practical step is to work with students to develop their ability to interpret the meaning of the standards description used in the system (Smith, Worsfold, Davies, Fisher, & McPhail, 2013). Validity is a matter of assessor's expert opinion, and this may vary from time to time. Communication is the means by which the student explicates their integrative thinking. Assessor judgment and student communication interact in the assessment process. For instance 'successfully justifies' relies both on the judgment of the assessor/s and on the novice-expert communication skills of the student in articulating the integration of theory and practical matters.

As mentioned above, the APP (Dalton et al., 2011) is one of the best-researched instruments for assessing competence. One striking feature is the emergence of justification in one of the performance indicators (descriptors of behavior used by assessors to make judgments about student performances) – under the heading "applies evidence to practice" is "Options for physiotherapy intervention are identified and justified…". Although it is not common throughout the APP, the emergence of this performance indicator is commensurate with the direction being advocated in this chapter; more needs to be done to develop assessment protocols that capture the integration of practice and knowledge in WIL settings.

In 2002, JAMA published a paper that was based on a thorough review of literature on clinical competence assessment (Epstein & Hundert, 2014) in which the authors target similar ideas for developing assessment of clinical competence to those espoused here for the assessment of competence in all fields in which WIL is conducted. Epstein and Hundert define competence as:

> …the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served. (Epstein & Hundert, 2014, p. 226)

Knowledge in this view is not separate from practice, but central to its competent enactment. Regarding the 'acquisition and use of knowledge' in evidence-based medicine Epstein and Hundert suggest that integrative thinking provides "an explicit means for generating an important answerable question, interpreting new knowledge, and judging how to apply that knowledge in a clinical setting" (Epstein & Hundert, 2014, p. 227).

They go on to say:

> [competence includes] … an *integrative* function—using biomedical and psycho-social data in clinical reasoning. (Epstein & Hundert, 2014, p. 227, emphasis added)

Although writing about clinical competence in medicine, this idea of integrative use of knowledge in practice is core to the notion that, regardless of discipline, WIL should provide opportunities of this kind, not just of work-experience or skill development.

They provide a table of the dimensions of competence with headings; under the heading *cognitive* their list includes "applying knowledge to real-world situations" and under *integrative* they list:

- Incorporating scientific, clinical, and humanistic judgment;
- Using clinical reasoning strategies appropriately (hypothetico-deductive, pattern-recognition, elaborated knowledge);
- Linking basic and clinical knowledge across disciplines;
- Managing uncertainty. (Epstein & Hundert, 2014, p. 227)

Thus their conceptualization of the outcomes of education focused on integration is commensurate with the idea of integration developed above – the use of knowledge to interpret, predict, and justify practice decisions in the real-world. Any of those elements listed could be applied in most, if not all, disciplines.

Assessments of integrative thinking, or integration, as a learning outcome of WIL need to focus attention on these moments in which knowledge and the world intersect through acts of interpretation, prediction and action. This area of pedagogical and curriculum development is in need of urgent and focused attention. After a review of the approaches to assessment Epstein and Hundert concluded:

> Although curricular designs increasingly integrate core knowledge and clinical skills, most assessment methods evaluate these domains in isolation. … Few reliably assess clinical reasoning. (Epstein & Hundert, 2014, p. 230)

The current state of play marks a turning-point in two respects. First, in some disciplines the idea of integration being the main aim of WIL curricula will be a new idea, perhaps even a challenging one; success implies the alignment of workplace experience with the goals of integrative learning, a condition that may take a lot of work to assure in some contexts. Second, where integrative learning outcomes are to be incorporated, assessment needs to follow suit, and this step may present further challenges; even in clinical medicine and allied health disciplines, where the idea seems eminently appropriate, there are challenges in the creation of valid and reliable assessments. Such challenges are being addressed gradually.

CONCLUSIONS

Assessment validity and reliability are necessary for the valid accreditation by universities of practitioners in any field. Clarity about the learning objectives being assessed is necessary for the construction of valid and reliable assessment protocols. Thus a first step in valid and reliable assessment of work-integrated learning outcomes is clarity in the articulation of the learning goals of WIL curricula.

There are three general classes of learning outcome from WIL curricula: experiencing the work-world; developing professional abilities; and applying canonical knowledge in work contexts. Each outcome class requires different kinds of assessment protocols, and raises different concerns about validity and reliability.

Integration of canonical knowledge with practice in the work context is a way forward for all cases of WIL. By focusing assessment on students' ability to *justify*; 1) their interpretations of a situation, context or case before them; 2) the possible range of actions they might take and their consequences; and 3) the action-decision they ultimately make, assessable artefacts are generated that show exactly how knowledge and practice have been integrated by the student.

Validity and reliability of the protocols used for the assessment of the pedagogical outcomes of WIL are functions of the characteristics of the assessment protocols used, and validity in particular depends upon the fit between the protocol and the pedagogical outcome.

The classes of outcomes related to experiencing the work world and developing skills for practice have well-developed protocols that exhibit varying degrees of validity and reliability. Integration of theory and practice is a key outcome class for WIL for which there is emerging focus; it is an area ripe for the development of new protocols and research into their validity and reliability.

REFERENCES

Bates, M. (2003). The assessment of work integrated learning: Symptoms of personal change. *Journal of Criminal Justice Education*, *14*(2), 303–326. Retrieved from http://proquest.umi.com

Bates, M. (2005). *Just practice? Towards a theory of professional education that uses the workplace as context*. School of Criminology and Criminal Justice. Griffith University, Brisbane. Retrieved from http://www4.gu.edu.au:8080/adt-root/public/adt-QGU20060112.141825/index.html

Bates, M. (2008). *Preparing professionals for autonomy workplace-based courses in professional education*. Saarbrucken: VDM Verlag Dr. Muller Aktiengesellschaft & Co. KG.

Biggs, J. (2003). *Teaching for quality learning at university* (p. 309). Buckingham: Society for Research into Higher Education & Open University Press.

Billett, S. (2001). *Learning in the workplace: Strategies for effective practice*. Crow's Nest: Allen and Unwin.

Billett, S. (2004). Workplace participatory practices Conceptualising workplaces as learning environments. *The Journal of Workplace Learning*, *16*(6), 312–324. doi:10.1108/13665620410550295

Billett, S. (2008). Learning throughout working life: a relational interdependence between personal and social agency. *British Journal of Educational Studies*, *56*(1), 39–58.

Boud, D., Keogh, R., & Walker, D. (1985). Reflection: Turning experience into learning. London: Kogan Page.

Boud, D., & Solomon, N. (2001). Work-based learning : A new higher education? Buckinghamshire: Society for Research into Higher Education & Open University Press.

Boud, D., & Walker, D. (1991). *Experience and learning: Reflection at work*. Melbourne: Deakin University Press.

Cannon, L. (1999). *Official negligence: How Rodney King and the riots changed Los Angeles and the LAPD*. Boulder, Colorado: Westview Press.

Cheetham, G., & Chivers, G. (2000). A new look at competent professional practice. *Journal of European Industrial Training*, *24*(7), 374–383.

Dalton, M., Davidson, M., & Keating, J. (2011). The assessment of physiotherapy practice (APP) is a valid measure of professional competence of physiotherapy students: a cross-sectional study with Rasch analysis. *Journal of Physiotherapy*, *57*(4), 239–46. doi:10.1016/S1836-9553(11)70054-6

Dewey, J. (1939). *Experience and education*. New York: Collier Book.

Epstein, R. M., & Hundert, E. M. (2014). Defining and assessing professional competence. *Journal of the American Medical Association (JAMA)*, *287*(2), 226–235.

Girot, E. A. (1993). Assessment of competence in clinical practice - a review of the literature. *Nurse Education Today*, *13*(2), 83–90.

Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development.* . Englewood Cliffs, N.J. : Prentice-Hall, Inc.,.

McAllister, S., Lincoln, M., Ferguson, A., & McAllister, L. (2010). Issues in developing valid assessments of speech pathology students' performance in the workplace. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *45*(1), 1–14. doi:10.3109/13682820902745461

McAllister, S., Lincoln, M., Ferguson, A., & McAllister, L. (2011). A systematic program of research regarding the assessment of speech-language pathology competencies. *International Journal of Speech-Language Pathology*, *13*(6), 469–79. doi:10.3109/17549507.2011.580782

Orrell, J. (2011). *Good practice report: Work-integrated learning*. Sydney: Australian Learning and Teaching Council.

Patrick, C. J., Peach, D., Pocknee, C., Webb, F., Fletcher, M., & Pretto, G. (2009). *The WIL (Work Integrated Learning) report: A national scoping study - Australian Learning and Teaching Council [ALTC] Final report*. Brisbane: Queensland University of Technology.

Polanyi, M. (1966). *The tacit dimension*. Chicago IL: University of Chicago Press.

Schon, D. (1983). *The Reflective Practioner: How professionals think in action*. New York: Basic Books.

Shanteau, J., Weiss, D., Thomas, R., & Pounds, J. (2003). How can you tell if someone is an expert? Empirical Assessment of Expertise. In *Emerging Perspectives on Judgment and Decision Research* (pp. 620–641). Cambridge, U.K.: Cambridge University Press.

Smith, C. D., & Worsfold, K. (2013a). Unpacking the learning–work nexus: "Priming" as lever for high-quality learning outcomes in work-integrated learning curricula. *Studies in Higher Education*, (November 2013), 1–21. doi:10.1080/03075079.2013.806456

Smith, C. D., & Worsfold, K. (2013b). WIL curriculum design and student learning: a structural model of their effects on student satisfaction. *Studies in Higher Education*, (ahead-of-print), 1–15.

Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: the case for explicitly developing students "assessment literacy." *Assessment & Evaluation in Higher Education*, *38*(1), 44–60. doi:10.1080/02602938.2011.598636

Wass, V., Vleuten, C. Van Der, Shatzer, J., & Jones, R. (2001). Medical education quartet assessment of clinical competence. *The Lancet*, *357*, 945–949.

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors*, *45*(1), 104–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12916584

Asia-Pacific Journal of Cooperative Education
ISSN 1175-2882
www.apjce.org

About the Journal

The Asia-Pacific Journal of Cooperative Education publishes peer-reviewed original research, topical issues, and best practice articles from throughout the world dealing with Cooperative Education (Co-op) and Work Integrated Learning/Education (WIL).

In this Journal, Co-op/WIL is defined as an educational approach that uses relevant work-based projects that form an integrated and assessed part of an academic program of study (e.g., work placements, internships, practicum). These programs should have clear linkages with, or add to, the knowledge and skill base of the academic program. These programs can be described by a variety of names, such as work-based learning, workplace learning, professional training, industry-based learning, engaged industry learning, career and technical education, internships, experiential education, experiential learning, vocational education and training, fieldwork education, and service learning.

The Journal's main aim is to allow specialists working in these areas to disseminate their findings and share their knowledge for the benefit of institutions, co-op/WIL practitioners, and researchers. The Journal desires to encourage quality research and explorative critical discussion that will lead to the advancement of effective practices, development of further understanding of co-op/WIL, and promote further research.

Submitting Manuscripts

Before submitting a manuscript, please unsure that the 'instructions for authors' has been followed (www.apjce.org/instructions-for-authors). All manuscripts are to be submitted for blind review directly to the Editor-in-Chief (editor@apjce.org) by way of email attachment. All submissions of manuscripts must be in MS Word format, with manuscript word counts between 3,000 and 5,000 words (excluding references).

All manuscripts, if deemed relevant to the Journal's audience, will be double blind reviewed by two reviewers or more. Manuscripts submitted to the Journal with authors names included with have the authors' names removed by the Editor-in-Chief before being reviewed to ensure anonymity.

Typically, authors receive the reviewers' comments about a month after the submission of the manuscript. The Journal uses a constructive process for review and preparation of the manuscript, and encourages its reviewers to give supportive and extensive feedback on the requirements for improving the manuscript as well as guidance on how to make the amendments.

If the manuscript is deemed acceptable for publication, and reviewers' comments have been satisfactorily addressed, the manuscript is prepared for publication by the Copy Editor. The Copy Editor may correspond with the authors to check details, if required. Final publication is by discretion of the Editor-in-Chief. Final published form of the manuscript is via the Journal website (www.apjce.org), authors will be notified and sent a PDF copy of the final manuscript. There is no charge for publishing in APJCE and the Journal allows free open access for its readers.

Types of Manuscripts Sought by the Journal

Types of manuscripts the Journal accepts are primarily of two forms; *research reports* describing research into aspects of Cooperative Education and Work Integrated Learning/Education, and *topical discussion* articles that review relevant literature and give critical explorative discussion around a topical issue.

The Journal does also accept *best practice* papers but only if it present a unique or innovative practice of a Co-op/WIL program that is likely to be of interest to the broader Co-op/WIL community. The Journal also accepts a limited number of *Book Reviews* of relevant and recently published books.

Research reports should contain; an introduction that describes relevant literature and sets the context of the inquiry, a description and justification for the methodology employed, a description of the research findings-tabulated as appropriate, a discussion of the importance of the findings including their significance for practitioners, and a conclusion preferably incorporating suggestions for further research.

Topical discussion articles should contain a clear statement of the topic or issue under discussion, reference to relevant literature, critical discussion of the importance of the issues, and implications for other researchers and practitioners.

Asia-Pacific Journal of Cooperative Education
ISSN 1175-2882
www.apjce.org

EDITORIAL BOARD