Quality of WIL assessment design in higher education: a systematic literature review

Michelle Lasen[a], Snowy Evans[b], Komla Tsey[c], Claire Campbell[d] and Irina Kinchin[e]

[a]Learning, Teaching and Student Engagement, James Cook University, Cairns, Australia

[b]College of Arts, Society and Education, James Cook University, Cairns, Australia

[c]College of Arts, Society and Education and Cairns Institute, James Cook University, Cairns, Australia

[d]College of Arts, Society and Education, James Cook University, Townsville, Australia

[e]Centre for Indigenous Health Equity Research, Central Queensland University, Cairns, Australia

CONTACT. Michelle Lasen. Email: Michelle.Lasen@jcu.edu.au

ABSTRACT

We investigated the quality of work-integrated learning (WIL) assessment design, in higher education programs, through review of peer-reviewed studies published internationally and in English, 1990–2015. Such a review is timely in light of vested interest from a range of WIL stakeholders; high-level endorsement of WIL across university programs; a regulatory environment requiring development and assurance of higher-order learning outcomes; and a WIL assessment literature that identifies a number of challenges and opportunities. We searched six electronic databases, yielding 20 intervention studies that met inclusion criteria. Findings reveal high-quality assessment design, albeit a need for greater involvement of industry/professional partners in assessment practices, and stronger alignment between reflective activities and students' WIL experiences. The evidence base under review largely comprised qualitative and mixed methods studies, with indication that the quality of study design had improved over time, although variably across disciplinary fields. The key recommendation from this review is that resources are needed to support research-active WIL academics and partners and students: a). to design and participate in assessment practices, which promote integration of student learning, across university and work settings, and achievement of higher-order learning outcomes; and b). to pursue a collaborative research agenda involving robust evaluation research, inclusive of quantitative studies.

Keywords: Assessment; systematic literature review; work-integrated learning

Introduction

Student learning, reflection, assessment and stakeholder relationships are identified foci for further WIL research (Zegwaard & Coll, 2011). WIL is 'an umbrella term for a range of approaches and strategies that integrate theory with the practice of work within a purposefully designed curriculum' (Patrick et al., 2009, p. 9). While international usage now favours the term WIL over cooperative education to describe the field, there is a proliferation of terms relating to WIL practices (Zegwaard, 2015). WIL occurs in a range of off- and on-campus settings through practicum, placement, professional experience, professional practice, internship, workplace learning, industry-based learning, project-based learning, fieldwork education, service learning, real world learning and experiential learning (AWPA, 2014; Patrick et al., 2009). WIL practices vary and evolve according to disciplinary and professional requirements and contexts (Bosco & Ferns, 2014; Hodges, Eames, & Coll, 2014). In contrast with longstanding and highly-regulated components of professionally accredited programs, recent WIL practices have emerged in largely unregulated contexts and are often the 'genesis of innovative and experimental design' (Orrell, 2011, p. 6). WIL tasks may involve 'simulations, case studies, role plays, ePortfolios, reflective journals, project work, mentoring from industry partners, and work-related presentations' (Bosco & Ferns, 2014, p. 285).

Increasingly, there are imperatives to pursue a WIL agenda across non-traditional WIL disciplines. Australia's Chief Scientist (Australian Government Office of the Chief Scientist, 2015) recently called for the embedding of industry placements and projects in all STEM degrees. Students have a growing expectation that a university qualification will equip them for the world of work (Smith, 2012). Employers, too, call for curricula that promote work readiness, as well as address skills shortages and realise productivity outcomes (Gamble, Patrick, & Peach, 2010; Universities Australia, 2008). For university academics, WIL is a potentially effective community engagement strategy and pedagogy, involving complex learning and reflective processes that promote the development of technical and transferable skills, as well as professional identities and values (Brown, 2010).

It is prudent that the sector gathers and reviews evidence of an educational agenda in which WIL provides 'qualitatively different' learning, assessment and feedback opportunities (Orrell, 2011, p. 8). Internationally, there are calls for reappraisal of assessment policy and practice (Boud & Falchiko, 2006; Crisp, 2012; Higher Education Academy, 2012). In Australian universities, assessment reform has ensued given the recently implemented *Higher Education Standards* (Australian Government DET, 2015), which require progressive

development and assurance of program-level learning outcomes targeting disciplinary-specific and generic knowledge, skills and applications, as well as capabilities required for successful transition to the workplace, further study and lifelong learning. Outcomes of this nature can only be realised through robust assessment profiles (Bosco & Ferns, 2014). Traditionally, higher education assessment has tended to focus on knowledge and conceptual understanding. However, a critical WIL curriculum has the potential to bridge the theory–practice divide and promote higher-order learning outcomes. By engaging in and reflecting on the complexity and ambiguity of real world practice, students can generate *new* understandings, skills and perspectives (Smith, 2012).

Nonetheless, assessment remains 'one of the biggest challenges in designing WIL programs' (Orrell, 2011, p. 9). The need to balance different stakeholder expectations and intended outcomes can lead to students perceiving WIL assessment as onerous and largely fulfilling compliance purposes rather than promoting learning and reflection (Patrick et al., 2009; Peach, Ruinard, & Webb, 2014). Further, while universities typically retain responsibility for WIL assessment given time and resource constraints, student learning outcomes are variably impacted by supervisor–student relationships, workplace dynamics and the levels of support provided (Fleming, 2015; Hodges et al., 2014). According to Zegwaard (2015), there is a 'pressing need for further work to develop truly authentic, robust, reliable and defendable assessment practices that measure and inform student learning whilst participating in WIL' (p. 94).

Our aim was to investigate the quality of WIL assessment design in higher education programs, through review of peer-reviewed intervention studies published internationally and in English, 1990–2015. In order to appraise the quality of WIL assessment design, we adopted Bosco and Ferns' (2014) *Authentic Assessment Framework* (AAF). The AAF was specifically developed to evaluate the 'potential veracity, range and relevancy' of WIL tasks within university programs (p. 282). It comprises four criteria:

1) the student is actively engaged in a workplace setting or with an authentic audience
2) the student is required to demonstrate high-quality intellectual engagement (i.e., analysing, evaluating, creating, performance enactment)
3) the student reflexively evaluates performance
4) industry contributes to assessment (E.g., establishment of marking criteria, direct marking).

These criteria align with features of other literature-informed assessment frameworks (Boud & Falchikov, 2006; Gulikers, Bastiaens, & Kirschner, 2004; Herrington & Herrington, 2006). In

their authentic assessment framework, Gulikers et al. (2004) also afforded consideration to the physical and social aspects of the context within which the task is undertaken (Criterion 1 in the AAF); the intellectual/cognitive and metacognitive requirements of the task (Criteria 2 and 3); and the criteria and standards that are applied (related to Criterion 4). There is also alignment with Boud and Falchikov's (2006) assessment practices for *longer-term learning*; in particular, those practices that emphasise the importance of context; involve authentic representations and productions; promote student agency; foster reflexivity; and allow students to identify, develop and engage with criteria and standards. While not exhaustive, the four criteria of the AAF comprise an evidence-based evaluative lens.

While recent systematic reviews investigated research quantity, diversity and quality of WIL quantitative studies (Bartkus, 2007) and WIL qualitative studies (Coll & Kalnins, 2009), it is important to highlight that there is *no* existing systematic literature review of the quality of assessment design in WIL. A systematic literature review offers a rigorous and transparent method for identifying, analysing and synthesising a body of research. It is an effective means for assessing the existing state of a diverse and dispersed field, informing practice guidelines, and identifying research gaps and future directions (Pickering & Byrne, 2014; Shamseer et al., 2015; Tranfield, Denyer, & Smart, 2003). We anticipate that this review will provide stakeholders with meaningful, transparent information to inform design of WIL assessment and related research. We addressed three research questions:

1) What are the characteristics of the included studies?
2) What is the quality of WIL assessment design?
3) What is the quality of study design of the included studies?

The following sections present method, results, discussion and research limitations and conclusions.

Method

An essential component of a trustworthy systematic review is a protocol that pre-defines the rationale, research questions and review methods, including eligibility criteria, search strategy and justification of study quality and reliability (Bearman et al., 2012; Shamseer et al., 2015). We followed the PRISMA and PRISMA-P guidelines (Shamseer et al., 2015) to develop a protocol for rigorous interrogation of the evidence base.

Eligibility criteria

A study was included if it was an 'intervention study' (Bailey et al., 2009) published globally, in a peer-reviewed English-language journal between January 1990 and December 2015, with a central focus on WIL assessment in higher education programs. An intervention study involved an assessment task (i.e., formative or summative, as defined by Sadler, 1989), assessment program, initiative or approach implemented in a WIL context. Theoretical papers focussing on WIL assessment practices or papers focussing on the development of an assessment scale, instrument or rubric in isolation of WIL assessment practices were excluded (Figure 1).

Search strategy and study selection

The WIL field traverses disciplines and is described by a range of strategies and terms. Relevant publications are located across a range of journals, indexed in various databases. To compile a list of suitable databases and keywords, three of the researchers undertook preliminary searches and cross-checks in consultation with the university social sciences liaison librarian. The search strategy was designed to capture all studies that met the eligibility criteria, taking into account nuances of different databases. Databases included Educational Research Abstracts, ERIC via Proquest, A+ Education via Informit, Web of Science, Proquest, and Sage Journals. Key search words (Figure 1), informed by the most frequently used WIL terms identified by Patrick et al. (2009) and other sources (E.g., AWPA, 2014; Bartkus, 2007), capture a relatively wide description of WIL. Study selection involved two researchers in database searching (Figure 1, Step 1) and duplicate removal (Step 2). Abstracts (n=400) were screened to determine inclusion or exclusion (Step 3). Where abstracts met eligibility criteria, full papers (n=240) were read (Step 4). Disagreement about inclusion of studies was resolved through discussion between all researchers at this step.

Data classification and review

Data from included studies (n=102) were organised in Excel sheets with pre-determined headings (Step 5). Four researchers classified studies according to: research type (intervention; 'other' excluded); originating country or continent of the first author; field of education (Australian Government DET, n.d.); and WIL referent in study's title (Step 5; inter-rater reliability check on 20% sample). Intervention studies (n=20) were retained for examination and further classified according to: intervention type (*description* of an intervention or *evaluation* of an intervention, as per Bailey et al., 2009) and broad research type (qualitative, quantitative or mixed methods) (Step 6). The fifth researcher moderated any classification

differences. Two researchers appraised the intervention studies for quality of assessment design against each of the four criteria of the AAF (Bosco & Ferns, 2014), scoring as follows: 0=Article includes no evidence; 1=Article includes implicit evidence/limited explicit evidence; 2=Article includes explicit evidence (Step 6; inter-rater reliability check on 50% sample). Two researchers appraised the evaluation studies (n=18) for study design quality, utilising the *Qualitative Research Checklist* (Critical Appraisal Skills Programme [CASP], 2013) and *Quality Assessment Tool for Quantitative Studies* (Effective Public Health Practice Project [EPHPP], 2009) (Step 7; inter-rater reliability check on 20% sample).

Results

Note that field of education types (Australian Government DET, n.d.) and disciplines (verbatim from papers) are referred to in this section.

Study characteristics

The geographical scope of first authors (Table 1) was as follows: North America, Australia and New Zealand, United Kingdom, Continental Europe, and South America. The most common fields of education were Health, Education, Society and culture, and Management and commerce. Ten different WIL terms appeared in study titles, with six terms used more than once: internship, experiential learning, work-integrated learning, work-based learning, professional practice, and simulation. Of the intervention studies, 18 were evaluations of interventions and two were descriptions of interventions. Of the evaluation studies, 10 were qualitative and seven employed mixed methods. There was only one quantitative study.

Assessment design quality

Fifteen papers involved students actively engaged in workplace settings (Table 2), scoring a 2 for Criterion 1. A further four papers, scoring a 2, saw students engaged with authentic audiences. For instance, interdisciplinary communication students collaborated with educators, members of a non-profit organisation, architects, architecture students and volunteer expert builders, in Second Life and real life, to design virtual, low-cost, sustainable urban homes (P7).

Fifteen papers involved assessment demanding high-quality intellectual engagement, scoring a 2 for Criterion 2. This subset included all papers from Health and Education, as well ones from non-traditional WIL disciplines (P7; P12; P14; P15; P16). Third-year public policy students negotiated the parameters and assessment criteria of a research project with a public or private agency; undertook the research in a 100-hour internship; and produced a 7500-word

research report. In order for these students to 'conceptualise, synthesise and integrate an assessment process' into their internship experience, they first completed a preparatory subject, designed to develop understanding of assessment and evaluation processes and critical thinking skills (P14, p. 61). Undergraduate sports management students also engaged in a staged experiential learning model, culminating in high-level assessment. They participated in: a) a site visit to an intercollegiate athletic department, interacting with team management and sales staff; b) skills development in the classroom, facilitated by sales staff, course instructors and peer mentors; c) a product knowledge assessment; and d) a five-week, lab-based sales call centre experience (P15). In four papers, scoring a 1, it was evident that reflective tasks needed refinement to maximise students' intellectual engagement. A final paper scored a 1 given that learning outcomes were aligned with 'beginning to medium level intellectual engagement' (Bosco & Ferns, 2014, p. 283). In the paper, first-year social work students, engaging in a role-played interview with service users and carers, were required to 'demonstrate basic communication and interpersonal skills and some understanding of the client's problem' (P19, p. 302).

Eleven papers, including all papers from Education, the majority from Health, as well three from non-traditional WIL disciplines (P7; P14; P19), involved students reflexively evaluating their performance, scoring a 2 on Criterion 3. In four papers, scoring a 1, it was evident that reflective tasks needed to better align with students' WIL experiences. The 'emotive articulations' of commerce students in survey data communicated 'strong resistance to reflective tasks' (P5, p. 111). Law students expressed mixed sentiments, in focus groups, regarding the effectiveness of online forum tasks. There was a sense that responses were somewhat contrived to maximise marks (P13). In two papers, scoring a 1, it was less evident that the *central* aim of reflective tasks was for students to reflexively evaluate their performance (P12; P16). For instance, science students undertaking international ecological research and conservation activities were required to complete post-field reflective tasks, designed primarily to promote integration and application of key scientific concepts (P12). In the final paper, scoring a 1, survey and focus group data showed that, 'reflective behaviour was not sufficiently promoted' among veterinary medicine students in small-group reflective meetings (P1, p. 7).

Only six papers involved industry contribution to assessment, scoring a 2 on Criterion 4. A gerontology internship involved development of mutually-determined learning goals and evaluation criteria, and 'structured and unstructured opportunities for feedback and evaluation from site preceptors, faculty supervisors and students themselves' (P9, pp. 302 & 303). In one of the three papers that scored a 1, clinical supervisors assessed undergraduate veterinary

medicine students formatively yet focus group data revealed dissatisfaction with their lack of influence over summative decisions (P1). Eleven papers scored a 0 on Criterion 4.

Study design quality

Of the 10 qualitative research designs, three were rated moderate and seven were rated weak in study design (Table 3). The one quantitative research paper was rated moderate (Table 4). Of the seven mixed methods publications, all were rated weak for the quantitative components and only one was rated strong for the qualitative component. Qualitative studies or components received unfavourable ratings due to limitations largely regarding research design, recruitment, data collection, details of researcher–participant relationships and ethical considerations, and data analysis. The quantitative components of the mixed methods studies received weak ratings for research design, confounders, blinding, and data collection methods. All five papers of moderate to strong study design quality (P1; P3; P5; P15; P17) clustered in Health and Management and commerce fields and were published from 2010–2015. The majority (5 out of 8) of the papers published in this recent period were of moderate to strong research quality.

Discussion

The evidence base, under review, comprising WIL assessment interventions involved largely qualitative and mixed methods studies. There was indication that study design quality had improved over time, albeit variably across disciplinary fields. Similarly, preliminary research findings from a 2015 review of the broader WIL literature (Hoskyn & Zegwaard) revealed an increase in qualitative and mixed methods studies and an increase in study design quality, over the 2000–2013 review period.

In terms of assessment design quality, the majority of the studies scored highly (i.e., 6 or over out of 8). This subset included all papers but one from Health and Education, as well as those from non-traditional WIL disciplines. The public policy internship, sport management experiential learning model and virtual communications design project – largely university-based WIL experiences – serve as examples of how students can be prepared to participate in high-level assessment through the purposeful design of multi-staged experiential learning, involving engagement with authentic audiences and enabling technologies. High-quality assessment design is an important finding. If assessment does not promote students' learning and engagement, 'it undermines the entire educational enterprise' (Boud, 2010, p. 4).

Only six papers scored 4 or less. Even the lowest-scoring paper displayed merit. Political science students' reflective accounts revealed high-level engagement in a simulated

political debate, wherein students ‘influenced others through cogent reasoning and developed integrative policy solutions’ (P16, p. 332 & 333). While this assessment design scored a 2 for high-quality intellectual engagement (Table 2), attention to the other AAF criteria would enhance its authenticity. The lowest-scoring subset included papers from business, commerce, law and engineering, where there was misalignment between reflective activities and students’ WIL experiences. There was a recognised need for ‘greater inclusion of students’ diverse learning experiences’ and a ‘balance of structured and unstructured reflective activities’ (P5, p. 111).

Relevance, flexibility and feedback were themes in high-scoring papers. For instance, largely mature-age professional instructors, undertaking a formal education course, were required to ‘patch together’ a portfolio of responses to critically-reflective tasks, based on key practice elements, including an ‘open patch’ determined by them or in negotiation with tutors (P4, Table 2). In interviews, students revealed that they were motivated by the element of choice and the opportunity to use theory as a critically reflective lens on practice. Tutors’ feedback on draft patches was found to be instrumental in facilitating the intended shift from ‘descriptive writing to reflective, discursive and analytic writing’ (P4, p. 48). While it was acknowledged that these feedback commitments were onerous and required some level of modification, other papers found senior students to play an important role in providing feedback for junior peers in practice contexts (P1; P15), and dialogue between WIL partners to serve as an important mechanism for feedback and student reflection and learning.

Certainly, students called for opportunities to participate in ‘communicative and reflexive spaces’ with university academics and industry/professional partners (Higgs, 2014, p. 257). Business students felt that a weekly debrief with the university-based internship coordinator and the site supervisor would have been more effective than the existing university-based video journaling tasks (P20). Medical interns showed ‘marked interest’ in extending the duration of assessment meetings, between themselves and university- and clinically-based supervisors, and the scope of practice under review (P2, p. 565). In one high-scoring paper, special education students selected from a range of artefacts and conceptual frameworks, in order to compile a professional practice portfolio, and participated in formative meetings and a summative appraisal process, with the university supervisor and cooperating teacher, to review and discuss progress and, ultimately, verify achievement of competencies (P10).

Without opportunity for dialogic review with other stakeholders and adequate professional development, papers showed a tendency for industry partners to award students high marks and provide limited constructive feedback (P1; P9). A lack of professional

development for clinical supervisors, in the provision of formative feedback, and moderation processes to support panel members' summative assessment decision-making were themes that emerged in a veterinary assessment program evaluation (P1). Community partners participating in the social work role play requested greater guidance regarding standards and structuring feedback; indeed, provision of professional development showed positive impacts in a second-round evaluation of the role-play assessment (P19).

There were papers that showed considerable investment in calibrating assessor and student understanding of assessment processes, requirement and standards (P10; P11). Notably, students were positioned as WIL partners in the gerontology and public policy internships, wherein they negotiated and developed assessment criteria with work- and university-based supervisors (P9; P14) – exemplifying participation in assessment for longer-term learning (Boud & Falchikov, 2006). In summary, robust *assessment partnerships* (Coll & Kalnins, 2009) between key stakeholders, involving opportunities for collaborative design, professional learning, dialogic review and student agency and reflexivity, are vital to developing high-quality WIL assessment practices.

Research limitations and conclusions

While journals are the most reliable and current outlets of research, we recognise that research on WIL assessment is published elsewhere and in languages other than English. Further, while the selected search words encompassed a relatively wide description of the field, they were not an exhaustive set. We also acknowledge that a comprehensive appraisal of assessment design quality would involve review of alignment between assessment tasks and learning outcomes, and the quality of task descriptions, marking criteria/rubrics and moderation strategies (Boud, 2010; Higher Education Academy, 2012). Nonetheless, the application of the AAF (Bosco & Ferns, 2014) was valuable in the context of this systematic review – facilitated by our adoption of a simple scoring system. Importantly, high-quality design of assessment was characteristic of the field, albeit with opportunity for promotion of more robust WIL assessment partnerships. Resourcing and professional development need to support *research-active* WIL academics, from all disciplines, and their professional partners and students a). to design and participate in assessment and reflective practices, which promote integration of students' learning, across university and work settings, and achievement of higher-order learning outcomes; and b). to pursue a collaborative research agenda involving robust evaluation research, inclusive of quantitative studies.

References

Australian Government Department of Education and Training [DET]. (n.d.). Field of education types. Retrieved from http://heimshelp.education.gov.au/sites/heimshelp/resources/pages/field-of-education-types

Australian Government Department of Education and Training [DET]. (2015). *Higher Education Standards Framework. Tertiary Education Quality and Standards Agency Act 2011.* Retrieved from https://www.comlaw.gov.au/Details/C2015C00622

Australian Government Office of the Chief Scientist. (2015). Bridging a gap between study and work. Media Release. Retrieved from http://www.chiefscientist.gov.au/2015/08/media-release-bridging-a-gap-between-study-and-work/

Australian Workforce and Productivity Agency [AWPA]. (2014). *Work integrated learning: AWPA scoping paper*. Retrieved from http://www.awpa.gov.au/our-work/Workforce development/Pages/Work-Integrated-Learning.aspx

Bailey, L., Sanson-Fisher, R., Aranda, S., D'este, C., Sharkey, K., & Scholfield, P. (2009). Quality of life research: Types of publication output over time for cancer patients, a systematic review. *European Journal of Cancer Care, 19*, 1–8.

Bartkus, K. (2007). A review and synthesis of scholarly research in cooperative education and internships: An analysis of quantitative research published outside the Journal of Cooperative Education and Internships. *Journal of Cooperative Education and Internships, 41*(1), 56–96.

Bearman, M., Smith, C., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. (2012). Systematic review methodology in higher education. *Higher Education Research & Development, 31*(5), 625–640.

Bok, H., Teunissen, P., Favier, R., Rietbroek, N., Theyse, L., et al. (2013). Programmatic assessment of competency-based workplace learning: When theory meets practice. *BMC Medical Education, 13*.

Bosco, A., & Ferns, S. (2014). Embedding of authentic assessment in work-integrated learning curriculum. *Asia-Pacific Journal of Cooperative Education, 15*(4), 281–290.

Boud, D. (2010). *Student assessment for learning in and after courses. Final Report for Senior Fellowship.* Australian Learning and Teaching Council.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation, 31*(4), 399–413.

Brown, N. (2010) WIL[ling] to share: An institutional conversation to guide policy and practice in work-integrated learning, *Higher Education Research & Development, 29*(5), 507-518.

Centeno, A., Campos, S., & Primogerio, C. (2004). Addressing ethical issues during assessment of internship. *Medical Education, 38*(5), 565–566.

Clarke, D., Litchfield, C., & Drinkwater, E. (2010). Supporting exercise science students to respond to the challenges of an authentic work-integrated learning assessment. *Asia-Pacific Journal of Cooperative Education, 11*(3), 153–167.

Coll, R., & Kalnins, T. (2009). A critical analysis of interpretive research studies in cooperative education and internships. *Journal of Cooperative Education and Internships, 15*(3), 189–207.

Crisp, G. (2012). Integrative assessment: Reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education, 37*(1), 33–43.

Critical Appraisal Skills Programme. (2013). CASP Checklists. Retrieved from http://www.casp-uk.net/#!casp-tools-checklists/c18f8

Dalrymple, R., & Smith, P. (2008). The patchwork text: Enabling discursive writing and reflective practice on a foundation module in work-based learning. *Innovations in Education and Teaching International, 45*(1), 47–54.

Dean, B., Sykes, C., Agostinho, S., & Clements, M. (2012). Reflective assessment in work-integrated learning: To structure or not to structure, that was our question. *Asia-Pacific Journal of Cooperative Education, 13*(2), 103–113.

Effective Public Health Practice Project. (2009). Quality assessment tool for quantitative studies. Retrieved from http://www.ephpp.ca/tools.html

Fleming, J. (2015). Exploring stakeholders' perspectives of the influences on student learning in cooperative education. *Asia-Pacific Journal of Cooperative Education, 16*(2), 109–119.

Gamble, N., Patrick. C., & Peach, D. (2010). Internationalising work-integrated learning: Creating global citizens to meet the economic crisis and the skills shortage. *Higher Education Research & Development, 29*(5), 535–546.

Griffin, C., & Kilgore, K. (1995). Framing the problems of practice: The effects of self-assessment in a study of special education students' internships. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children, 18*(1), 56–71.

Gulikers, J., Bastiaens, T., & Kirschner, P. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67–86.

Herrington, J., & Herrington, A. (2006). Authentic conditions for authentic assessment: Aligning task and assessment. In A. Bunker & I. Vardi (Eds.), Critical Visions. *Proceedings of the 29th HERDSA Conference* (pp. 146–151). Milperra, NSW: HERDSA.

Higgs, J. (2014). Assessing the immeasurables of practice. *Asia-Pacific Journal of Cooperative Education, 15*(3), 253–267.

Higher Education Academy. (2012). *A marked improvement: Transforming assessment in higher education.* Retrieved from https://www.heacademy.ac.uk/knowledge-hub/marked-improvement

Hodges, D., Eames, C., & Coll, R. (2014). Theoretical perspectives on assessment in cooperative education placements. *Asia-Pacific Journal of Cooperative Education, 15*(3), 189–207.

Hoskyn, K., & Zegwaard, K. (2015). Trends in research methods and approaches in cooperative and work-integrated education research. *WACE World Conference on Cooperative and Work-integrated education*, Sangyo University, Kyoto, Japan, 18–21 August.

Jarmon, L., Traphagan, T., Mayrath, M., & Trivedi, A. (2009). Virtual world teaching, experiential learning, and assessment: An interdisciplinary communication course in Second Life. *Computers & Education, 53*(1), 169–182.

Jones, E. (2013). Practice-based evidence of evidence-based practice: Professional practice portfolios for the assessment of work-based learning. *Quality in Higher Education, 19*(1), 56–71.

Karasik, R. (2009). Learning from internships in gerontology and geriatrics: Assessment and program evaluation. *Gerontology & Geriatrics Education, 30*(4), 300–316.

Kossar, K. (2003). Graduate practicum – Special education: Assessment through portfolio Development. *Teacher Education and Special Education, 26*(2), 145–149.

Levett-Jones, T., Gersbach, J., Arthur, C., & Roche, J. (2011). Implementing a clinical competency assessment model that promotes critical reflection and ensures nursing graduates' readiness for professional practice. *Nurse Education in Practice, 11*(1), 64–69.

McLaughlin, J., & Johnson, D. (2006). Assessing the field course experiential learning model: Transforming collegiate short-term study abroad experiences into rich learning environments. *Frontiers: The Interdisciplinary Journal of Study Abroad, 13*.

McNamara, J., & Brown, C. (2009). Assessment of online discussion in work-integrated learning. *Campus-Wide Information Systems, 26*(5), 413–423.

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

O'Toole, K. (2007). Assessment in experiential learning: The case of a public policy internship. *Education Research and Perspectives, 34*(2), 51–62.

Patrick, C., Peach, D., Pocknee, C., Webb, F., Fletcher, M., & Pretto, G. (2009). *The Work Integrated Learning report: A national scoping study.* Brisbane: Queensland University of Technology.

Peach, D., Ruinard, E., & Webb, F. (2014). Feedback on student performance in the workplace: The role of the workplace supervisors. *Asia-Pacific Journal of Cooperative Education, 15*(3), 239–249.

Pickering, C., & Byrne, J. (2014). The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research & Development, 33*(3), 534–548.

Pierce, D., Petersen, J., & Meadows, B. (2011). Authentic assessment of experiential learning in sport sales. *Sport Marketing Quarterly, 20*(2), 75–83.

Rackaway, C., & Goertzen, B. J. (2008). Debating the future: A social security political leadership simulation. *Journal of Political Science Education, 4*(3), 330–340.

Ramm, D., Thomson, A., & Jackson, A. (2015). Learning clinical skills in the simulation suite: The lived experiences of student nurses involved in peer teaching and peer assessment. *Nurse Education Today, 35*(6), 823–827.

Rompelman, O., & De Vries, J. (2002). Practical training and internships in engineering education: Educational goals and assessment. *European Journal of Engineering Education, 27*(2), 173–180.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ*, 349, g7647.

Skilton, C. (2011). Involving experts by experience in assessing students' readiness to practise: The value of experiential learning in student reflection and preparation for practice. *Social Work Education, 30*(3), 299–311.

Smith, C. (2012). Evaluating the quality of work-integrated learning curricula: A comprehensive framework. *Higher Education Research & Development, 31*(2), 247–262.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management, 14*, 207–222.

Universities Australia. (2008). A national internship scheme: Enhancing the skills and work-readiness of Australian university graduates. *Position Paper 3/08.* Canberra: Author.

Wilkinson, K. (2008). Using Breeze for communication and assessment of internships: An exploratory study. *Journal of Educators Online, 5*(2), 15.

Zegwaard, K. (2015). Building an excellent foundation for research: Challenge and current research methods. *Asia-Pacific Journal of Cooperative Education, 16*(2), 89–99.

Zegwaard, K., & Coll, R. (2011). Exploring some current issues for cooperative education. *Journal of Cooperative Education and Internships, 45*(2), 8–15.

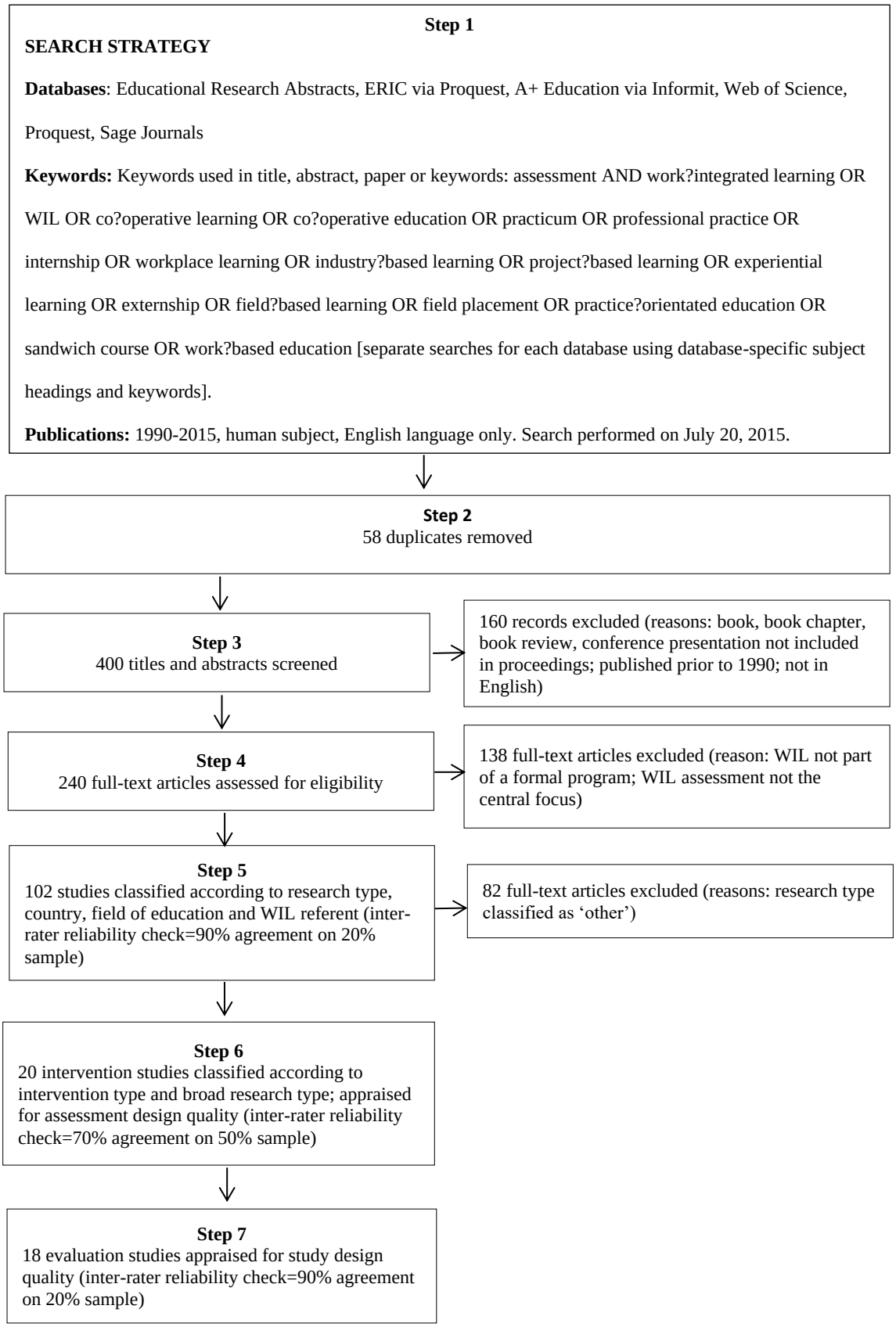

Figure 1. Study selection log

Table 1. Characteristics of assessment in WIL interventions

Paper	First author, year	Country/continent of first author	Field of education type	WIL referent in paper title	Intervention type	Study design
P1	Bok, 2013	Netherlands	Health	Workplace learning	Evaluation	Mixed methods
P2	Centeno, 2004	South America	Health	Internship	Description	N/A
P3	Clarke, 2010	Australia	Health	Work-integrated learning	Evaluation	Qualitative
P4	Dalrymple, 2008	United Kingdom	Education	Work-based learning	Evaluation	Qualitative
P5	Dean, 2012	Australia	Management & commerce	Work-integrated learning	Evaluation	Qualitative
P6	Griffin, 1995	North America	Education	Internship	Evaluation	Mixed methods
P7	Jarmon, 2009	North America	Information technology	Experiential learning	Evaluation	Mixed methods
P8	Jones, 2013	New Zealand	Education	Work-based learning; professional practice	Evaluation	Mixed methods
P9	Karasik, 2009	North America	Health	Internship	Evaluation	Qualitative
P10	Kossar, 2003	North America	Education	Practicum	Evaluation	Qualitative
P11	Levett-Jones, 2011	Australia	Health	Professional practice	Evaluation	Mixed methods
P12	McLaughlin, 2006	North America	Natural & physical sciences	Experiential learning	Evaluation	Mixed methods
P13	McNamara, 2009	Australia	Society & culture	Work-integrated learning	Evaluation	Mixed methods
P14	O'Toole, 2007	Australia	Society & culture	Experiential learning; internship	Description	N/A
P15	Pierce, 2011	North America	Management & commerce	Experiential learning	Evaluation	Quantitative
P16	Rackaway, 2008	North America	Society & culture	Simulation	Evaluation	Qualitative
P17	Ramm, 2015	United Kingdom	Health	Simulation	Evaluation	Qualitative
P18	Rompelman, 2002	Netherlands	Engineering & related technologies	Practical training; internship	Evaluation	Qualitative
P19	Skilton, 2011	United Kingdom	Society & culture	Experiential learning	Evaluation	Qualitative
P20	Wilkinson, 2008	North America	Management & commerce	Internship	Evaluation	Qualitative

Table 2: Appraisal of assessment design quality, using the AAF (Bosco & Ferns, 2014)

Paper	Criterion 1. Student actively engaged in a workplace setting or with authentic audience	Criterion 2. Student required to demonstrate high-quality intellectual engagement	Criterion 3. Student reflexively evaluates performance	Criterion 4. Industry contributes to assessment	Overall score
P2	2 Engaged in medical internship	2 Participated in small-group assessment meetings focusing on clinical practice	2 Reflected upon ethical problems, identifying principles and courses of action	2 University and clinically-based supervisors conducted meetings to assess achievement of outcomes	8
P6	2 Engaged in special education internship	2 Planned, implemented and evaluated teaching units	2 Provided written responses to probing questions and participated in post-observation conferences	2 Cooperating teacher completed performance reviews and met with student and university supervisor	8
P9	2 Engaged in gerontology internship	2 Devised learning objectives and evaluation methods and fulfilled learning contract	2 Reflected on activities and learnings, in weekly journal reports, and knowledge and skills and identifiable gaps, in final report	2 Faculty supervisor assigned grades based on site preceptor's assessments and student's written work	8
P10	2 Engaged in special education practicum	2 Evidenced achievement of competencies in portfolio	2 Compiled artefacts and reflections in portfolio and completed Performance Evaluation and Appraisal instrument	2 University supervisor, cooperating teacher and student validated competency attainment	8
P14	2 Engaged in public policy internship	2 Negotiated research project with supervisors, implemented action plans, and produced report	2 Drew upon reflections in learning journal to analyse workplace challenges and actions	2 Workplace and academic supervisors monitored drafts and assessed final report	8
P11	2 Engaged in nursing clinical practice	2 Participated in full-day, holistic clinical competence assessment	2 Responded to questions regarding clinical practice, reflected on feedback, and negotiated strategies for improvement	1 Assessors were highly-qualified registered nurses employed by the university	7

P1	2 Engaged in veterinary medicine rotations	2 Undertook formative tasks and evidenced achievement of competencies in digital portfolio	1 Reflected on feedback to analyse strengths and weaknesses and participated in small group reflective sessions	1 Clinical supervisors assessed formatively and facilitated small group sessions to set learning goals	6
P3	2 Engaged in exercise science professional placement	2 Designed, implemented and evaluated action research project	2 Posted fortnightly reflective blogs and produced evaluation report	0	6
P4	2 Engaged in professional education	2 Compiled critically reflective Patchwork Text and integrating summary	2 Reflected upon experiences as learner and teacher and practice	0	6
P7	2 Engaged with range of stakeholders	2 Collaborated to virtually design urban model homes	2 Participated in discussions and reflected on experiences and emerging views regarding interdisciplinary communication	0	6
P8	2 Engaged in professional education	2 Evidenced achievement of graduate learning objectives by selecting and compiling artefacts in portfolio	2 Reflected on artefacts in relation to personal philosophy and literature	0	6
P15	2 Engaged with prospective clients	2 Participated in written product knowledge assessment and sales calling	0	2 Sports sales expert assessed student's sales calls	6
P19	2 Engaged with service users and carers	1 Demonstrated basic communication and interpersonal skills in interview role play	2 Reviewed filmed role plays to reflect on strengths and weaknesses	1 Service users and carers and module leaders provided feedback on interviews and module leaders decided grades	6
P12	2 Engaged in science fieldwork	2 Participated in pre-field tasks, ecological research and conservation work and post-field tasks	1 Documented observations in field journal and undertook post-field reflective tasks	0	5

P5	2 Engaged in commerce internship	1 Reflected on experiences in daily eLog, responded to modules, and compiled journal focusing on skills development and future actions	1 *Reflective tasks needed refinement to align with WIL experiences*	0	4
P13	2 Engaged in law work placement	1 Contributed fortnightly reflective posts in online forum on numerous topics	1 *Reflective tasks needed refinement to align with WIL experiences*	0	4
P17	2 Engaged with first-year nursing students	2 Taught and assessed first-year students' clinical nursing skills in simulation suite	0	0	4
P18	2 Engaged in international engineering internship	1 Produced self-evaluation report and participated in debrief with teacher regarding professional and cultural insights	1 *Reflective tasks needed refinement to align with WIL experiences*	0	4
P20	2 Engaged in business internship	1 Responded to weekly questions in video journal	1 *Reflective tasks needed refinement to align with WIL experiences*	0	4
P16	0	2 Adopted stakeholder perspectives and developed solutions to problems in simulation of political debate	1 Participated in debriefing session and reflected on key concepts in post-simulation paper	0	3

Table 3. Appraisal of study design quality, using the Qualitative Research Checklist (CASP, 2013)

Paper	Clear statement of research	Qualitative methodology appropriate	Research design appropriate for aims	Recruitment strategy for aims	Data collection addresses research issue	Researcher–participant relationship considered	Ethical considerations accounted for	Rigorous data analysis	Clear statement of findings	Research is valuable	Overall score
Qualitative study											
P3	Yes	Yes	Yes	Yes	Yes	No	Yes	Can't tell	Yes	Yes	Moderate
P4	Yes	Yes	No	Can't tell	Can't tell	No	No	No	Yes	No	Weak
P5	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Moderate
P9	Yes	Yes	Can't tell	Yes	Yes	No	No	Yes	Yes	Yes	Weak
P10	Yes	Yes	Can't tell	Can't tell	Can't tell	No	No	No	Yes	No	Weak
P16	Yes	Yes	No	No	No	No	No	No	Yes	No	Weak
P17	Yes	Yes	Yes	Yes	Yes	No	Yes	Can't tell	Yes	Yes	Moderate
P18	Yes	Yes	No	No	No	No	No	No	Yes	No	Weak
P19	Yes	Yes	Can't tell	Can't tell	Yes	No	No	Can’t tell	Yes	Yes	Weak
P20	Yes	Yes	Can't tell	Can't tell	Yes	No	No	No	Yes	Yes	Weak
Qualitative component (mixed method study)											
P1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Strong
P6	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Weak
P7	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	Weak
P8	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Weak
P11	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes	Weak
P12	Yes	Yes	Can't tell	No	Yes	No	No	No	Yes	Yes	Weak
P13	Yes	Yes	Yes	No	Yes	No	No	No	Yes	Yes	Weak

Table 4. Appraisal of study design quality, using the Quality Assessment Tool for Quantitative Studies (EPHPP, 2009)

Paper	Selection bias	Research design	Confounders	Blinding	Data collection methods	Withdrawals and dropouts	Intervention integrity	Analyses	Overall score
Quantitative study									
P15	Q1. 1 Q2. 1 Strong	Moderate	Q1. 1 Q2. Strong	Q1. 1 Q2. 3 Weak	Q1. 1 Q2. 1 Strong	Q1. 2 Q2. 1 Strong	Q1. 4 Q2. 3 Q3. 6	Q1. Individual Q2. Individual Q3. 1 Q4. 1	Moderate
Quantitative component (mixed method study)									
P1	Q1. 1 Q2. 5 Moderate	Weak	Q1. 3 Weak	Q1. 1 Q2. 3 Weak	Q1. 3 Q2. 3 Weak	Q1. 4	Q1. 4 Q2. 3 Q3. 6	Q1. Individual Q2. Individual Q3. 3 Q4. 3	Weak
P6	Q1. 1 Q2. 5 Moderate	Weak	Q1. 3 Weak	Q1. 1 Q2. 3 Weak	Q1. 3 Q2. 3 Weak	Q1. 3 Q2. 4 Weak	Q1. 4 Q2. 3 Q3. 6	Q1. Individual Q2. Individual Q3. 1 Q4. 1	Weak
P7	Q1. 1 Q2. 5 Moderate	Weak	Q1. 3 Weak	Q1. 1 Q2. 3 Weak	Q1.3 Q2. 3 Weak	Q1. 3 Q2. 4 Weak	Q1. 4 Q2. 3 Q3. 6	Q1. Individual Q2. Individual Q3. 1 Q4. 1	Weak
P8	Q1. 1 Q2. 5 Moderate	Moderate	Q1. 3 Weak	Q1. 1 Q2. 3 Weak	Q1. 3 Q2. 3 Weak	Q1. 3 Q2. 4 Weak	Q1. 4 Q2. 3 Q3. 6	Q1. Individual Q2. Individual Q3. 3 Q4. 3	Weak
P11	Q1. 1 Q2. 5 Moderate	Weak	Q1. 3 Weak	Q1. 3 Q2. 3 Weak	Q1. 3 Q2. 3 Weak	Q1. 4	Q1. 4 Q2. 3 Q3. 6	Q1. Individual Q2. Individual Q3. 3 Q4. 3	Weak
P12	Q1. 1 Q2. 5 Moderate	Weak	Q1. 3 Weak	Q1. 3 Q2. 3 Weak	Q1. 3 Q2. 3 Weak	Q1. 4	Q1. 4 Q2. 3 Q3. 6	Q1. Individual Q2. Individual Q3. 3 Q4. 3	Weak
P13	Q1. 1 Q2. 5 Moderate	Weak	Q1. 3 Weak	Q1. 3 Q2. 3 Weak	Q1. 3 Q2. 3 Weak	Q1. 4	Q1. 4 Q2. 3 Q3. 6	Q1. Individual Q2. Individual Q3. 3 Q4. 3	Weak