



## A Work-Integrated Learning Framework: the what, where and how of evaluating WIL

Karen Young<sup>1</sup>, Sophie McKenzie<sup>1</sup>, and Jill Thomas<sup>2</sup>

Corresponding author: Sophie McKenzie ([sophie.mckenzie@deakin.edu.au](mailto:sophie.mckenzie@deakin.edu.au))

<sup>1</sup> Faculty of Science, Engineering and Built Environment, Deakin University, Geelong, Victoria, Australia

<sup>2</sup> J.A Thomas and Associates, Melbourne, Victoria, Australia

### Abstract

Through a case study at Deakin University, we present our approach for evaluating Work-Integrated Learning (WIL) during a pilot of a WIL Evaluation Framework (WEF). This paper focuses on our approaches and decisions relating to what, where and how to evaluate WIL, offering considerations for a sector-wide approach to WIL evaluation more generally. The findings from the pilot inform ongoing institution-specific WIL evaluation projects. We anticipate that our study will provide insight into the comprehensive processes involved in WIL evaluation, and the impact findings can make to inform WIL and employability strategies to supplement current sector-wide discussions on WIL indicators and measures of success for graduate outcomes.

### Keywords

Work-Integrated Learning; WIL evaluation; employability; Higher Education

### Introduction

This paper presents, as a case study, an investigation of a fit-for-purpose WIL Evaluation Framework (WEF) for the organisational unit of the Faculty of Science, Engineering and Built Environment (SEBE) at Deakin University (Young et al., 2023). Within SEBE there are four discipline areas corresponding to the four schools that delivery undergraduate and postgraduate courses: School of Information Technology, School of Life and Environmental Science, School of Engineering and School of Architecture and Built Environment. This work was undertaken as part of an action-research project titled: 'WIL Evaluation Project' (WEP) as a part of the Faculty's *Employability Guiding Plan* (Young et al., 2022). The SEBE *Employability Guiding Plan* (SEGP) was a strategy to provide a scaffolded, embedded, multiple and varied WIL opportunities (multiplicity-WIL) across 21 undergraduate courses in the Faculty. For the strategic aspiration of enhanced employability to gain traction across all undergraduate programs, twelve Faculty and school-based projects were endorsed. 80 Deakin staff (academic and professional) were directly involved across the 12 curriculum-based projects during 2021-2022, resulting in the re-vision of courses. The WIL Evaluation Project had three deliverables:

1. design of a bespoke Evaluation Framework capable of measuring the impact of holistic WIL offerings within the Faculty;

2. trial the effectiveness of the WIL Evaluation Framework by undertaking a test-case evaluation;
3. recommend an evaluation process and possibly a refined framework to enable a sustainable WIL evaluation for the Faculty (Young et al., 2023).

This case study presents findings relating to the first of these objectives – the effectiveness of the design of the WIL Evaluation Framework. The paper will conclude with outcomes relating to the testing phase as a way of informing ongoing work and operations to ensure sustainability, noting that the data arising in the trial of the test-case is out of scope for this study (Young et al., 2023; Young, 2023).

As background, the WEP conceptualised six guiding questions, referred to as the 6W's of WIL evaluation (WE) – the Why, Who, Where, What, When and hoW (Young et al., 2023). These were instrumental in guiding a process of evaluation that measures what matters - to the evaluators and the stakeholders of the evaluation, as well as the stakeholders of WIL more broadly. The application and the examination of the 6W's of WE (Young et al., 2023) is confined in this paper to the What, Where and hoW. The original guiding questions for these three considerations are shown in Table 1.

**Table 1. Guiding Questions for Evaluating Work-Integrated Learning**

<p>3. <b>What</b> is being evaluated?</p>	<p><b>Guiding Sub-questions</b></p> <p>3.1 What is the grounding approach taken to evaluate?</p> <p>3.2 What are the WIL activity types included in the WEF?</p> <p>3.3 What are the chosen indicators for success?</p> <p>3.4 What are the data types/tools used to evaluate the activities?</p>	<p>Original question:</p> <p>What are the key underpinnings, considerations, requirements, and deliverables for WIL evaluation?</p>
<p>4. <b>Where</b> can we find the data?</p>	<p><b>Guiding Sub-questions</b></p> <p>4.1 Where is the data relating to this activity coming from (existing and future)?</p> <p>4.2 Where is (should) the data (raw, cleaned, de-identified and analysed) being stored for use?</p>	<p>Original question:</p> <p>Where should data (existing and future) be collected, stored, analysed, and reported, and by whom?</p>
<p>6. <b>How</b> will a WIL Evaluation Framework be repeatable and therefore sustainable?</p>	<p><b>Guiding Sub-questions</b></p> <p>6.1 remain a repeatable process (BAU)?</p> <p>6.2 become a longitudinal study?</p> <p>6.3 contribute to reports of institutional impact?</p>	<p>Original Question:</p> <p>How can we measure what matters to SEBE in terms of the innovation and impact of WIL across all courses?</p>

Our first study, the 'why' of WE (Young et al., 2023), reported on the importance of each institution finding their purpose for WE (Young et al., 2023). Our second study, the 'who' of WE (Young et al., 2024), reported on the dimensions we (the 'who') collectively valued about WIL and the challenges associated with macro-evaluations of WIL (those involving a large-scale network of invested stakeholders across varied and multiple approaches to WIL offered across a school/faulty/institution).

This case study, and the reason why this paper focuses in on three of the six guiding questions, is because they relate to the application of the WIL Evaluation Framework design (informed by all 6W's) and thus encapsulate the parts that needed to be, and could be, tested via data capture and analysis. Further to this, when representing and examining our 'what', this paper focuses on our decisions relating to the types of WIL included in our evaluation. The 'where' focuses on the agreed destination (data sources) for mining, and challenges and solutions of data wrangling when evaluating the impact and effectiveness of WIL types included in our bespoke WIL Evaluation Framework. The 'how' was concerned with process and lessons learnt during our trial of implementing the WIL Evaluation Framework. The data and analysis relating to the 'how' is out of scope for this paper, however, some of the initial findings are addressed by way of conclusion as they support our future iterations of evaluation and are likely to contribute to the nascent field of WIL evaluation more broadly.

In summary, the testing of our WIL Evaluation Framework needed to deliberately consider what to include and measure (predominantly item 3.2 and 3.4 in Table 1), to know where to find relevant data (see 4.1 in Table 1) and, whether employability-related outcomes could be triangulated (see 4.2: 'analysed' in Table 1). To report on this study ethics approval was sought from the authors institution ethics panel with approval code: SEBE-2023-01-MOD01.

## Evaluation Approach

The WIL evaluation project was incubated in direct response to the National Priorities Industry Linkage Fund (NPILF) for enabling job ready graduates. NPILF, as part of the job ready graduates package established in 2020, gave rise to the need for an approach to evaluating WIL activities in STEM. The STEM case study was designed to test a bespoke evaluation of WIL, with the goal of sharing the lessons across the sector, particularly to support a university-wide evaluation of WIL, needed for ongoing institutional reporting of 'good WIL'. The framework was developed to evaluate WIL courses and/or students enrolled in undergraduate STEM across the entire Faculty, as a way testing the frontier for others more broadly in the sector.

The WEP (and the intended outcome of the WEP – the design, testing and implementation of a WIL Evaluation Framework and related process) responds to regulatory imperatives for substantiating sound pedagogical practices, valuable learning experiences, beneficial outcomes for all stakeholders, and institutional processes for quality assurance and quality improvement of WIL (Tertiary Education Quality and Standards Agency (TEQSA), 2017; Department of Education, 2021). Jackson (2024) highlights the need for ongoing evaluation of WIL to ensure quality and inclusive practice. WIL, and the evaluation of WIL, is a strategic priority for the Faculty (Young et al., 2022), given that our goal (via the SEGP) was to actively enhance student employability to enable them to create work, move into careers and make an effective contribution to their world. Our strategy, and thus the WEP, aligned with the institutional employability strategy (Deakin University, 2022), and so our vision for a bespoke WIL Evaluation Framework (WEF) was a timely priority.

WIL is often positioned as the vehicle for developing graduate capabilities beyond discipline specific knowledge and skills (Green et al., 2009; Jorre de St Jorre et al., 2016). Universities are now obligated to develop graduate capabilities, often prepared via institutional-centric definitions of graduate employability (noting the paucity of universal sector-wide agreement of the definition of employability particularly in relation to WIL). In this paper, employability is defined as:

*Employability means that students and graduates can discern, acquire, adapt and continually enhance the skills, understandings and personal attributes that make them more likely to find and create meaningful paid and unpaid work that benefits themselves, the workforce, the community and the economy (Oliver, 2015).*

There is no universal method of enhancing employability, and any number of variables could affect one's achievement and enactment of it (Cranmer, 2006). When it came to considering students' employability in our Faculty, we first looked towards ways to determine the potential impact of WIL

interventions as the intentional and designated primary vehicle for enhancing student employability for our undergraduate students. Our approach to evaluation was cemented by a need to measure and report on the effectiveness and impact of individual WIL programs, as well as the entire suite of undergraduate WIL offerings across the Faculty.

The project team understood the ambitious nature of both the design and testing of this macro-evaluation. Given it was the Faculty's first encounter into the practice of a large-scale evaluation, the volume of the 'what' in scope (while risky), was driven by our desire to translate our varied practices of WIL into a shared understanding of the context, delivering a fit for purpose WIL evaluation, alongside the value of trialling a macro-evaluation. Faculty-specific stakeholders of WIL, along with data custodians and policy makers, wanted to know what was best to evaluate and then measure when considering the effectiveness and impact of diverse WIL types across an institution. The project team, in consultation with broader WIL stakeholders at Deakin University (Young et al., 2024), highlighting the intent to measure effectiveness and impact, refined the initial WIL evaluation guiding questions (see Table 1) to encapsulate the nuances of our context, purpose and agreed value (Rowe et al., 2018).

WIL stakeholders (Young et al., 2024) agreed on the areas that needed to be encompassed in our WEF process. Our WIL stakeholders included senior staff, teaching academics and professional staff involved in WIL from the Faculty, senior staff from the university executive, staff from the university careers service, industry partners and student alumni. Further detail about 'who' was consulted for WIL evaluation is described in (Young et al., 2024). The planned areas that were particularly relevant to the pilot of the WEF have been summarised below in Figure 1. The priorities were:

1. **Evaluation Orientation:**  
Define the evaluation model type (related to purpose) and acknowledge the typical outcomes (positive and negative) expected for the evaluation model type and the criteria of merit.
2. **Quality Standards:**  
Deciding on, and framing, the set of standards used to frame the evaluation purposes.
3. **WIL types:**  
Describing the programs (WIL-types), the dimensions of in-scope activities and devising a hierarchy/schema for classifying the WIL-types to facilitate data mining and analysis.
4. **Indicators and Measures:**  
Deciding on the criteria of merit, identifying and selecting the range of indicators and setting the measures of success for what 'good' looks like.
5. **Methodology:**  
Design the data (i.e. mixed) methods including and considering 'how to use' existing data tools and mechanisms (e.g. Graduate Outcome Survey (GOS), (GOS-L)), framed within the appropriate methodology (i.e. inclusive of process and outcome-oriented evaluation models with multi-dimensional set of indicators, array of indices and/or a single indexes).
6. **Data:**  
Considering where Institutional, Faculty-level, school-level, unit level data can be found and where it should be stored, how and by whom it will be collected, mined and analysed (Including the perceptions and experiences of all stakeholders).
7. **Outcomes:**  
Re-acknowledge the typical outcomes (positive and negative) and identify the diverse outcomes (positive and negative), including reporting on the reliable data typical of the evaluation model type.

**Figure 1: WEF Planning Priorities**

## What: Our approach to WIL evaluation

The ‘what’ of our evaluation initially focused on an overarching guiding question: ‘What are the key underpinnings, considerations, requirements, and deliverables for WIL evaluation?’ We had to learn how to tackle the ‘what’ of our evaluation, so we looked to models devised for evaluating education - Stufflebeam & Coryn’s (2014) ‘Context, Input, Process, and Product (CIPP)’ model was selected. The ‘what’ guiding question evolved to the sub-evaluation questions (see Table 1), to unpack and better grapple with how to understand what to evaluate. The most challenging area here was, what should our indicators for success be? To address success, the project team, in consultation with the WIL evaluation advisory group, developed the following criteria of merit (CoM) (see Table 2 below) to represent the dimensions deemed to be of value for our context (Young et al., 2023). The criteria of merit were identified through analysis of the program logic, evaluation priorities and values and the definitions of; Effectiveness, Impact and Learnings became the basis by which the success of the WIL program would be judged (Davidson, 2005). We proposed key evaluation questions (KEQs) that were aligned to the criteria of merit, to provide insights into the WEF, as shown in Table 2. The evaluation questions were also grounded in practical considerations, reflecting resources, data availability and systems.

**Table 2: Criteria of merit (CoM)**

Criteria of Merit	Definition
<b>Effectiveness</b>	The WIL program has been implemented as intended and adheres to priority WEF Operational Plan quality standards across the dimensions of capability, participation, student experience and stakeholder engagement and outcomes. Students, staff and partners value WIL experiences. Critical success factors and barriers are identified which can inform WIL future planning.  KEQ 1: Effectiveness: How well is WIL being implemented?
<b>Impact</b>	The WIL program results in positive changes for students including academic outcomes, staff and industry partners. Participants value what they have gained from the program.  KEQ 2: Impact: To what extent do of the WIL modes contribute to the students’ journey towards employability?
<b>Learnings</b>	Learnings and opportunities for continuous improvement, to increase the sustainability and repeatability of WIL activities and the WEF, are identified and provide an evidence base for future planning to ensure WIL, all WIL programs and the WEF are continually evolving within available resources, to maximise the quality of the WEF and impact of WIL for students, staff and other stakeholders.  KEQ3: Learning: What learnings and opportunities for continuous improvement can inform future growth of employability initiatives in the Faculty?

## Quality Standards

The criteria of merit and key-evaluation questions (Table 2) and the associated sub-questions (Table 1) ensured a logic for the collection of data and analysis of outcomes specific to our context-sensitive approach to investigating the impact of WIL. Enacting the WEF required a considered process of setting quality standards so that data could be drawn to a validated evaluation frame. We looked to other established models to tackle how we might begin a design for a context-specific set of quality

standards. There was no better place to start than the WIL Quality Assurance Framework (Campbell et al., 2021). The four key quality pillars; capability for WIL, participation in WIL, student experience and stakeholder engagement were instrumental foundations for our WEF. The validated elements provided a monumental support so that we could review (rather than design) and then adapt a final set of critical dimensions for our context. We landed on a matrix of quality standards 17 in total (see 'WIL Evaluation Framework KEQ1 and (adapted) quality standards' in Table 3 below). Our WIL evaluation quality standards specifically related to key evaluation question 1 on effectiveness.

**Table 3: WIL Evaluation Framework KEQ1 and (adapted) Quality Standards**

<b>KEQ 1: Effectiveness: How well is WIL being implemented?</b>	<b>WIL Evaluation Framework Quality Standards</b>
1.1 To what extent are staff engaged and have the capacity and resources to deliver WIL?	Academic and professional staff are: <ul style="list-style-type: none"> <li>a) involved in the support and or delivery of WIL</li> <li>b) supportive of WIL as a tool to enhance student employability and value it</li> <li>c) confident in their own ability to teach and support WIL</li> <li>d) provided with adequate professional development opportunities to deliver WIL</li> <li>e) provided with adequate resources to ensure they have the capacity to deliver WIL.</li> </ul>
1.2 To what extent are students engaged and participating in a range of scaffolded WIL opportunities?	Scaffolded WIL activities enable students: <ul style="list-style-type: none"> <li>a) to participate in WIL</li> <li>b) to be ready for each WIL experience</li> <li>c) to successfully complete each WIL unit</li> </ul>
1.3 To what extent do students value WIL experiences?	Students identify value in their WIL activities in areas such as: <ul style="list-style-type: none"> <li>a) relevance to their knowledge and skill professional interest</li> <li>b) securing future employment</li> <li>c) career planning and employment requirements</li> <li>d) students identify WIL opportunities as positive, motivating and beneficial means for enhancing student employability.</li> <li>e) students are satisfied with their overall placement experience.</li> </ul>
1.4 To what extent do partners value WIL experiences?	Partners are: <ul style="list-style-type: none"> <li>a) involved in WIL</li> <li>b) supportive of WIL as a tool to enhance student employability</li> <li>c) adequately prepared and supported to be involved in WIL</li> <li>d) positive about their WIL experience and identify benefits.</li> </ul>

While the CoM enabled a clear path for setting our key evaluation questions, a slightly different logic for the collection of data and analysis of outcomes was applied for each key evaluation question. For the pilot, a pragmatic scoping decision was made – only the effectiveness of key evaluation question 1 would be tested/applied during the pilot. Quality standards that relate to key evaluation questions 2 and 3 were proposed but not operationalized in the WEF due to limits on capturing longitudinal data.

### **What: The WIL activity types to be included**

Our decision for a macro-evaluation approach, determined to measure the effectiveness of both individual WIL programs, as well as the entire suite of offerings across the Faculty required the first what sub-question (see 3.2 in Table 1 above) to consider: What are the WIL activity types included in the WIL Evaluation Framework? For our Faculty, our stakeholders had already asserted that WIL evaluation needed to encapsulate more than ‘placement-based’ WIL (P-WIL) (Young et al., 2023, 2024). Due to the wealth of WIL-types already occurring, it would have been remiss, and a disservice to our stakeholders, not to be inclusive of the breadth of innovation and offerings that makes our WIL history for our Faculty unique (Young et al., 2022). And Winchester-Seeto (2024) describe the importance of making active choices about the most effective model of WIL for a particular set of students and circumstances, to ensure programs and offering can best suit student need.

Solving the problem of how to best measure the effectiveness of these varied WIL types was unlikely to be straight forward. But given our strategy was to ensure undergraduate students had multiple opportunities to engage in WIL, macro-evaluation was necessary. Further to this, as WIL was implemented across the schools to accommodate the strategy, a mapping of where the various and numerous WIL-types were occurring across curriculum was vital to the framing of our WEF. The TEQSA definition of what WIL encompasses was useful as a starting point for classifying our Faculty’s WIL offerings:

*Work-Integrated learning (WIL) encompasses any arrangement where students undertake learning in a work context as part of their course requirements. WIL can be undertaken as part of coursework or research training. WIL activities may include:*

- *professional workplace placements (also known as internships, clinical placements, fieldwork, practicums) whether local, interstate or international*
- *online or virtual WIL (e.g. telehealth) with real clients or industry input*
- *industry-partnered projects in the classroom (e.g. hackathons, incubators/start-ups) that involve industry, community or professional partners*
- *a simulated work environment with industry input, consultation or assessment, or*
- *activities in other contexts involving industry or community partners (TEQSA, 2022 ‘What does work-integrated learning encompass?’).*

The TEQSA definition were adapted to suit a typology of WIL types that we referred to in the pilot as WIL activity. Table 4 summarises our WIL activity types within our curriculum.

**Table 4. Types of WIL Activity Within the Curriculum**

<b>Mode 1: Placement Units (P-WIL)</b>	Placement (internships or IBL) experiences involve an ethos of ‘working to learn’, undertaken with host from industry, a ‘workplace’ where there is a host and student agreement and the student is supervised by the host organisation.
<b>Mode 2: Authentic Assessment (AA)</b>	Authentic Assessment focuses on the development, assessment, and communication of transferrable employability skills contextualised within a realistic, real-world environment, yet delivered within a university setting. The inclusion of industry/community in the co-design, co-delivery, and co-assessment of students is considered the ideal.
<b>Mode 3: Career Education (CE)</b>	Career Education is learning designed to develop students’ knowledge of future careers, their evidence of employability skills and their ability to transition to graduate employment. Career education in this mode of activity is embedded into the curriculum, often combined with discipline specific subjects/units.
<b>Mode 4: Professional Practice</b>	Professional practice occurs via project (team) work undertaken with and for industry or community, that involves problem-based learning. Professional practice is facilitated within the curriculum, often scaffolded as a later stage subject that focuses on a project of larger scope.
<b>Mode 5: Simulated WIL</b>	Workplace simulations involving experiences that imitate professional practice and/or ‘workplace’ environments. Simulated WIL can be an individual or team-based activity, often conducted within the curriculum.
<b>Mode 6: AA &amp; CE Assessment</b>	Some subjects in a student’s course include both authentic assessment and career education and constitute a different form of WIL activity.

Our version of the ‘what to evaluate’ was important as it involved the inclusion of the sum parts of single WIL programs, discipline-specific approaches, and the total of course-wide offerings for Faculty-wide reporting. The modes also supported the requirement of finding a means to compare the impact of a single WIL unit as well as multiple WIL initiatives (a WIL program) across an entire undergraduate course. The ‘modes’ of activity offered both discrete and comparative processes for evaluation, enabled us to position a data framework or schema for evaluation. Without this, our broader learnings from evaluation would not be clear. It also meant that our diverse offering of WIL activities would set up to include multiple measures for each WIL-type being evaluated. Lessons from Rowe et al., (2018, p. 280) revealed that variable considerations needed to encapsulate an understanding of context such as the ‘nature of the program, the program components being evaluated, and how success, impact and quality is defined and measured, before determining the methods.’ As pointed out by Winchester-Seeto, (2019, p. 20), when it comes to evaluation, especially across a university or program, it becomes necessary to capture data on, measure, and then celebrate, the quality of each WIL program type. As such, the purpose of each WIL activity type was clearly defined (see Table 4) and then applied to the specific units that would be in scope for evaluation (noting that the names of the units and unit codes have been omitted from this paper).



### What: The data types/tools used to evaluate the activities

To address the 'what' evaluation guiding question, and particularly, the sub-question 3.3: What are the data types/tools used to evaluate the activities we collected data from a variety of institutional systems, drawn from three teaching periods (trimesters) in 2022 to support the WEF, as listed in Table 5. We connected our data through the teaching unit or discipline/ school where it was collected. Our WIL activity types (Table 4) enabled us to analyse the data in relevant groups. We did not collect external data, such as the graduate outcomes survey, as a part of our data set. The graduate outcome survey was deemed out of scope as we could not connect to our WIL activity types.

**Table 5. Data Sources for the WIL Evaluation Framework**

Data Sources
Student Management System, unit enrolment
Unit level completion data
Student Survey Evaluation of Teaching and Learning
Student survey at completion of placement
Supervisor/ host survey at completion of placement
Placement Application Management System, placement location and host engagement
Research Survey with staff and students on Career Education
Research Survey with staff and students on Authentic Assessment
Student survey pre and post unit completion
Student Survey at completion of Individual Unit
Survey of Academic and Professional Staff Involved in WIL
Student Survey Evaluation of Teaching and Learning
Student Survey at completion of Individual Unit

There are numerous forms of data currently available across the sector for evaluating WIL, most of which is informative at the institution level, but less so at the degree/program-level. For example, most universities use Quality Indicators for Learning and Teaching (QILT) - a suite of national, higher education surveys endorsed by the Australian government that covers student life cycle from commencement to employment for degree-level quality assurance, using proxy measures such as student completion, satisfaction or self-reported information on achievement of generic skills (as collected from the Graduate Outcomes Survey) (Jorre de St Jorre et al., 2016). More recently the QILT data has incorporated questions on WIL, querying the student experiences across a variety of WIL types. Jackson & Dean (2023) analysed this WIL data, highlighting how the different types of WIL impact on student skill development and preparedness for employment. Aside from the sector-wide data mechanisms available, Rowe et al., (2018, p. 280) suggest that collecting data from multiple stakeholders across the whole institution provides a unique opportunity to measure outcomes on a large scale, as well as the ability to drill down to assess which program components are the most effective and for whom. We can gain some insight into the effectiveness of the courses and programs through student satisfaction surveys and graduate employment and destination instruments, but these are not granular enough, or timely enough, to understand and measure the quality of WIL programs and courses, especially for improvement purposes. Nonetheless, we can use the existing systems to think about the best ways to approach this task (Winchester-Seeto, 2019).

## Application of the WIL Evaluation Framework

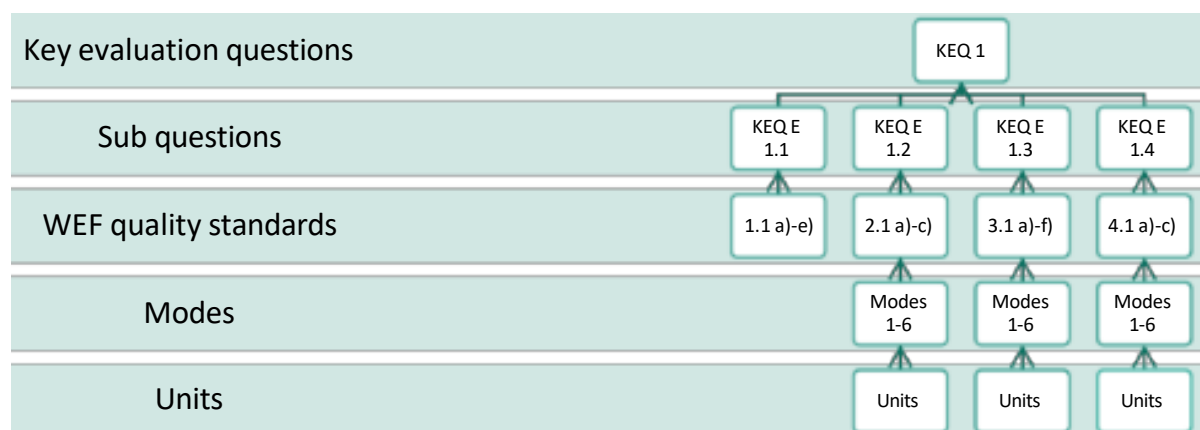
The description of the application of our WEF also directly addresses our final ‘where’ guiding questions via the sub questions: 4.1 Where is the data relating to this activity coming from (existing and future)? and 4.2 Where is (should) the data (raw, cleaned, de-identified and analysed) being stored for use? To describe our WEF, we provide a high-level description of the method of analysis, the logic applied to understanding and aggregating the data, and how this has been synthesised to generative evaluative findings for WIL.

### Where: Data Storage

As the WIL evaluation occurred on pre-existing student performance and experience data (see Table 5), data storage was consistent with our institutions information management standards and procedures. All evaluation deliverables were managed and stored in accordance with Deakin Universities knowledge management systems and ICT standards.

### Where: Data analysis

The data sources were collated into a master dataset in Microsoft Excel. We listed data at the unit (subject) level to enable aggregation towards modes (or activity type). Data in our analysis was restricted to quantitative (numerical) data in this application. This was due to time and funding limits available for evaluation. Then we mapped the individual data items from each data source to the WEF quality standards, evaluation sub questions and key evaluation questions (Table 2). It should be noted that the mapping of data at unit and activity/ mode level reflects the specific pedagogical structures of Deakin University. Other contexts would need to map data according to their own pedagogical structures. The definition of Quality Standards and key evaluation questions is more universal and may be usefully translated directly into other contexts verbatim. Figure 2 shows the data mapping process that supported the application of the WEF.



**Figure 2** Aggregation of Data

### Data synthesis

In addition to analysing different sources of WIL data, how these would be compared, combined and aggregated to answer the evaluation questions and sub questions needed to be defined. An explicit evaluative synthesis method has been used to aggregate evaluation ratings from different data sources and generate logical evaluative conclusions for each criteria of merit, key evaluation questions and quality standard. The quality standards effectively defined what ‘good’ WIL looks like (Davidson, 2005) in the context of this program (for the modes and the units).

When completing the data analysis a rubric was employed to define performance of the quality standards. The rubric defined the parameters for grading performance against expectations. We set threshold standards (Table 6) to determine the extent that the evaluation question was achieved. Standards were generally determined to meet or exceed expectations when they achieved a threshold of 70%. But thresholds can be adjusted based on institutional standards. These standards were internally validated through consultation with the WIL advisory group. A resultant standard received a rating using the following traffic light coding. The traffic light coding was used because it is universally understood and could be implemented in monitoring and evaluation reports in MS Word or Excel to good effect.

**Table 6: Rubric Traffic Light Coding**

	Met or exceeded expectations
	Partially met expectations
	Not yet demonstrated
	Data not yet available
	Data available, no judgement can be made due to lack of base line data

Table 7 below provides an example of the rubrics used to address the criteria of merit for key evaluation question 1 on effectiveness. The synthesis rubrics were shared and tested with the WIL Evaluation Advisory Group before being finalised. The key evaluation sub-questions were considered to be of equal value/weight and were aggregated to provide an overall judgement about the related key evaluation questions (Figure 2). An overarching program evaluation rubric was also used to capture the performance of each of the criteria and merit and to show broadly how the program performed.

**Table 7 KEQ 1.1 Rubric Example: To what extent are staff engaged and have the capacity and resources to deliver WIL?**

High	Medium	Low
% Staff are involved in the delivery of WIL.	% Staff are involved in the delivery of WIL.	% Staff are involved in the delivery of WIL.
% Staff are supportive of WIL as a tool to enhance student employability.	% Staff are supportive of WIL as a tool to enhance student employability.	% Staff are supportive of WIL as a tool to enhance student employability.
% Academic staff are confident in their own ability to teach WIL.	% Academic staff are confident in their own ability to teach WIL.	% Academic staff are confident in their own ability to teach WIL.
% Academic and professional staff are provided with professional development opportunities.	% Academic and professional staff are provided with professional development opportunities.	% Academic and professional staff are provided with professional development opportunities
% Adequate resources to ensure they have the capability and capacity to deliver WIL.	% Adequate resources to ensure they have the capability and capacity to deliver WIL.	% Adequate resources to ensure they have the capability and capacity to deliver WIL.

## Where: Data Limitations (availability)

As we worked through evaluation, we found several limits and challenges when gathering and working with the data. The recorded challenges at the conclusion of the pilot included:

- No staff data was available for one teaching period, which means no judgement can be made about WEF Quality Standard 1.1
- Some student data sources have not been made available by survey owners due to response rates of less than ten.
- Survey, and individual survey questions, response rates may not meet the liberal conditions for required response rates described by (Nulty, 2008). Low response rates limit the reliability and representativeness of the data cited in this report. Whether or not the data source is considered reliable is noted in each figure caption.
- Where survey data was available and response rates were less than five, then these results were suppressed due to the potential to identify respondents, even though the surveys are confidential, and results are anonymous. The suppression rule is applied unless the total number of students enrolled in a unit was less than ten, in which case the data was suppressed or excluded at the discretion of the analysis team. It should be noted that the suppression rules used in this evaluation are extremely liberal compared to the Graduate Outcome Survey (GOS) suppression rule of 25 responses or less. Given the size of the survey field in GOS compared to SEBE data sources, this difference is justifiable.
- To address data limitations, a variety of data sources were collected (as described Table 5). However, this results in much variation in data items due to inconsistency in data collection approaches across the university.
- A lack of qualitative survey data and inability to gather further insights through interviews or focus groups limits the depth and validity of the findings in this report.

The most difficult challenges relating to the ‘where’ of our WIL evaluation fall into the resource bucket. The availability of evaluation experts, data integrity, availability and sourcing of custodians of data (Young et al., 2024), during nascent enterprise-wide platform solutions, can prevent expedient data outcomes. We identified the need for a designated and expert task force/team as an enterprise-wide operation/investment (beyond WIL practitioners/champions who opt to invest time in scholarly approaches to WIL) to ensure that the data (capture, analysis, reporting) is repeatable.

Despite the data limitations our WEF set out to encapsulate an iterative process of evaluation, rather than a recipe for collecting WIL data. The action of testing the WEF organically brought to the surface many lessons. Our key evaluation question 3 asked: ‘What learnings and opportunities for continuous improvement can inform future growth of employability initiatives in our Faculty?’ Our vision for a bespoke WEF when used as a synthesis rubric approach to ascertain what ‘good’ WIL looked like (Davidson, 2005) found that WIL evaluation should be:

- **About Learning** – driving improvements of WIL programs for all stakeholders.
- **Repeatable and Pragmatic** – ensuring there is an institutional plan to produce evidence-based benchmarked findings within the resources available.
- **Beneficial** to all involved not just in WIL programs, but in the process of evaluating the WIL activities.

Therefore, if WIL evaluation framed as an opportunity for learning (by inference, transformation), but presented as something familiar (by inference, a repeatable process), then a good practice of quality assurance and quality improvement of WIL is likely to be of great benefit to all stakeholders.

## Our How

Whilst the testing of the WEF has highlighted many challenges with data, there are indeed important learnings related to the paucity of evaluation processes and models. This final section reframes some of these the barriers as insights, so that likely similar concerns relating to data for others can potentially be overcome if anticipated (based on our trial) and therefore planned from the outset of a large-scale evaluation of this sort. If repeatable and sustainable mechanisms for measuring what matters to the university can be agreed up, then there are great opportunities to generate live/current knowledge and data (via analytic dashboards) about impact and effectiveness. As mentioned in the 'where' barriers section above, a realistic allocation of resources required for WE endeavours is needed.

Despite the challenges, the process of designing the WEF and conducting a test of the macro-evaluation, particularly in relation to the criteria of merit on effectiveness, as described above, was an invaluable and transformative undertaking that is informing our ongoing processes today. To this end, the pilot continues to frame our 'How' guiding sub-questions which resonate our philosophical stance of wanting the WEF process to be about learning from the ambitious initiative and sharing with others the nascent field of WIL evaluation. As introduced, our project team reached a consensus - we needed an approach to WIL Evaluation that would be repeatable, sustainable and scalable, that could be used and refined in future cycles of WIL evaluation and that could capture, pre-determined and unexpected outcomes, both positive and negative, that may arise (Ellington, 1993). If we were to hold 'learning' as a true north principle (Young et al., 2023, p. 17), then the WEF design needed to be robust and flexible to enable it to evolve over time.

Our first ever attempt to design a large-scale evaluation of this sort, tackled and not yet tamed, revealed the demonstrated complexities (and limitations) of evaluating the impact of WIL in terms of 'value' for the different stakeholders. For example, many of our lessons revealed how the different WIL types have different levels and types of impact. As such, we recommend for others to carefully examine (as we did) if and how one framework could encapsulate disparate types of impact for an overarching evaluation of WIL. The pilot provided evidence of the necessity of using the 6W guiding questions to determine the intended long-term outcomes expected from the evaluation process (Young et al., 2023). The process of data collection during the pilot highlighted the need for continuous evaluation cycles to generate long-term impact (see sub-question 6.2). In addition, limitations in the data sources (Table 6) and data types (all quantitative) highlight a need for more time to evaluate qualitative outcomes from a variety of stakeholders to make the WEF process business as usual (BAU – see sub-question 6.1). Further, the application of the WEF in our Faculty highlighted a need for an increase in staff engagement and understanding of WIL outputs and outcomes. Limited data from staff on WIL process and outcomes evidenced an opportunity for increased staff engagement and continuous data collection on not just the student but staff experience of WIL, adding to the potential for institutional impact (see sub-question 6.3).

Another major factor in our context was that the pedagogical practices and associated assessments for each type were not always identical. Other specificities of each WIL-type that influence both participation and engagement data, such as enrolment factors in compulsory versus elective unit offerings suggest some degree of pre-conceived value for the student. The trickle on effect is that this may affect participation rates and engagement experience for the network of stakeholders – academic, hosts and other students for example. Other worthy considerations is where the WIL units are wholly WIL focused versus the WIL experiences that made up a smaller component of the unit, and whether the WIL unit was standalone or scaffolded. Understanding the details of WIL activity, particularly when constructively aligned to other WIL activities, was a serious consideration for us given our strategy to ensure students had multiple opportunities to engage in WIL within their course. Understanding this impact has yet to be resolved but is anticipated when the evaluation findings can be shared and unpacked with stakeholders.

So, in summary, we can note that our decision to formulate the WEF in the way we did was, and still is, transformative. This was because, as Kerr et al. (2016) suggested, we spent considerable time identifying the purpose of evaluation, what we valued knowing more about, who the audience might be and how we might use this information. Not only did we need to understand the value, effectiveness and impact of our WIL offering, but we also had a strong desire to learn from the evaluation findings and use these learnings in future decision making and WIL design. We worked to maximise the process and outputs (findings) of our WIL evaluation to be relevant, accessible and meaningful to our wide range of stakeholders (Young et al., 2024). Prioritising and then formalising 'learning' in our WEF underscores our desire for WE to become a repeatable process in our normal business operations, with goals for both longitudinal studies and reporting on institutional impact. We welcomed and laid the foundations for the many lessons likely to be learned about future of employability initiatives in our Faculty, as well as the continuous improvement of a WEF to both enable and support the iterative enhancements of WIL innovations.

## Limitations

The WEF, which included a data schema to enable both the data to inform our context specific learnings about the WIL offering, and data about the process of designing and testing an Evaluation Framework, resulted in significant learnings about the risks and constraints associated with sustainable and scalable evaluations of WIL. Our pilot highlighted some of the limitations of comprehensive evaluations of this type. For any given evaluation to be effective at capturing data for robust analysis leading to informed outcomes, an obvious place to commence are considerations relating to how and where the representational data is best collected. This study affirms that for data analyses and findings to provide adequate benchmarking, the evaluation process must account for difficult to come by, but nevertheless, consistent data. Our study found that the collection of consistent data was difficult to come by; this finding was obscured at the onset of the research design. The degrees to which the required data were buried in the operational were not anticipated. The cautionary note for others undertaking similar trials for WIL evaluation, is that possible methodological limits (relating to the quality, quantity, or diversity of the data), empirical limits (relating to the representativeness, validity, or reliability of the data), as well as likely analytical limits (relating to the completeness of the hypotheses and findings), are best made explicit during the scoping stage of an evaluation.

## Conclusion

The findings from the WEF can be used to inform ongoing institution-specific WIL evaluation projects, and sector-wide critical conversations relating to success measures and impacts of WIL. The findings highlight the challenges associated with mapping, accessing, and analysing data that relate to institutional impact as well as seeing the challenges of supporting a sustainable, repeatable/ ongoing process of WIL evaluation. As the WIL programs and modes of activity (Table 4) included in our Faculty-wide WEF typify WIL across the sector, this study calls for careful consideration of both the enablers and the barriers intrinsic to large-scale data approaches to WIL evaluation. For example, considerations of the scale of data types for inclusion (Table 6), as well as the CoM and KEQ decisions (Table 2 and 3), need to be determined before measures of 'good' WIL can be set, let alone, reported on, across the sector. Quality outcomes from a WIL program typically hinge on frequent evaluation against a carefully selected framework/model based on a given program(s)' desired outcomes. Such an evaluation gives the educational institution a clearer understanding of the needs of all participants. It also measures the degree to which a program is achieving its goals and helps identify specific shortcomings and solutions.

Whilst not fully resolved, future contemplation of data, including the processes/approaches to WE, as well as enterprise-wide/sector-wide systems for WIL evaluation, to supplement current sector-wide

reporting of graduate outcomes are needed (Young, 2023; Young et al, 2024). The potential is that via collaboration and even similar action-research approaches across the sector, crucial and timely agreements relating to the key elements of what a universal WE framework may include, can be addressed. Our approach and decisions relating to what, where and how to measure what mattered in our context, are transferable factors for those with an appetite to bring an evidence-based approach to the impact that diverse WIL types have on student employability. The hope is for the insights from this paper to guide others embarking on a similar large-scale, context-specific, evaluation of varying and diverse WIL types.

## Acknowledgements

The authors disclose that they have no actual or perceived conflicts of interest. Whilst the authors did not receive any funding for this manuscript, beyond resourcing for academic research allocations, funding was provided via an internal teaching and learning grant as part of the large-scale WIL evaluation project. The authors wish to acknowledge the Associate Dean Teaching and Learning in Faculty of Science, Engineering and Built Environment, and Senior Deputy Vice Chancellor, Academic, Deakin University for sponsoring the WIL evaluation project and supporting of the related research endeavours

## References

- Campbell, M., Russell, L., Thomson, K., Tunny, R., Smith, L., & McAllister, L. (2021). The Construction and Testing of a Framework to Assure the Institutional Quality of Work-Integrated Learning. *International Journal of Work-Integrated Learning*, 22(4), 505–519. <https://files.eric.ed.gov/fulltext/EJ1338160.pdf>
- Cranmer, S. (2006). Enhancing graduate employability: best intentions and mixed outcomes. *Studies in Higher Education*, 31(2), 169–184. <https://doi.org/10.1080/03075070600572041>
- Davidson, E. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Sage.
- Deakin University. (2022). *Deakin 2030 Ideas to Impact*. [https://www.deakin.edu.au/\\_data/assets/pdf\\_file/0012/2758719/Strategic-Plan\\_2030\\_Refresh\\_March\\_2024.pdf](https://www.deakin.edu.au/_data/assets/pdf_file/0012/2758719/Strategic-Plan_2030_Refresh_March_2024.pdf)
- Department of Education (2021, April 24). *Higher Education Standards Framework*. Australian Government. <https://www.education.gov.au/higher-education-standards-panel-hesp/higher-education-standards-framework>
- Ellington, H. (1993). *Handbook of educational technology*. Nichols Publishing.
- Green, W., Hammer, S., & Star, C. (2009). Facing up to the challenge: why is it so hard to develop graduate attributes? *Higher Education Research & Development*, 28(1), 17–29. <https://doi.org/10.1080/07294360802444339>
- Jackson, D. (2024). Work-integrated learning: opportunities and challenges in Australia. *Higher Education Research & Development*, 43(3), 767–773. <https://doi.org/10.1080/07294360.2024.2307929>
- Jackson, D., & Dean, B. A. (2023). The contribution of different types of work-integrated learning to graduate employability. *Higher Education Research & Development*, 42(1), 93–110. <https://doi.org/10.1080/07294360.2022.2048638>
- Jorre de St Jorre, T., Johnson, L., & Oliver, B. (2016). *Deakin Hallmarks: principles for employability credentials*. <https://hdl.handle.net/10536/DRO/DU:30097295>
- Kerr, G., Stirling, A., Banwell, J., MacPherson, E., & Heron, A. (2016). *A Practical Guide for Work-integrated Learning: Effective Practices to Enhance the Educational Quality of Structured Work Experiences Offered through Colleges and Universities*. [https://heqco.ca/wp-content/uploads/2020/03/HEQCO\\_WIL\\_Guide\\_ENG\\_ACC.pdf](https://heqco.ca/wp-content/uploads/2020/03/HEQCO_WIL_Guide_ENG_ACC.pdf)
- Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education*, 33(3), 301–314. <https://doi.org/10.1080/02602930701293231>
- Oliver, B. (2015). Redefining graduate employability and work-integrated learning: Proposals for effective higher education in disrupted economies. *Journal of Teaching and Learning for Graduate Employability*, 6(1), 56–65. <https://doi.org/10.21153/jtlge2015vol6no1art573>

- Rowe, A. D., Nay, C., Lloyd, K., Myton, N., & Kraushaar, N. (2018). Telling Your Story of Work-Integrated Learning: A Holistic Approach to Program Evaluation. *International Journal of Work-Integrated Learning*, 19(3), 273–285. <https://files.eric.ed.gov/fulltext/EJ1196752.pdf>
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications* (Vol. 50). John Wiley & Sons.
- Tertiary Education Quality and Standards Agency. (2022, May 4). *Guidance note: Work-integrated learning* (Version 2.0). Australian Government. <https://www.teqsa.gov.au/guides-resources/resources/guidance-notes/guidance-note-work-integrated-learning>
- Tertiary Education Quality and Standards Agency (2017, October 11). *Guidance note: Academic quality assurance* (Version 2.2). Australian Government. <https://www.teqsa.gov.au/guides-resources/resources/guidance-notes/guidance-note-academic-quality-assurance#toc-heading-2>
- Winchester-Seeto, T. (2019, March). Quality and standards for work integrated learning. *Australian Council of Deans of Science*. ACDS Teaching and Learning Centre. <https://www.acds.edu.au/wp-content/uploads/Winchester-Seeto-Literature-Review-Quality-and-Standards.pdf>
- Winchester-Seeto, T. (2024). Intensive Work-Integrated Learning (WIL): The benefits and challenges of condensed and compressed WIL experiences. *Journal of University Teaching and Learning Practice*, 21(2). <https://doi.org/10.53761/1.21.2.06>
- Young, K. (2023, 19-20 October). WIL Evaluation: It is both what you know and who 'knows', that matters [Paper presentation]. *National Association of Field Experience Administrators (NAFEA) Conference*, Cairns, Queensland, Australia. <https://nafea.org.au/wp-content/uploads/2023/10/2023-NAFEA-Conference-Abstracts.pdf>
- Young, K., Campbell, M., Coldwell-Neilson, J., & Harvey, M. (2022). *Strong-WILed STEM: from placement-WIL to Faculty-wide multiplicity-WIL*. Australian Collaborative Education Network Case Studies. <https://acen.edu.au/resources/strong-wiled-stem-from-placement-wil-to-faculty-wide-multiplicity-wil/>
- Young, K., McKenzie, S., & Harvey, M. (2024). WIL Evaluation: It is both what you know, and who 'knows' what, that matters. *WIL in Practice* 2(1). <http://wilinpractice.nafea.org.au/index.php/WILIP/article/view/23>
- Young, K., Semple, A.-L., Harvey, M., & McKenzie, S. (2023). Theory and practice of why to evaluate WIL: A context-sensitive approach. *Advancing Scholarship and Research in Higher Education*, 4(1). <https://doi.org/10.59197/asrhe.v4i1.7841>