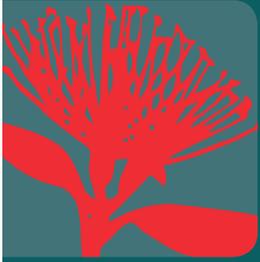


Central Regional Hub-funded project



Multi-stakeholder Feedback on Learning and Teaching in Professional Programmes and Work-Integrated Learning Contexts

Barbara Kennedy, Michael Philipp, Sara Herkt,
and Shanika Yoshini Koreshi

Research undertaken at Massey University

The evaluation tool outlined in this booklet was developed on the basis of teaching evaluation literature and was refined using feedback from iterative surveys of enrolled students, recent graduates, field supervisors, and work supervisors involved in Massey University's Postgraduate Diploma in Psychological Practice.

The evaluation tool was designed as a standardised tool that can be adapted for evaluations of other professional training programmes and work-integrated learning contexts. The intention is that such programmes obtain quality feedback from many programme stakeholders. This ensures that best practice teaching and pastoral care is provided to students and that quality assurance and accreditation standards are met.

This resource booklet is the outcome of the project 'Innovation in evaluating professional training programmes in psychology: A multi-stakeholder approach'.

The project was co-funded by Ako Aotearoa's Central Hub, through the Regional Hub Project Fund, and Massey University. The funding model has since been reviewed and changed.

Go to the Ako Aotearoa website project page for the full report at: <https://ako.ac.nz/knowledge-centre/multi-stakeholder-learning-and-teaching-evaluation/>

Publishers

Ako Aotearoa – The National Centre
for Tertiary Teaching Excellence
PO Box 756
Wellington 6140
www.ako.ac.nz
0800 MYAKONZ
info@ako.ac.nz

December 2022

ISBN: 978-1-98-856244-5 online



Creative Commons 4.0 New Zealand Attribution Non-commercial Share Alike Licence. Under this licence you are free to copy, distribute, display and perform the work as well as to remix, tweak, and build upon this work non-commercially, as long as you credit the author/s and license your new creations under the identical terms.

Table of Contents

- Introduction 2
 - Background..... 2
 - What is this instrument?..... 2
 - Who is this resource for?..... 3
- A multi-stakeholder evaluation tool is needed 4
 - What is in the literature?..... 4
 - What were the goals in developing the tool? 5
- Scope of this resource booklet 6
 - Sustainable benefits to learners..... 6
- Designing the multi-stakeholder evaluation tool..... 8
 - What sort of programme was this tool designed for? 8
 - Where does the teaching & learning occur in this internship programme? 9
- Measuring facets of the programme..... 10
- Psychometric properties of the tool 12
- Adapting the teaching evaluation instrument 14
 - Adaptation for professional Psychology programmes..... 14
 - Adaptation for other professional programmes 16
- References 19
- Appendix A: Glossary of useful terms..... 21
- Appendix B: Worksheet for adapting the instrument 24

Introduction

Background

Learning and teaching are interdependent. To teach well, it is essential to understand our students' experiences of our teaching and how it helps – or hinders – learning. That understanding informs the content we select, the design of learning experiences, our choice of delivery strategies, and our own professional development.

The standard approach to monitor teaching effectiveness in tertiary institutions is the teaching evaluation survey that focuses on student perspectives. However, courses that partner with work settings to achieve learning outcomes do not solely focus on the teacher–student nexus. Rather, such courses involve complex partnerships and multiple stakeholders, which may also include workplace supervisors, placement staff, and employers.

Training courses for trades, professions, and other work-integrated learning settings require extended workplace learning alongside formal tertiary study. However, typical instruments used to gather feedback about courses are designed for more traditional classroom-based teaching. Such evaluations will often not meet the needs of training programmes where tertiary courses and workplaces must work in partnership – such as professional programmes, apprenticeships, and internships. Equally, although training staff in tertiary courses may have expertise in their field of practice, their preparation may not have included advanced formal study in curriculum development, programme evaluation, or psychometric instrument development. To meet the above needs, we have developed a multi-stakeholder evaluation instrument.

What is this instrument?

The Multi-stakeholder Feedback on Learning and Teaching (M-FoLT) tool serves as an evaluation tool that gathers feedback from students and multiple stakeholders who are involved in work-integrated learning (WIL). The tool was originally developed to evaluate the impact of teaching practice on post-graduate learners undertaking professional psychology internships. The tool can be adapted to meet the programme needs of other professional and vocational internship programmes that are engaged in WIL, where multiple stakeholders contribute to a tertiary-level learning endeavour.

Who is this resource for?

This booklet is for subject-matter experts, teachers in professional training, teachers in trades, and those who oversee courses with substantive workplace components. It may be useful to teachers who want to gather feedback from their students, those who oversee their students in the workplace, and even workplace managers who make such placements possible.

To successfully adapt such an instrument for other programmes—and retain acceptable reliability and construct validity – requires both expertise in psychometrics and detailed knowledge of the programme in which teaching feedback is to be sought. This could be achieved in partnership between a psychometrician and a subject expert. But our hope is this booklet provides the rationale for item and scale construction so that professionals from other programmes can more easily identify adaptations that may be required to suit their needs.

Also available is a separate booklet that details the specific instrument we have used with our professional programme and provides technical details of the instrument's psychometric properties (see the supplementary booklet 'Technical Reports: Multi-Stakeholder Feedback on Learning and Teaching in Professional Programmes,' referred to as the *Technical Reports booklet*).

A multi-stakeholder evaluation tool is needed

What is in the literature?

To improve the quality of the work-integrated learning environment, work coordinators should use evaluation to better meet stakeholders' needs and education quality (Stirling, Kerr, Banwell, MacPherson, & Heron, 2016). Evaluations lead to processes that improve the way we are thinking and, consequently, inform how we develop, implement, and change programmes (Fitzpatrick, Sanders & Worthen, 2011).

Existing teaching evaluation instruments are not geared towards the specific needs of professional training programmes (Broomfield & Bligh, 1998; Freestone, Williams, Thompson, & Trembath, 2007). This applies particularly to the work-integrated learning component of professional programmes (Carless, Robertson, Willy, Hart, & Chea, 2012; Gross, 2006; Smith, 2012). Evaluation efforts for professional programmes have tended to be purpose-built for specific programmes, limited in scope, and not easily transferable to other programmes (e.g., Fortune & McKinstry, 2012; Nedeljkovic, Chaffey, Murray, & Brennan, 2014; Yap, 2012). Even among research that considers multiple stakeholders (e.g., Khuong, 2015; Welch, Vo-Tran, Pittayachawan, & Reynolds, 2012; Zahra & Pavia, 2012), no existing instruments examine multiple stakeholders' perspectives. Furthermore, we found no such tools specific to our own needs for broadly evaluating professional training programmes in psychology.

Each type of stakeholder group has different needs and priorities (e.g., potential for continual improvement, funding, to refine tools of thought, endorsements, and growth). For academics there are always tensions in deciding what to include in the curriculum, and there is a strong drive for research-informed choices when overseeing a programme or teaching a course. Students may prefer certain content or experiences and yet not always be in the position to know now what will best serve their development—but they can say how the teaching-learning experiences are working (or not) for them. Field supervisors are in an advantageous position to comment on how well the learning-teaching structure is helping students navigate the day-to-day realities of working with clients in a particular employment context. Employers may, at a theoretical level, value research-informed teaching. Yet they have service-delivery parameters, contract requirements, and client satisfaction to navigate, not all of which are necessarily aligned.

Seeking the perspectives of all stakeholders—current students, field supervisors, employer supervisors, and graduates—allows for a more complete picture of a programme's effectiveness. Programme improvements can therefore be based on a much more informed understanding of the nature of any gaps, and best address the needs of all stakeholders.

What were the goals in developing the tool?

We set out to design a standardised evaluation tool for learning and teaching that will:

- capture multiple stakeholders' perspectives;
- gather high quality data capable of driving ongoing programme development and curriculum improvement;
- have sound reliability and validity;
- provide the basis for the development of a standardised evaluation tool for other professional and vocational internship programmes and work-integrated learning contexts.

We acknowledge that developing a psychometric instrument is an extensive process that requires clarity about what you wish to measure. It also involves many considerations including: distinguishing outputs from outcomes, deciding what can be directly measured and what may be inferred from more accessible proxies, choosing appropriate response formats, and making trade-offs between what is desirable and what is practicable. Having worked through this process over four years, we feel this instrument is now sufficiently robust to adapt it to other groups to inform further development.

Scope of this resource booklet

This guide is not a formal psychometric instrument manual. Rather, we describe in plain language how the tool can benefit learners, outline the underlying measurement scales and their rationale, summarise the tool's psychometric properties, and provide guidance and some resources for adapting the tool to other professional and work-integrated learning programmes.

Sustainable benefits to learners

Collecting learning and teaching-focussed feedback assists us to provide the best quality training possible, based on accurate information. The benefits for the learner stems both from improving the quality of a programme and from the process of participating in providing structured feedback (as illustrated in Figure 1).

Participating in learning and teaching evaluations may benefit learners by:

- Encouraging self-reflection on their learning;
- Increasing their awareness of their own learning processes;
- Gaining insights into how to get the most out of the programme;
- Having a safe and autonomous way to express their own learning and internship experiences;
- Providing a medium for resolving issues they may have encountered;
- Providing an opportunity to consider their preparation to enter the workplace; and
- Contributing to course improvements for future students.

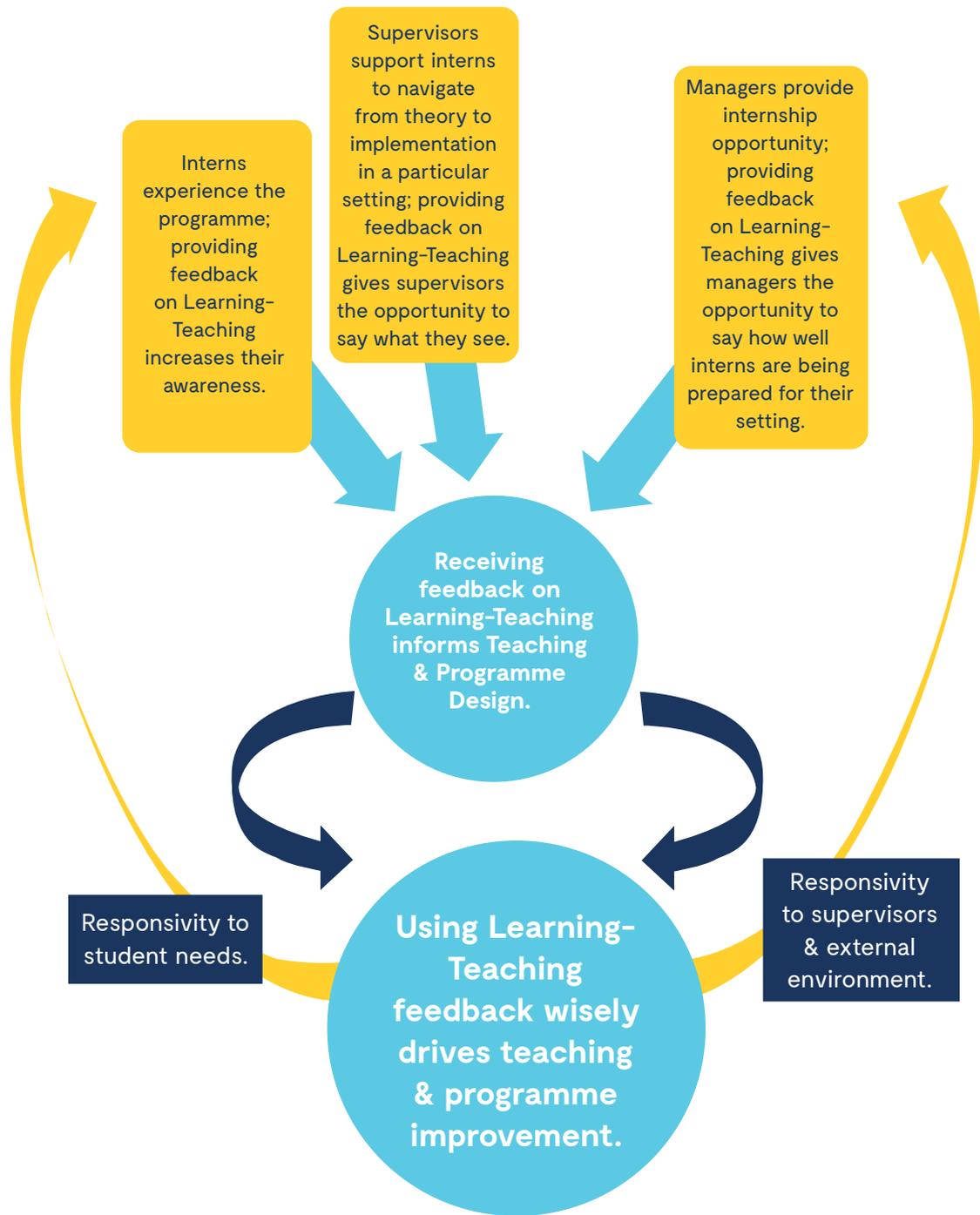


Figure 1: Multi-stakeholder feedback driving improvement in learning and teaching

Designing the multi-stakeholder evaluation tool

In higher education, the tools available for evaluating teaching fall mostly into three categories of purpose:

- (i) evaluation of large undergraduate programmes in which respondents might be completing diverse groups of courses over multiple years,
- (ii) individual lecturer feedback (data from these are also frequently applied to selection and promotion), and
- (iii) highly specific tools for bespoke programmes.

None of these categories address the types of issues pertinent to professional training, and even the individual lecturer feedback instruments provide little actionable feedback.

To design an evaluation tool, we must consider a programme's breadth and the important elements of a programme. We then develop a unique collection of questions (e.g., a scale) that help us evaluate the different facets of each element.

What sort of programme was this tool designed for?

The information is provided for context; in some instances, it is easier to identify how to adapt a tool if you can see how your situation is similar to, and different from, the one for which the tool was developed.

The professional psychology programme for which this instrument was developed is a one-year, post-masters internship programme which leads to professional registration. Students must complete a minimum of 1,500 hours of supervised practice. Typically, that equates to four or five days per week for a calendar year, in employment as an intern psychologist working with clients. In this psychology internship programme, alongside their workplace practice, students also have some academic learning which occurs in a cohort model under an integrated, competency-based curriculum. The internship is a transitional year in which students learn, for the first time, how to apply the scientific knowledge of the discipline in meeting the needs of clients. Work with clients occurs in the workplace, under the close supervision of an experienced psychologist. In addition, each intern has a University Supervisor who is also an experienced psychologist, who oversees the internship learning experience for each intern.

Where does the teaching and learning occur in this internship programme?

- In class teaching (condensed blocks of face-to-face teaching throughout the year)
- Knowledge and skills audits (informing development of a learning plan)
- Construction of a personal internship learning plan
- Workplace experience (working with clients and other professionals in the workplace)
- Independent learning
- Academic assignments
- Supervision-based learning with a university supervisor and field supervisor
- Participation with colleagues in semi-structured peer supervision meetings
- Viva exam

Measuring facets of the programme

Within an evaluation tool, each scale helps us address the different facets of the course elements. Each scale can also be adapted to assess the course elements from the perspective of different stakeholders.

For example, a set of questions asking *current students* about the suitability of the course's workload (e.g., the workload scale) can also be adapted to assess how appropriate the course workload is from the perspective of *supervisors* and *employers*.

By designing scales that are adaptable to different stakeholders, the tool's capacity to evaluate elements of the course from multiple perspectives is improved.

Our evaluation instrument considered 17 elements of the course, each of which was assessed for all stakeholders. The full instrument with the list of items assessing each element and overall psychometric properties of each scale is provided in the supplementary *Technical Reports booklet*. The 17 elements are:

- Teaching
- Block Course / Workshop
- Assessment
- Examination
- Workload
- Expectations
- Learning Outcomes
- Self-care and well-being
- Field supervisor
- University supervisor
- Psychological safety
- Relationships
- Pre-enrolment stage
- Enrolment process
- Resources
- Internship
- Overall evaluation

Depending on the complexity of a course element, each scale may have a different number of questions. For example, the Workload scale presented to students consisted of three questions (also referred to as 'items'; see Figure 2). A scale for each course element was presented to each stakeholder group. Where necessary the wording and selection of items was adjusted to account for the differing needs and perspectives of each stakeholder.

Figure 2: Three questions evaluating the workload element of the course.

In our evaluation tool, each stakeholder group (students, supervisors, employers) responded to these identical items.

To what extent do you agree or disagree with the following statements:

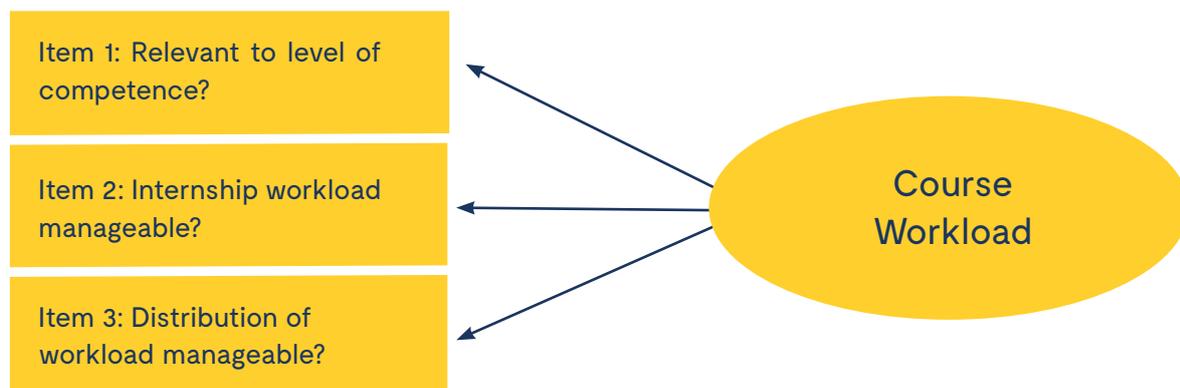
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	Not Applicable
In this programme, each of the tasks required (e.g. Reviews, Logs, Learning Cell, Video-Process Report) is relevant to reaching the level of competence of a beginning psychologist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Though demanding, the internship workload is manageable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The workload distribution across the year is manageable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Each item assesses a different facet of the course, so each item should be a distinct question about a specific facet. Yet, if the items are all related to a common course element (e.g., workload), the ratings of each item will be correlated to some extent. We have a good scale where the individual items (e.g., each of the three questions related to workload) are moderately correlated, but each separately describes a unique facet of the course element (see Figure 3 below).

In addition to the items that are each assessed on a quantitative basis, we also include open-ended questions that invite participants to comment on the programme's overall strengths and weaknesses – as well as programme improvements they would recommend.

Figure 3: Scale items representing different facets of a common course element.

If developed well, we can infer that each of the items we measure in our scale (rectangles) reflects a broader element of the course that is not directly measured (oval; in this case Course Workload). If the items reflect the course element well, the item scores may be averaged to calculate an overall evaluation score for the course element.



Psychometric properties of the tool

Each item of the scale is evaluated by respondents using a 6-point quantitative scale. As each item reflects a quantitative assessment, basic statistical analyses can be used to understand the psychometric properties of each scale. In our descriptions of each scale, we summarise item scores and overall scale scores using the mean and standard deviation. The level of relatedness between scale items (e.g., reliability) is described with the bi-variate correlations and the scale's overall reliability (using Cronbach's α (alpha); see glossary for further explanation of these statistical analyses).

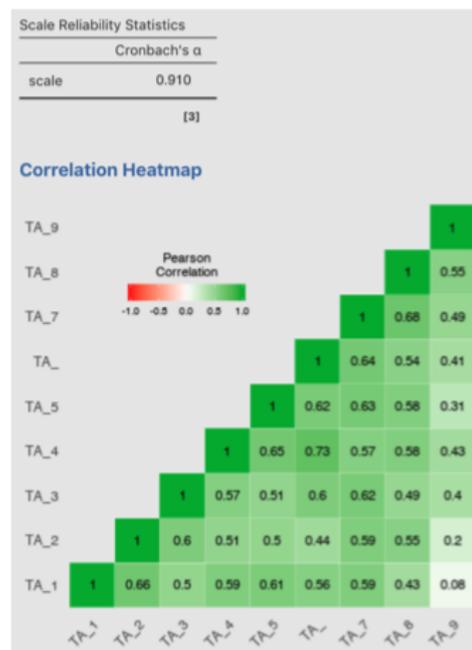
Below is an example of one of the larger scales, which explores graduate perceptions of the teaching they experienced in the programme (see Figure 4). The table on the left lists each item making up the scale followed by the average score (mean), variability of scores (standard deviation), and number of ratings made for the item (count; where students skipped answering a question, the count is lower). The shaded figure on the right examines the statistical relationship between the items. At the top is the overall Cronbach's α (alpha). Scores closer to 1.0 suggest the scale items are more related; typically scores above 0.70 are indicative of a reliable scale. The green matrix at the bottom depicts the correlations between each pair of items. Like Cronbach's α , correlations closer to 1.0 suggest scores on the two items are more strongly related; stronger correlations are coloured a darker shade of green.

Figure 4: Example analysis of a scale (Teaching).

Past Student Teaching Report – 2019

Teaching – Higher scores are greater agreement.

#	Question	Mean	Std Deviation	Count
TA1	The teaching staff were good at facilitating my learning during the block courses.	4.67	0.71	66
TA2	Incidental teaching (e.g. issues that came up in emails, phone calls or on-site visits) is helpful.	4.53	0.81	66
TA3	Individual teaching (e.g. if I was struggling with something) was helpful.	4.37	0.85	62
TA4	Teaching staff guided me and supported me to learn what I needed.	4.35	0.91	65
TA5	Teaching staff modelled critical thinking and critical evaluation of evidence.	4.67	0.73	66
TA6	Teaching staff modelled being respectful of different points of view.	4.50	0.86	66
TA7	The feedback I received on Reviews was helpful.	4.21	0.94	66
TA8	The feedback I received on my Reflective Activity Log was helpful.	4.21	0.95	66
TA9	I was confident that teaching staff would correct me if I was wrong (e.g. if I had misunderstood something or was not executing a skill appropriately).	4.58	0.68	66
TA_Scale	Median= 4.56	4.47	0.63	62



One of the challenges in achieving scale and item reliability from a multiple stakeholder tool, is that although finding parallels across the experiences of all stakeholders is possible, deriving parallel measures is more difficult. Supervisors, for example, have only indirect exposure to teaching that occurs during block courses, and few managers would have more knowledge of block courses than approving the leave to attend.

A further challenge is that professional programmes, unlike larger undergraduate programmes, have smaller student numbers. This makes robust testing of items and scales harder to ensure.

Adapting the teaching evaluation instrument

Adaptation for professional Psychology programmes

The instrument can be adapted for use by other professional psychology programmes. The scales and items for each stakeholder survey are available in the accompanying *Technical Reports booklet*, which reports statistical analyses of the items and scales to date. The wording of items can be customised to suit programme needs without substantively altering item meaning.

The evaluative potential of a standardised instrument created specifically for the needs of professional psychology training programmes can therefore be applied in a wide range of settings and allow programmes to gather quality information to inform any programme review. In this way, an increasing number of students will potentially gain benefits from continued improvements to professional psychology training.

As a discipline, psychology has the technical expertise to measure aspects of human experience and performance; we are often called on to provide that expertise to other professions, including education. As a profession, psychology is committed to striving to have evidence-driven practice. Developing this instrument constitutes one small step toward a more systematic and systemic use of evidence to drive our teaching of professional psychology.

Within the Scientist-Practitioner framework for the practice of psychology, the ideal (the top sequence coloured gold in Figure 5) is that theory drives research, producing evidence which drives practice, upon which we conduct research to refine theory and then further refine practice. The reality (coloured bronze in Figure 5) often falls sort of that goal.

Similarly, the ideal (the second sequence coloured gold in Figure 5) would be to identify any relevant theory of how to teach professional psychology to drive research; produce evidence about how best to teach professional psychology; and, implementing that, we would conduct research to further refine our practice of teaching. The reality more often resembles the second bronze sequence in Figure 5, in which how we were taught professional psychology combined with how we (personally) practise psychology is transformed through something of an internal 'black box' into how we teach professional psychology.

Similarly, *the ideal* practice model for teaching Professional Psychology would be for teachers to embody careful consideration of theory to drive research that could translate into evidence-based teaching practice. Evaluation of teaching practices will then allow one to refine theory (if necessary) and pursue further research to inform teaching. However, *the reality* resembles the idea that the way we teach is largely informed by how/what we were taught, how we personally practice, and our individual sense-making 'black box'.

By developing the Multi-stakeholder Feedback on Learning and Teaching (M-FoLT), we hope to have taken a step in moving the teaching of professional psychology towards the sequence represented in green in Figure 5.

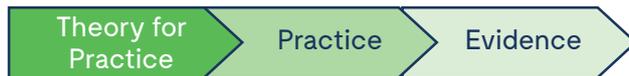
Figure 5: Scientist-Practitioner for Practice; Scientist-Practitioner for Teaching Professional Practice

Practice driven by evidence from research of a theory (Argyris & Schon, 1974).

The ideal

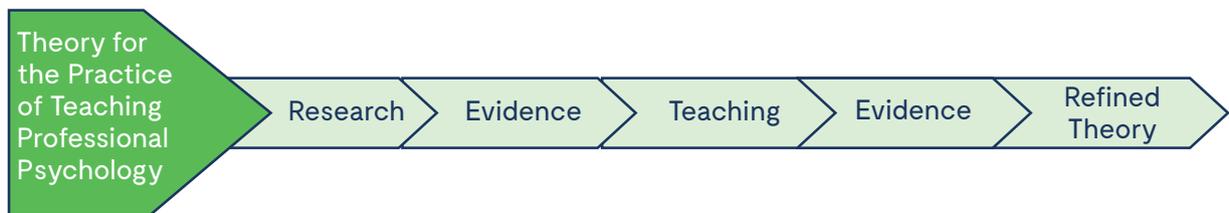


Often the reality



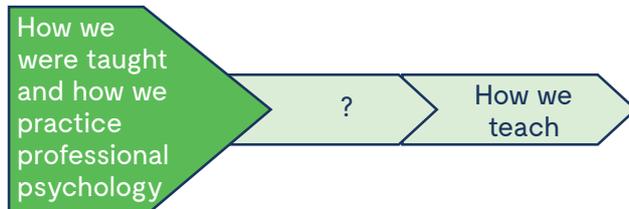
Teaching of Professional Psychology driven by evidence from research of theory for Teaching Professional Psychology:

The ideal

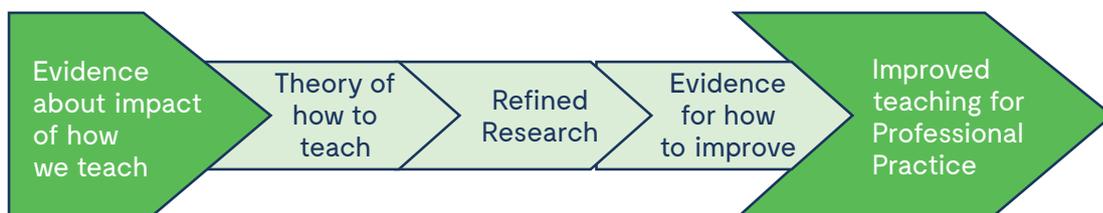


Teaching of Professional Psychology driven by whatever happens in the 'black box' in which we have personally made sense of how we were taught and how we practise:

Often the reality.



Working toward improving teaching for professional practice based on evidence:



Adaptation for other professional programmes

When considering how the tool can be best adapted, it is critical to reflect on the dimensions of the teaching programme's learning context so that the adapted measure can be fit for purpose. To gather data that can truly inform the development of your programme, your teaching, and yourself as a teacher, start by thinking about the dimensions of the learner, learning, teaching, and the contexts in which the learning and teaching occur (see Figure 6).

Begin by characterising your programme according to the following dimensions:

Learning Environment

- Key characteristics of the learning environment (e.g., in person/distance/mixed, directed/scaffolded, prescribed/facilitated, individual/group/cohort)

Learners

- Relevant characteristics of the learners (e.g., prior experience, other life commitments, values, learning preferences)
- Aim in completing this course/programme
- Experience of the key characteristics of the learning environment (which aspects are easy/difficult, (dis)liked, (un)helpful)

Learning and Assessment

- Types of learning (e.g., knowledge, understanding, critical thinking, skill, competency)
- Expectations of learners (learning aims to be achieved, scope of learning, tasks, types and amount of assessment, criteria for the standard to be achieved, time frames)

Learners' experience of the Learning and Assessment

- Alignment of aims of learner and course/programme
- Achievability of the learning aims through engagement with the course, teaching/supervision offered, and completion of the required tasks
- Clarity of expectations
- Extent to which assessment tasks help/hinder achieving personal/course learning aims
- Incidental learning (neither taught nor part of the design of the teaching-learning context)

Teaching

- Relevant characteristics of teachers/supervisors (e.g., qualifications, practice experience, warmth, availability, fairness)
- Strategies and techniques that help/hinder learning (e.g., culturally matched practices such as opening and closing with *karakia*; icebreakers; didactic/Socratic/interactive/demonstration/role play/multimedia; feedback that is timely, fair, kind, constructive, useful)
- Behaviours that facilitate/hinder/might be irrelevant to learning (e.g., inclusive versus judgemental language, examples from experience versus books, addressing individual learning barriers)

Context

- Where learning happens (in class, alone, independent group work, with clients, during reflection, in supervision and knowledge, techniques, skills, insight, etc.)
- Where teaching happens (in class, in 'hallway' conversations, individual consultations, in how feedback is written, and knowledge, techniques, skills, insight, etc.)
- Dimensions of context that might influence the learning (e.g., programme culture, relationships with other learners or work colleagues, workload, scheduling)?
- The resources that are needed and how available are they to learners

In summary, there are many factors to consider:

- about the learners themselves,
- what is to be learned and how,
- by whom and how it is taught and assessed,
- in what contexts, and,
- how all of these may interact.

Figure 6: The learner sits at the intersection of teaching and learning – both within and beyond the learning context.

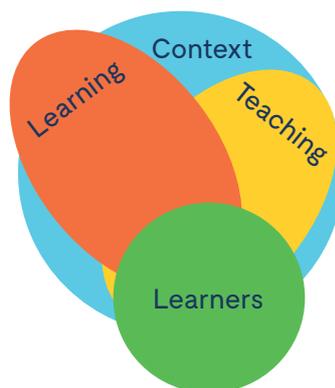


Table 1: Comparison of scales for professional programmes

Original version of M-FoLT	M-FoLT for Nursing	M-FoLT for Culinary arts
Teaching	Teaching	Teaching
Writing tasks	Writing tasks	
	Technical skills (e.g., injection)	Technical Skills (e.g., knife skills)
Examination	Examination	Examination
Workload	Workload	Workload
Expectations	Expectations	Expectations
Learning objectives	Learning objectives	Learning objectives
Supervision	Supervision	Supervision
Internship experience	Practicum experience	Apprenticeship experience
Internship setting	Rotations	Apprenticeship setting
Relationships (peers, teaching staff, supervisor, clients, workplace colleagues)	Relationships (teaching staff, supervisor, patients, nurses, doctors)	Relationships (teaching staff, chef, kitchen staff)
Enrolment	Enrolment	Enrolment
Resources	Resources	Resources
Overall satisfaction	Overall satisfaction	Overall satisfaction

To adapt the instrument for your own programme, consider each programme element and describe how it is manifested. The worksheet in Appendix B has been developed to help you identify the more (and less) critical elements of your teaching programme that should be evaluated and which stakeholders are most critical to include in designing your instrument.

In addition to integrating all the recommendations provided above, we would like to allude that to facilitate honest and committed participation it is essential to communicate the purpose and benefits of multi-stakeholder evaluations to the audience. Also, it is useful to report back to the audience with a summary of the feedback gathered. Working in partnership with students and the workplaces will empower a truly integrated learning environment.

References

- Argyris, C., & Schon, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. Jossey-Bass.
- Broomfield, D., & Bligh, J. (1998). An evaluation of the 'short form' course experience questionnaire with medical students. *Medical Education*, 32(4), 367–369. DOI:[10.1046/j.1365-2923.1998.00232.x](https://doi.org/10.1046/j.1365-2923.1998.00232.x)
- Carless, S. A., Robertson, K., Willy, J., Hart, M., & Chea, S. (2012). Successful postgraduate placement experiences: What is the influence of job and supervisor characteristics? *Australian Psychologist*, 47(3), 156–164. DOI: [10.1111/j.1742-9544.2012.00085.x](https://doi.org/10.1111/j.1742-9544.2012.00085.x)
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011) *Program evaluation: alternative approaches and practical guidelines* (4th ed). Pearson Education.
- Fortune, T., & McKinstry, C. (2012). Project-based fieldwork: Perspectives of graduate entry students and project sponsors. *Australian Occupational Therapy Journal*, 59(4), 265–275. DOI:[10.1111/j.1440-1630.2012.01026.x](https://doi.org/10.1111/j.1440-1630.2012.01026.x)
- Freestone, R., Williams, P., Thompson, S. M., & Trembath, K. (2007). A quantitative approach to assessment of work-based learning outcomes: An urban planning application. *Higher Education Research and Development*, 26(4), 347–361. DOI:[10.1080/07294360701658575](https://doi.org/10.1080/07294360701658575)
- Gross, S. M. (2006). The student perspective of psychology practica training. *Administration and Policy in Mental Health Services Research*, 33(3), 264–6. DOI:[10.1007/s10488-005-0023-1](https://doi.org/10.1007/s10488-005-0023-1)
- Jiang, Y. H., Lee, S. W. Y., & Golab, L. (2015). Analyzing student and employer satisfaction with cooperative education through multiple data sources. *Asia-Pacific Journal of Cooperative Education*, 16(4), 225–240.
- Khuong, C. (2015). *Internationalising tourism education in Vietnam: An evaluation of the work-integrated learning process in tourism training programs*. RMIT University. <https://researchrepository.rmit.edu.au/esploro/outputs/9921861961701341>
- Nedeljkovic, M., Chaffey, L., Murray, G., & Brennan, C. (2014). Postgraduate clinical psychology placements in Victoria: The experience of students and supervisors. *Australian Psychologist*, 49(6), 348–357. DOI:[10.1111/ap.12067](https://doi.org/10.1111/ap.12067)
- Scott, T. L., Pachana, N. A., & Sofronoff, K. (2011). Survey of current curriculum practices within Australian postgraduate clinical training programmes: Students' and programme directors' perspectives. *Australian Psychologist*, 46(2), 77–89. DOI:[10.1111/j.1742-9544.2011.00030.x](https://doi.org/10.1111/j.1742-9544.2011.00030.x)
- Smith, C. (2012). Evaluating the quality of work-integrated learning curricula: A comprehensive framework. *Higher Education Research & Development*, 31(2), 247–262. DOI:[10.1080/07294360.2011.558072](https://doi.org/10.1080/07294360.2011.558072)
- Stirling, A. E., Kerr, G., Banwell, J., MacPherson, E., & Heron, A. (2016). *A practical guide for work-integrated learning: Effective practices to enhance the educational quality of the structured work experiences offered through Colleges and Universities*. Toronto, ON, Canada: Higher Education Quality Council of Ontario.
- Welch, B., Vo-Tran, H., Pittayachawan, S., & Reynolds, S. (2012). Crossing borders: Evaluating a work integrated learning project involving Australian and Vietnamese students. *Australian Academic & Research Libraries*, 43(2), 120–134. DOI:[10.1080/00048623.2012.10722265](https://doi.org/10.1080/00048623.2012.10722265)

Yap, C. (2012). Bridging the theory–practice gap: Work-based business learning. *International Journal of Learning*, 18(8), 127–139.

Zahra, A., & Pavia, C. (2012). A unique approach to work-integrated learning: Meeting industry, student and educator needs through academic and experiential elements. *ACPET Journal for Private Higher Education*, 1(2), 41–50.

Appendix A: Glossary of useful terms

Some terms in this glossary apply to the Technical Report booklet available as a companion resource.

Aggregated Data

Data combined from multiple respondents (or a single respondent over multiple occasions) and expressed as a statistically calculated summary (summary statistic is the technically more accurate term but for purposes of this booklet, we have adopted the more common usage).

Construct

An abstract, explanatory idea formed from the synthesis of simpler ideas that are related to each other as parts of a larger whole. Contrasted with concrete variables that are easy to observe (e.g., height, colour), constructs are more difficult to observe and measure (e.g., intelligence, satisfaction). Consequently, measuring constructs usually entails observing simpler, but related, phenomena.

Construct Validity

The degree to which a psychometric test or instrument is capable of measuring what it claims to measure. For example, in relation to a questionnaire developed to measure student satisfaction with a course, the construct validity would be the extent to which the questionnaire measures students' course satisfaction with minimal influence from other variables that are unrelated to course satisfaction.

Correlation (see also Inter-item correlation)

A statistical summary of how strongly two scale items are related to each other expressed as a score between -1 to +1. Positive values suggest that as ratings go up (or down) for one item, the other item's ratings also go up (or down). Negative values suggest that as ratings go up for one item, the other item's ratings go down. Values closer to zero suggest the items are less related. See also, Cronbach's alpha.

Cronbach's alpha

A statistical summary of a quantitative measurement scale's internal consistency expressed as a score between 0 and 1. Scores closer to 1 are indicative of more cohesive scale. Conceptually, Cronbach's α is the average correlation between all possible pairs of items contained within a set of items. If an item on the scale does not correlate well with the other items, the Cronbach's α will be reduced. As a rule of thumb, a Cronbach's α of 0.70 or higher is generally considered good.

Data Custodian

The person who has the responsibility of looking after a dataset, ensuring its safe-keeping and that data access is available only in accordance with the Research Ethics' approval for the project. Often this person conducts statistical analysis and then passes on only de-identified, and/or aggregated data to other research project team members.

Domain

In this context, the term is used in the sense derived from set theory and relates to the set of elements over which a function is defined.

Evaluation

In this context, see **Programme Evaluation**

Inter-item correlation (see also Correlation)

Used as a measure of the internal consistency of a test; a measure of the extent to which each test item in a scale is correlated with the other items in that scale.

Internal Consistency

The extent to which items on a test are interrelated or homogenous. Internal consistency is an index of a test's measurement reliability. Correlations, Cronbach's alpha, and McDonald's omega are quantitative indicators of internal consistency.

Item analysis

Statistical process for evaluating the validity of a scale by examining how well individual items from the scale relate to—and discriminate from—one another. Often inter-item correlations are used in this process.

Item loading

Extent to which one item (question on a questionnaire) is related to the other items in a particular Scale. See **item analysis**.

McDonald's omega

Similar to Cronbach's alpha as a measure of internal consistency. McDonald's omega may be a more robust estimate of reliability than Cronbach's alpha under some conditions.

Norms (statistical)

Allow for comparison of an individual score with a relevant population. Development of norms for a psychometric instrument require accumulation of data from a relevant population. For example, in this context, the norms for psychology internship programmes may differ from norms for undergraduate-level Work-Integrated Learning programmes.

Norm-referenced test

A norm-referenced test (or psychometric instrument) compares individual scores (in this instance, scores from the responses of one student, supervisor or manager, or one programme) with a large group of similar students, supervisors or managers, or programmes. It is distinguishable from a criterion-referenced test in which it is possible to articulate standards which are un/acceptable (e.g., can/cannot execute a parallel park on a busy street in a single move).

Pilot Study

In this instance, a preliminary study undertaken with a small number of participants, designed to evaluate the measurement capacity of the questionnaires. It is undertaken in an effort to determine the appropriateness and effectiveness of items selected for the questionnaire. It may lead to modification or deletion of some items or identification of gaps to be addressed in future research.

Programme Evaluation

May be formative or summative. When summative, it is an appraisal process designed to determine the extent to which a Programme (of teaching, social intervention, etc.) is achieving the ends it claims.

Psychometric Instrument

A standardised instrument to measure one or more psychological constructs (e.g., satisfaction with a service).

Reliability

Overall consistency of a measure, meaning that it will yield similar results over subsequent trials. See also **Internal consistency**.

Robust (statistical)

A psychometric test is said to be robust if it provides useful insights into that which it is intended to measure, even if administered under conditions in which not all of its underlying assumptions are met.

Scale (in a psychometric instrument)

A Scale is comprised of a number of related items which together measure a particular construct.

Standard Deviation

A statistical measure of how dispersed quantitative data around the average value is. Very crudely, the standard deviation can be interpreted as the average distance of individual scores from the mean.

Standardised Instrument

In this context, a psychometric assessment instrument, the reliability and validity of which have been established through thorough empirical research and analysis with a sufficiently large and relevant population. Standardised instruments have clearly defined norms making it possible to compare similar individuals (or, in this instance, professional training programmes).

Subject Matter Experts

Individuals who possess expert, in-depth knowledge of the practical expression of a specific topic. For example, a Professional Programme Director is a Subject Matter Expert (SME) in that specific approach to professional training; to find an SME for the training in the building trade, you might consult a registered builder who has overseen many apprentice builders. The identification of the SME is defined in part by the specific purpose of the research.

Teaching Feedback

In this context, is a type of formative evaluation which is intended to improve or guide development of a course.

Unidimensional construct

A simple construct with only a single attribute. In contrast, a multidimensional construct has multiple facets making it more complex to study. For example, the construct of academic aptitude is multidimensional in that we often measure different facets including verbal aptitude and mathematical aptitude.

Validation Study

A research study to establish whether a psychometric instrument measures what it claims to be designed to measure; typically involves comparison with other known valid measures of the phenomenon under scrutiny. For example, a new measure to assess mathematical aptitude in students applying to study engineering might be administered to a group of newly graduated mathematicians whose mathematical aptitude is therefore known and provides a straightforward validation (and to a group of, for example, fine arts students who might reasonably be expected to achieve lower scores than students applying to engineering; a discriminant validation).

Validity

In reference to psychometric test construction, validity typically refers to construct validity (see **Construct Validity** above). Criterion and content validity contribute to achieving construct validity.

Appendix B: Worksheet for adapting the instrument

Below are listed the different programme elements we considered when developing the Multi-stakeholder Feedback on Learning and Teaching in Professional Programmes (M-FoLT) for our own programme. To adapt the instrument for your own programme, consider each programme element and describe how it is manifested. This should help you identify the more (and less) critical elements of your teaching programme that should be evaluated and which stakeholders are most critical to include in designing your instrument.

Programme Element	Description of element in our base professional psychology programme	Absence - or description of element in your programme
Programme Length	Stand-alone, one-year psychology internship.	
Overall Activity Types	1500 hours supervised practice + academic work.	
Academic Eligibility (min)	NZ Master's in Psychology	
Other Eligibility	Substantive experience in a similar role.	
Teaching Year	December to December.	
Delivery	Distance	
Required Reading	Set readings + texts (whole cohort) + individualised for field of practice.	
Teaching	Pre-internship (December prior to internship year), February, July; occasional webinars.	
Scope preparing for	Psychologist.	
Internship roles	Wide range of roles, settings, types of clients.	
Curriculum	Competency-based + Integrated.	
Key documentation for organising learning	Knowledge & Skills Audit.	
	Personal Internship Learning Plan.	
	Self-Care Plan.	

Supervision arrangements	Field Supervisor – 1 hour per week minimum, oversight of day-to-day work with clients.	
	University Supervisor – as needed, oversight of internship learning opportunities and progress, additional supervision for specific needs or in emergency.	
Assessable tasks	6 x 3000-word assignments (Review linking theory and practice via examination of Core Competency).	
	2 x Video Process Reports (Linking theoretical base and practice in relation to session video).	
	Reflective Activity Log (1 x Critical Reflection per week).	
	Examination - viva (no mock exam).	
Non-assessed tasks	Cell Group – monthly meeting.	
	In-class activities.	
Monitoring	Regular contact with University Supervisor.	
	Supervision Report – every 450 hours; Self-Assessment + Field Supervisor Assessment.	
Add elements critical to your programme		Describe additional elements



MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Ako
AOTEAROA