Assessment in Education: Principles, Policy & Practice

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/caie20

# Authentic assessment training for university teachers

Verónica Villarroel, Daniela Bruna, Carola Bruna, Gavin Brown & David Boud

**To cite this article:** Verónica Villarroel, Daniela Bruna, Carola Bruna, Gavin Brown & David Boud (08 May 2024): Authentic assessment training for university teachers, Assessment in Education: Principles, Policy & Practice, DOI: 10.1080/0969594X.2024.2350395

**To link to this article:** https://doi.org/10.1080/0969594X.2024.2350395

ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE
https://doi.org/10.1080/0969594X.2024.2350395

Routledge
Taylor & Francis Group



# Authentic assessment training for university teachers

Verónica Villarroel [a], Daniela Bruna [b], Carola Bruna [c], Gavin Brown [d] and David Boud [e]

[a]Facultad de Odontología y Ciencias de la Rehabilitación, Universidad San Sebastián, Concepción, Chile; [b]Facultad de Psicología, Universidad del Desarrollo, Concepción, Chile; [c]Facultad de Ciencias Biológicas, Universidad de Concepción, Concepción, Chile; [d]Faculty of Education and Social Work, University of Auckland, Auckland, New Zealand; [e]Centre for Research in Assessment and Digital Learning, Deakin University, Melbourne, Australia

**ABSTRACT**

A training course grounded in the principles of authentic assessment was implemented with 24 academics from two Chilean universities. Through a single-group pre-test/post-test design, the change in the evaluated parameters was analysed to determine the strengths and weaknesses of the conducted training. By the end of the course, the assessments were more contextualised, requiring the application of analysis and transfer skills, aligning with the competencies outlined in the graduation profile. Teachers included more open-ended questions focused on knowledge construction and fewer closed-response questions that required memorisation. Additionally, they reduced their beliefs about teaching as transmitting information, and the idea of provoking conceptual change in their students grew. This study shows that training in authentic assessment facilitates a change in the teacher's beliefs about teaching and assessment, towards teaching focused on conceptual change and assessment for and as learning. It also promotes the transformation of assessment instruments to be more realistic, challenging, and feedback-oriented.

**ARTICLE HISTORY**
Received 2 September 2023
Accepted 24 April 2024

**KEYWORDS**
Assessment; authentic assessment; higher education; professional practice

Improving the quality of assessment practices among university teachers and influencing their pedagogical mindset has emerged as a prevailing trend in academic development (Chalmers & Gardiner, 2015; Postareff et al., 2008). Assessment is one of the most persistent aspects of university education, and changing it is a formidable challenge (Sambell et al., 2020).There is a general weakness regarding alternative student-centred assessment methods and instruments (Barton et al., 2020; Pereira et al., 2016; Serrano et al., 2018).Unless university teachers effectively enact recommendations from assessment research, then poor practices will continue. The effectiveness of training courses depends on active learning and collective participation, as well as close alignment to the curriculum, existing teaching realities, and sufficient duration and continuance (Kaynardağ, 2019; Valiandes & Neophytou, 2018). It also depends on time and

**CONTACT** Verónica Villarroel veronica.villarroel@uss.cl Facultad de Odontología y Ciencias de la Rehabilitación, Universidad San Sebastián, Campus Tres Pascualas, Lientur 1457, Concepción 4080870, Chile
Supplemental data for this article can be accessed online at https://doi.org/10.1080/0969594X.2024.2350395.

opportunity to transfer knowledge into practice, with time for reflection (Hood & Houston, 2016; Mak, 2019).

While numerous courses and workshops aim to foster changed assessment practices, they are often not well-designed pedagogically and contextually, lacking impact evidence (van der Pol et al., 2020).The weakness in assessment is exacerbated because professors tend to work alone (Barton et al., 2020; Meijer et al., 2020),and there is a lack of systematic evaluation and monitoring (Valiandes & Neophytou, 2018). Indeed, few programmes use well-developed instruments to determine impact (Gibbs, 2013; Stes et al., 2013).

This article presents a training programme in authentic assessment in higher education, describing the conceptualisation that underpinned its design, implementation, and evaluation, and highlighting the obtained results.

## The problem of assessment in higher education

Research in higher education has highlighted the shortcomings of current assessment practices and the discontent expressed by students regarding their evaluation outcomes (Dunn & Sambell, 2018). Despite efforts made by institutions to instigate change, an examination-centric culture is still pervasive, particularly so in the ones in which this study was conducted. Conventional written examinations continue to dominate in higher education, primarily focusing on remembering and understanding, despite the existence of alternative approaches and methods (Constantinou, 2020; Davey et al., 2007; MacLellan, 2001; Matos et al., 2009; Mawa et al., 2019; Whittle et al., 2018). For example, evidence from 73 programmes in 14 UK universities indicates that two types of assessment predominate: essays and examinations (Jessop & Tomas, 2017). In a sample of 250 courses from five academic disciplines, exams were the most heavily weighted assessment within the final average, accounting for 78.8% of the grade (Lipnevich et al., 2020). In Spain, 1693 syllabi were studied from the first to the fourth year of undergraduate studies, revealing that traditional assessment approaches are still the most prevalent (Panadero et al., 2019).

The limitations of examinations and the need to address a broader range of outcomes have spurred a general shift in assessment to focus on learning effects rather than certifying functions (Carless, 2015). Assessment can now encompass scenarios that challenge and stimulate the acquisition of complex knowledge (Ibarra-Sáiz et al., 2020, 2021) and serve as learning opportunities for students, fostering their active and reflective participation in the assessment process (Allen & Parks, 2018; Yan & Boud, 2021).

The importance of promoting the transfer of learning from educational institutions to work has been emphasised (Jackson et al., 2019). Universities must offer opportunities for practice and simulation that consider the intricacies and shifts within the professional landscape (Garraway & Christopher, 2020; Garraway et al., 2015; Kulgemeyer & Riese, 2018). Assessment can align with university course objectives and the demands of the workplace, thus, providing context and significance to professional learning (M. Campbell et al., 2019).

## Authenticity in assessment

In authentic assessment, students typically engage in assessment tasks that replicate the complexities of the professional sphere. Through their performance on such tasks, students not only demonstrate the attainment of the intended learning outcomes but also showcase their grasp of the subject matter's knowledge and skills. Authentic assessments equip students with a range of competencies that greatly enhance their employability, offering them practical skills to be well-prepared for the workforce, and providing opportunities to apply practical and higher-order cognitive skills effectively (Akbari et al., 2022).

Authentic assessment can encompass a range of written and performance-based tasks, requiring attributes of realism, contextualisation, and problem-solving. These approaches share three key characteristics (Villarroel et al., 2017, 2019, 2021):

(1) Realistic Context:Assessments are contextualised closely resembling real-life or professional situations requiring students to apply skills and knowledge typically utilised in professional workplaces and real-world scenarios.
(2) Cognitive Challenge: Assessments involve tasks that demand the transfer and utilisation of higher-order cognitive skills, such as analysis, criticism, evaluation, or creation.
(3) Evaluative Judgement and Feedback: Students engage with criteria and standards that define high-quality performance and receive ongoing feedback based on the established quality standards and how their performance aligns with them.

Nonetheless, numerous higher education institutions have encountered challenges when attempting to embrace authentic assessment methodologies, especially given the prevalent culture of testing that prioritises rote memorisation and recall skills (Endedijk & Vermunt, 2013; İlhan-Beyaztaş & Senemoğlu, 2015). Educators often exhibit reluctance to overhaul their assessment practices, primarily due to the substantial demands placed on their time, energy, and intellectual resources (Brush & Saye, 2008; Deneen & Boud, 2014; Fletcher et al., 2022). Resistance to change can also be observed among students, who may not readily embrace shifts in assessment methods, especially if they have experienced success using memorisation-based approaches (Struyven & Devesa, 2016).

Limitations in the use of authentic assessment have been observed. Hobbins et al. (2021) found that authentic assessments were more frequently implemented in smaller advanced courses, although they were not completely absent from introductory courses with larger enrolments. Performance-based tasks and activities were notably more authentic than traditional tests, although tests continued to have a prominent place in the curriculum. Additionally, teachers often did not naturally or intentionally design authentic assessments in the classroom, as it was not a concept at the forefront of their thoughts.

To carry out a transition in the assessment system, academics require specific levels of expertise and knowledge that are not only related to their disciplinary expertise (Brown et al., 2022). The training of students to become professionals needs a well-defined, coherent, and integrated curriculum between disciplinary and professional knowledge (Karunanayaka & Naidu, 2021). This involves

providing and explaining to teachers the graduation profile of the programme they teach, the professional field of future graduates, the experiences, and issues that students face in their career practice to infuse teaching with realism and promote discussions about the necessary changes in assessment (Villarroel et al., 2021).

## Research hypotheses

Following participation in the authentic assessment training:

(1) Teacher-made assessments would be more realistic, contextualised, and closely aligned with the world of work.
(2) Teachers would create more challenging assessments, requiring application and transfer of knowledge.
(3) Teacher-made assessments would be less reliant on memorisation.
(4) Teachers would develop a student-centred teaching approach that focuses on student learning.
(5) Teachers would incorporate more feedback processes into the assessment.

This article aims to describe the principles, features, and implementation conditions of an intervention/training in authentic assessment for university teachers, as well as its main outcomes regarding beliefs, teaching approach, and assessment instruments.

## Method

### *Design*

A one-group pre-test-post-test design was evaluated with inferential statistical data analysis. Because all participants were included in the treatment, with no control group, this is a pre-experimental design (D. T. Campbell & Stanley, 1963) because despite having the logic of an experiment, it lacks randomisation.

### *Sample*

A sample of 24 university teachers, all with master's degrees,participated (14 females and 10 males). The average age was 45.6 years and the average length of academic experience was 16.7 years (range from 14 to 19 years). The pool of participants included four from each of six programmes (i.e. Psychology, Nursing, Business and Administration, Medical Technology, Bioengineering, and Speech Therapy). All were responsible for third- or fourth-year undergraduate courses, having a minimum of 5 years of teaching experience in that course.

Half of the group were affiliated with a medium-sized private university in Chile, while the other half were from a large, public, comprehensive Chilean university.

### Data collection process

Participants were recruited through programme directors of the undergraduate programs.The two inclusion criteria were: having taught in the undergraduate programme for 10 years or more and having the availability to participate in a training programme. Participants attended 10 weekly three-hour sessionsand dedicated 3 hours of independent work each week, meaning 60 hours of work over three months. The intervention included both group and individual tutoring to ensure that teachers applied the methodology and received help and feedback as needed. The pre-test data were collected at the start of the first training session and the post-test data at the end of the tenth training session. As baseline data, participants were asked to provide eight written tests (including both quizzes and exams) and four performance-based tasks (including instructions and rubrics) as examples of the assessments they administered to their students over the two previous years.

### Instruments

For data collection, four instruments were utilised.

#### Teachers' Perception of Authenticity in the Assessment System (TPAAS)

The TPAAS inventory was created for this study and measured perceptions of the authenticity of an assessment system. A sample of 238 university teachers had previously completed the questionnaire, which has 16 questions, with four questions per dimension. Five ordinal response options ranged from 'Never' to 'Always'. Principal Component Analysis (PCA) suggested there were four dimensions. The factors explained 46% of the variance andCronbach's alpha estimate of reliability averaged 0.76, ranging from 0.70 to 0.82. The four dimensions were:

(1) Authenticity and Contextualisation of Assessment.Links between learning assessment and activities carried out in the world of work. Tasks were framed in a context that had value beyond the classroom. For example: 'Design tests in which students have to apply what they have learned to situations that could occur in real life'.
(2) Responsibility for Authentic Assessment Feedback. Provides inputs that allow students to engage in continuous comparison of their work with others, using exemplars, self-assessment, and peer assessment activities as well as allowing students to have prior knowledge of the scoring criteria. For example: 'In the assessment planning, I consider activities in which students evaluate each other using guidelines that I provide to them'.
(3) Complex Knowledge Construction. Development of cognitively complex assessments. These instruments focus on the original application of knowledge and skills, rather than the routine use of facts and procedures within the specific disciplinary area. These tasks promote the analysis of problems to build knowledge and to generate a product that is significant beyond the classroom. This construction requires students to use organisation, interpretation, assessment, and

synthesis skills. For example: 'In written assessments, I construct questions in which students must analyse and apply knowledge'.

(4) Diversity of Assessment Strategies. Multiple indicators and diverse sources are used to assess and judge students' performance. Students may participate in the development of group work focusing not just on individual performance, in addition to the development of performance-based tasks, which go beyond tests. For example: 'I assess students' learning through various strategies, instruments, and methodologies, such as tests and performance-based tasks'.

### *Approaches to Teaching Inventory (ATI)*

The ATI (Trigwell & Prosser, 2004) used in this study had two dimensions: (a) Teaching aimed at conceptual change (student-focused); and (b) Teaching concerned with information transmission (teacher-focused). This scale is widely used in research in higher education and has demonstrated adequate psychometric properties (Prosser & Trigwell, 2006). In the Chilean validation (Montenegro & González, 2013), the conceptual change scale ($\alpha = .80$) explained 17.74% of the common data variance, while the teacher-focused factor ($\alpha = .79$), usingten items explained 16.55% of the common variance. The two factors had a moderate fit to the data (CFI = 0.90; RMSEA = 0.068).

### *Authentic assessment rubric to judge written tests*

Based on Villarroel et al. (2017) model of authentic assessment, the authentic assessment rubric was used to determine the degree of authenticity in teacher-made written tests after the intervention. See Supplementary Material Appendix 1.

Each item was rated for five criteria, using a 3-point scale:

(1) Context presence: degree to which the assessment had a rich context, providing information to solve the task, and locating students in a realistic, professional problem.
(2) Alignment with specific skills from the graduation profile: degree to which what was measured in the test related to the specific skills in the graduate profiles.
(3) Alignment with generic skills of the graduation profile: degree to which the assessment reflected the generic skills declared in the graduate profile.
(4) Knowledge construction: level of complexity of the cognitive ability that the item required to be solved, built from a range of cognitive taxonomies. The lowest level was given for recognition of information/memorisation, the middle level for information management/analysis, and the highest level for information transfer/decision-making.
(5) Global authenticity: contextualisation, problematisation, and realism present in the task to ascertain the degree to which the situation-problem was authentic to a real or professional life situation, in which curricular contents had to be applied.

Five expert judges had an inter-judge agreement of 92% indicating good consistency of scoring (Stemler, 2004).

### *Authentic assessment teachers' task*

To measure whether participants had achieved the skills expected in the training, a written performance test was designed. Participants were given a test scenario and asked to create a range of authentic tasks, using the variety of task-types taught in the course. Expert judgement indicated that the teachers' tasks could be accurately rated (i.e. consensus rate = 87% and internal consistency $\alpha = 0.89$).

## *Training program*

The training programme design was evaluated, and content validated by 10 recognised worldwide specialists. Experts were selected based on their extensive academic and research expertise in the field. Seven were from universities in Europe, two from universities in North America, and one from Oceania. Among these experts, the authors of this article did not participate.

The inter-rater correlation coefficient among judges was $r = 0.82$. All agreed with the content and the authenticity of the activities of the training (100% approval). The judges' comments were considered, and the training programme was adjusted. Adjustments were made related to rephrasing the training's learning outcomes (83.3% approval) and the sequence of some sessions (71.43% approval). The evaluation guidelines are included in Supplementary Material Appendix 2. The training programme aimed to meet 10 conditions:

(1) Teacher training in situ. The training was carried out adjusted to the disciplines of the participants, considering eachprogramme's graduate profile, and contextualised in local cultures and workplaces (Boud & Brew, 2013).
(2) Theory & evidence-focused training. Based on a systematic review of the authentic assessment literature (Villarroel et al., 2019) the course developed the capacity to implement a model of authenticity.
(3) Voluntary participation but with clear commitments. From the beginning of the training, the time commitment to participate was clearly established, along with associated tasks that were to be completed after the training.
(4) Teacher learning community. During the training, there were opportunities for group work, peer review, presentations of products and assignments, participation, and dialogue in an atmosphere of respect, acceptance, and learning, which facilitated the development of an authentic learning community.
(5) Use of prior diagnostic evidence. Before commencing training, the participating teachers' degree programmes and subject programme outlines were requested, along with tests and performance tasks used by the teachers in the last 2 years. With these inputs, a diagnosis was made regarding the level of realism, cognitive complexity, curricular alignment, and alignment with the graduate profile of the tests and tasks submitted. In this way, this diagnostic overview was presented, collectively analysing opportunities for improvement following the principles of authentic assessment.
(6) Realistic tasks linked with the workplace.Before training, data regarding employers and recent graduates of each programme were requested from the participating teachers' degree programmes. Interviews were conducted with these

individuals to inquire about prototypical challenges faced in the world of work. These examples were shared during the training to facilitate the development of realistic assessments.

(7) Challenging and transferable tasks. Participants completed tasks involving the progressive improvement of their own written test items, performance-based tasks, rubrics, and feedback processes.
(8) Development of evaluative judgement. Participants engaged in self-assessment and peer assessment activities of their assessment products. The aim was to develop the ability to independently judge their assessment designs after the training.
(9) Feedback and formative assessment. Throughout the entire training process, both descriptive feedback for improvement and dialogical feedback related to the products that the teachers were constructing were provided.
(10) Follow up.In the subsequent semester, participants met with members of the training team. In these, they showed their assessment instruments, received feedback, discussed the problems they had in terms of application, and received advice.

In Supplementary Material Appendix 3, more information is provided about the intervention with a brief description of each session.

### *Data analysis*

Inferential statistical analyses were performed to address the hypothesis of the study with a significance level set at $p < .05$. The analyses included: (a) Descriptive statistics, (b) *t*-Test to Hypothesis 1, (c) Generalised Hierarchical Linear Model (GHLM) regression and Mixed Effects Model tested hypotheses 2 and 3, (d) Wilcoxon test of difference tested hypotheses 4 and 5. The data was processed using IBM SPSS V20.0.

### *Ethical safeguards*

An appropriate institutional review board approved this research.

## Results

The results of the measurements taken in the intervention group, before and after training are presented. It is important to remember that the study lacks a control group, which represents a limitation when interpreting the causality of the intervention.

### *Influence on teachers' attitudes about assessment and teaching*

Because survey response data were not normally distributed, a Wilcoxon Test of difference was used on the change score. Table 1 shows statistically significant changes in teachers' overall orientation towards teaching. An increase in the score of the teaching dimension for conceptual change and a decrease in the score in the dimension of

**Table 1.** Change in participants' Approaches to Teaching Inventory (ATI) and Teachers' Perception of Authenticity in the Assessment System (TPAAS) scores.

| | Pre to Post Score Change | | |
|---|---|---|---|
| Variables | *M* | *SD* | Cohen's*d* |
| ATI | | | |
| Teaching centred on conceptual change/Student | 0.27 | 0.38 | 0.51** |
| Teaching centred on information transmission/Teacher | −0.22 | 0.39 | −0.38** |
| TPAAS | | | |
| Authentic assessment feedback responsibility | 0.40 | 0.94 | 0.44* |
| Authenticity and Contextualisation of Assessment | 0.29 | 0.60 | 0.43 |
| Diversity of authentic assessment strategies | 0.32 | 0.79 | 0.43 |
| Complex Knowledge construction | 0.09 | 0.50 | 0.17 |

* = $p < .05$; ** = $p < .01$; *** = $p < .001$.

teaching as information transmission is observed. This is a change in the level of teachers' perception regarding the approach that underpins their teaching practices.

Regarding the authenticity scale of the assessment system, there was only one statistically significant change in the 'feedback responsibility' dimension following the training. A comprehensive analysis of Table 1 allows us to interpret that the results support Hypothesis 4, as teachers shifted their teaching focus towards students and their learning, reducing teacher-centred information transmission. In contrast, Hypothesis 5 regarding the change in feedback incorporation was only partially supported because there was a change in only one of the four dimensions of the authenticity scale.

## *Impact on assessment practices*

To measure changes in written tests due to the training programme, tests used before and after the programme, developed by the same person for the same course, were compared using an ordinal regression hierarchical linear model. Table 2 shows the changes in the type of items used before and after the intervention, with Cohen's *h* effect size for independent ratios or probabilities. No statistically significant changes were detected in the use of each type of open-ended items individually, but when analysed together, their use increased by more than 20%.The use of multiple choice decreased by 10% and in total, closed-ended items reduced by more than 20%. Together these shifts indicate the

**Table 2.** Changes in item type used during the assessments, before and after the training programme.

| Item type | % Before | % After | Cohen's *h* |
|---|---|---|---|
| Total open response | 59.8 | 82.4 | 0.51*** |
| Short written response | 25.3 | 38.0 | $0.27^{ns}$ |
| Case analysis | 21.4 | 30.4 | $0.21^{ns}$ |
| Simple problem solving | 3.7 | 4.1 | $0.02^{ns}$ |
| Long written response | 9.4 | 9.8 | $0.01^{ns}$ |
| Total closed response | 39.4 | 16.8 | −0.51*** |
| Fill in the blank | 2.0 | 1.3 | $-0.06^{ns}$ |
| Multiple choice | 20.5 | 10.5 | −0.28** |
| True or false | 16.9 | 5.0 | $-0.40^{ns}$ |

Items not classified are deleted from table; ns = not significant; * = $p < .05$, ** = $p < .01$; *** = $p < .001$.

**Table 3.** Changes in authentic assessment dimensions, before and after the training programme.

| Dimension | Before *M* (*SD*) | After *M* (*SD*) | Cohen's *d* |
|---|---|---|---|
| Relationship with specific skills of the graduate profile | 1.60 (.54) | 1.96 (.44) | 0.65*** |
| Relationship with generic skills of the graduate profile | 1.23 (.40) | 1.49 (.47) | 0.64*** |
| Context presence | 1.62 (.73) | 1.98 (.65) | 0.50** |
| Knowledge construction | 1.67 (.58) | 1.93 (.43) | 0.46*** |
| Global authenticity | 1.53 (.49) | 1.83 (.42) | 0.63*** |

** = $p < .01$; *** = $p < .001$.

training has supported greater use of open-response tasks over closed-response format. These results support Hypotheses 2 and 3, in that the programme helped increase the use of open-ended response items associated with knowledge construction and decreased the use of closed-ended response items associated with memorisation skills.

Table 3 shows changes in the dimensions of Authentic Assessment. All dimensions increased by medium to large effect sizes following the programme. Thus, after the intervention, participants constructed and used written tests with more contextualised items, linked with skills from the graduate profile, which promoted knowledge construction in their students.These results support Hypotheses 1 and 2. Consistent with Hypothesis 1, after training the teachers seem to construct more realistic, contextualised, and problem-oriented assessments. Likewise, for Hypothesis 2, the increase in 'Knowledge construction' after training indicated that the assessments were more challenging, aiming for application and transfer of knowledge.

## *Performance in authentic assessment design*

The exam in Session 10 required participants to create authentic written test items, complex tasks, and rubrics. Table 4 shows that the teachers performed well in all assessed items, except for those related to problem-based learning, one form of authentic assessment. Participants were capable of drafting contexts to place students in real scenarios in which they had to solve problems using higher-order thinking skills. Creating authentic short written response items and rubrics to mark them was also a strong general ability. Constructing a problem-based learning item and a rubric to assess it were the skills least acquired.This result appears to be stronger than that seen in Table 2 analysis of participant tests. Overall, teachers incorporated the dimensions of authentic assessment (realism, cognitive complexity, and feedback) into their written tests and performance-based tasks.

**Table 4.** Degree of authenticity in teachers' design of written tests, performance-based tasks, and feedback after their participation in the programme.

| Item | M | SD |
|---|---|---|
| I. Context | 2.92 | 0.27 |
| II. Short written response | 2.74 | 0.42 |
| III. Rubric for short written response | 2.44 | 0.78 |
| IV. Case analysis | 2.22 | 0.94 |
| V. Problem based learning (PBL) | 2.05 | 0.90 |
| VI. Rubric for problem-based learning | 1.81 | 1.17 |

**Table 5.** Compliance with authentic assessment activities by teachers.

| Activity | Expected Number | *M* | *SD* | Max | % Accomplishment |
|---|---|---|---|---|---|
| Performance-based Task Feedback | 1 | 0.76 | 0.44 | 1 | 76 |
| Performance-based Tasks | 2 | 1.48 | 0.75 | 2 | 74 |
| Performance-based tasks Rubrics | 2 | 1.43 | 0.68 | 2 | 72 |
| Written Test Feedback | 1 | 0.71 | 0.46 | 1 | 71 |
| Written exams | 4 | 2.67 | 1.39 | 4 | 67 |
| Written Test Rubrics | 2 | 0.86 | 0.91 | 2 | 43 |

### *Compliance with the application of the authentic assessment methodology*

As coursework, participants were asked to transform three written tests and two performance-based tasks from their courses. In addition, they built two analytic rubrics to grade tests, and two to review performance-based tasks, along with performing two feedback processes, one for each of these types of instruments.

The level of compliance with these requirements was high (Table 5). On average, teachers completed 7.9 ($SD = 3.15$; median = 8) of the 12 planned activities, which gave 66% compliance. The results showed that the activity with the highest compliance was the realisation of a feedback process for a performance-based task, followed by feedback for a written test and the execution of two performance-based tasks. The least successful activity was the creation of rubrics for written tests with a large difference to the overall average (using probit method for proportions Cohen's $d = .98$). These results can be explained by the complexity of the assigned task and the infrequency with which it was carried out in the prior practice of these teachers.

## Discussion

This study investigated how pedagogical training and a shift towards authentic assessment can influence the assessment practices of university teachers. Results show changes related to the design and implementation of authentic assessment training. Two aspects changed: the quality of assessment instruments and teachers' beliefs regarding teaching approaches.

The teachers achieved greater contextual authenticity, presenting realistic problem situations closer to everyday life and the professional world that students needed to address. Additionally, the teachers' tests required application of higher-order cognitive processes, involving analysis, application, using knowledge to solve problems, decision-making, designing, diagnosing, intervening, proposing changes, and criticising solutions based on the learned curriculum. Participants also created more open-ended questions and decreased the use of closed-response questions. These changes are relevant because open-ended questions tend to focus on the construction of knowledge and the transfer of skills better than closed-response items, which often focus on the recall of memorised material. These results may be due to the training's design, which was based on a clear and well-founded model of authentic assessment, allowing a profound exploration of key dimensions of authenticity. Instances of practice, collaborative work, self-evaluation, and co-evaluation were provided, fostering the development of evaluative judgement and the incorporation of quality criteria regarding authentic assessment.

Teachers expressed a shift from teacher-centric approaches to more student-centric approaches. They moved from understanding teaching as oriented towards information transmission to teaching focused on provoking conceptual change in students. This change is related to the design and implementation of authentic assessment training, which began by presenting teachers with a diagnosis of their assessment practices, allowing them to identify opportunities for improvement through ongoing reflection and timely feedback.

Unfortunately, two elements did not change: teachers' beliefs about the authenticity of their assessments and the implementation of feedback activities with guiding rubrics. In this sense, it is important to note that the change in beliefs about teaching was substantially greater than the change in the perception of authenticity in assessment. That is, teachers increased their perception of their responsibility to provide feedback to their students, but they did not progress in their perception of how realistic, diverse, and challenging their assessments were after training. Despite significant changes were observed in the cognitive complexity and contextualisation of questions when analysing the tests, teachers did not perceive significant changes in these areas. These results may be due to the need for more time and practice to display self-efficacy in the assessment area than what was included in the training. The same happens with rubrics; it requires more time and practice to feel skilled in constructing them.

The results suggest that authenticity faces challenges in implementing certain kinds of performance-based tasks (i.e. problem-based learning methodology, the use of rubrics, and feedback). Therefore, according to the results, teachers are more likely to make progress in making more authentic written tests than in other aspects. The transformative effect of the training is likely limited to that practice, at least immediately after completing the programme. This may be related to the fact that the training design provided a detailed diagnosis of written tests but not of performance-based tasks and rubrics, allowing for greater insight into the former.

## *Practical implications*

There are several practical implications of this study. Some are organisational, specific to the educational institution, potentially impacting academic work and university teaching. Furthermore, it underscores valuable lessons concerning the design and implementation of teacher training programmes.

Primarily, it is important to note as a practical implication the benefit of supporting teacher training programmes in tangible evidence drawn from the teachers' own practices. This approach fosters introspection, allowing teachers to see themselves reflected in the showcase data, analyse their practices, engage in self-assessment, and acknowledge the imperative for change, alongside the associated challenges and impediments. Moreover, the information can be escalated to university authorities who can make evidence-based decisions to direct the efforts to facilitate the expected change.

In the presented teacher training programme,a previous analysis of the assessment instruments employed by teachers over the preceding two years was conducted.In the analysis of the tests, before the intervention, a notable prevalence of questions focused on the retrieval and recognition of data was found, predominantly observing closed-response assessment formats such as multiple-choice, true/false items, completion, and

matching terms. It is important to note that the prevalence of objective, closed-response tests does not signify a sole emphasis on memorisation skills, since when multiple-choice questions are properly formulated, they can allow the measurement of higher-order skills.

The weaknesses observed in the diagnostic evaluation could be a product of work overload due to the participation of teachers in multiple courses in the same semester, needing to grade tests quickly, as well as not having received formal training in university teaching or assessment instrument design. Having this pre-intervention diagnosis, based on evidence, facilitated understanding among teachers about the imperative to transform their assessment practices because, on one hand, they mainly measured literal repetition of information, and on the other hand, they were misaligned with the competencies delineated in the graduation profiles. Using evidence from their own practices is also consistent with the principles of authenticity and acknowledges local contextual disparities as proposed by Boud and Brew (2013). This diagnosis allowed each participant to analyse the strengths and areas necessitating improvement in assessment in relation to an authentic assessment model, providing a contextual framework to initiate work and efforts toward a paradigmatic shift in assessment practices.

The second relevant practical implication underscores the need for higher education institutions to ensure working conditions for teachers, including enough time for the design, implementation, and grading of assessments. Constructing instruments under the principles of authentic assessment requires time for conceptualisation and the design of contexts that resonates with the course content and undergraduate profiles. Additionally, it requires time to review and grade student responses. To transition away from mechanised assessment, mandates time and opportunities for peer collaboration. Investing time in formulating contextualised, realistic, and cognitively challenging assessments, in which teachers provide feedback on student performance, contributes to deeper student learning, and promotes the retention of acquired knowledge. Furthermore, it also develops competencies aligned with graduate profiles, thereby enhancing students' preparedness for the demands of the workplace.

The third important practical implication involves ensuring the internal coherence of the implemented teacher training programme. In this study, the training addressed each dimension of the theoretical model of authentic assessment that was used. A deliberate instructional was employed to target instruction on the areas slated for transformation, with products that would be used in the course after the training being constructed during the sessions, as well as in the required assignments, providing prompt feedback. This reinforces the obvious: to improve assessment practices in higher education, adequate training opportunities for academics must be provided.

The fourth implication is related to the authenticity of the training. Teacher training poses a challenge as not all teachers exhibit interest in self-improvement, and even when they do, they may not always perceive the training as beneficial. However, training initiatives can be efficacious if they adhere to the authenticity guidelines outlined earlier. It reflects their practices and reality, adjusts to their needs, fosters active participation, promotes collaboration among peers, includes evaluative judgement, provides timely feedback, includes challenging activities transferable to other teaching contexts, supervises and follows the process of implementing change, integrates theory and practice in

a way that makes sense to the teacher and establishes a connection with the competencies that students must demonstrate in the world of work.

The fifth important implication, particularly relevant from the authenticity model employed in our study, is the possibility for higher education institutions to manage and provide information concerning real-world problems that graduates encounter in the workplace. This approach can provide ideas for an instructional design aligned with future professional demands. In this context, in the conducted training, participating teachers were provided with 'prototypical problems' from each discipline, constructed from interviews with employers, alumni, and practice supervisors from different professions. These problems served as an inspiration to create new problem scenarios, thereby contextualising and enriching knowledge.

### *Limitations and future directions*

The limitations of this study relate to four primary aspects. Firstly, the disadvantages of pre-experimental research designs for establishing the impact of the intervention. Secondly, the small sample size diminished the generalisability of findings. Thirdly, the lack of post-intervention follow-up on teachers regarding their practices and evaluative beliefs hinders a comprehensive understanding of sustained effects. Lastly, the study lacks measures evaluating the impact of authentic assessment methodology on student learning. For future studies, it is necessary to transition towards research designs conducive to establishing causality (i.e. experimental random allocation to training conditions, including a control group and post-intervention change monitoring).

Creating authenticity in assessment, however, may create problems in terms of the effort needed to prepare and grade authentic assessments. Hence, research is needed to determine the feasibility of these assessment approaches within contexts of low staff-student ratios and strong administrative compliance requirements. In research-intensive contexts, the path of least resistance (i.e. reliance on multiple-choice testing) is a strong siren call that may act against authenticity in assessment. Further research in such contexts may help develop manageable solutions.

While this small-scale intervention study shows positive shifts in approaches to assessment, it does not show that students will necessarily exercise deep learning or higher-order thinking skills when faced with such assessments. Theoretically, this type of assessment ought to prepare students for making decisions and solving problems in real professional contexts. However, the current study does not yet demonstrate that this is the case.

## Conclusion

If teachers are authentically instructed to assess their students' learning under the principles of authentic assessment, they are capable of learning to do so. This article illustrates that authenticity can be taught and applied across various disciplines, generating positive outcomes in teachers, at least immediately after completing the training. Trained teachers presented a notable shift in their assessment tools, successfully applying the principles of authenticity.

The design and implementation of the authentic assessment training programme presented here posed a significant challenge: the need to gather data from various sources, including assessment tools, graduate profiles, graduates, and employers. This involved a substantial educational management effort at the institutional level. In essence, authenticity was emphasised byincorporating realistic inputs. This allowed teachers to analyse their practices within their disciplinary context, identifying and addressing existing gaps. This approach could be a relevant aspect to consider in facilitating the application and transfer of their acquired knowledge to practice, providing elements to foster the development of high-order cognitive skills aligned with the specific demands of the workplace demands, and ensuring the fulfilment of the desired graduate profile.

The primary strength of the research presented in this article lies in the coherence between the characteristics of teacher training and the principles of authentic assessment that guided its implementation. The training approach was characterised by its realism, contextualisation, and adaptability to address the weaknesses identified among participants, considering prior diagnoses and graduate profiles for each respective programme. It was challenging, involving numerous performance-based activities, and providing continuous feedback to the teachers, including self-assessment and peer assessment activities. This intrinsic coherence between the training programme and its epistemological and conceptual foundations may seem evident. However, dissatisfaction from educators regarding the lack of alignment between the intended content of teacher training programmes and their delivery methods is not uncommon. We highlight this coherence, as it likely played a pivotal role in the achievement of the desired outcomes regarding the alignment of the new assessment practices with the principles of authenticity, which despite being expected based on the intended design of the training, might not have been achieved otherwise.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Notes on contributors

***Dr. Verónica Villarroel*** holds a PhD in Psychology from the Pontifical Catholic University of Chile. Currently, she works as a researcher and Director od the Master´s in Education for Health Sciences at Universidad San Sebastián. Her research interests include atuthentic assessment for learning, as well as dialogical teaching and remote education process in higher education.

***Dr. Daniela Bruna*** holds a PhD in Psychology from Universidad de Concepción. Currently, she serves as the director of the undergraduate Psychology program at Universidad del Desarrollo in Concepción. Her research focuses on the study and evaluation of teaching practices and interventions for students self-regulation in learning.

***Dr. Carola Bruna*** holds a PhD in Biological Sciences from the Universidad de Concepción. She currently works as faculty member in the Faculty of Biological Sciences and she is part of University Teaching Directorate. Her research focuses on training in effective pedagogical practices al university level.

***Dr. Gavin Brown*** is the Director of Quantitative Data Analysis and Research Unit and Professor of Learning, Development, and Professional Practice at the Faculty of Education and Social Work. He

is an Honorary Professor at University of Hong Kong and an Associate Professor at Umea University in Sweden. His research focuses on cross-cultural assessment.

***Dr. David Boud*** is Alfred Deakin Professor, and Founder & Co-Director of the center for Research in Assessment and Digital Learning (CREADLE) at Deakin University. He han been involved in research on assessment in higher education for over 40 years, and is one of the mos cited authors in this fileds.

## ORCID

Verónica Villarroel http://orcid.org/0000-0002-3000-2248
Daniela Bruna http://orcid.org/0000-0001-7424-2959
Carola Bruna http://orcid.org/0000-0003-0554-7147
Gavin Brown http://orcid.org/0000-0002-8352-2351
David Boud http://orcid.org/0000-0002-6883-2722

## References

Akbari, M., Nguyen, H. M., McClelland, R., & van Houdt, K. (2022). Design, implementation and academic perspectives on authentic assessment for applied business higher education in a top performing Asian economy. *Education + Training*, *64*(1), 69–88. https://doi.org/10.1108/ET-04-2021-0121

Allen, T. L., & Parks, M. A. (2018). Improving student learning through assessment: A mixed-methods study of the impact of exam wrappers in a large undergraduate course. *Teachers & Teaching*, *24*(5), 530–547. https://doi.org/10.1080/13540602.2017.142190

Barton, G., Baguley, M., Kerby, M., & MacDonald, A. (2020). Investigating the assessment practices within an initial teacher education program in an Australian university: Staff perceptions and practices. *Australian Journal of Teacher Education*, *45*(3), 34–47. https://doi.org/10.14221/ajte.2020v45n3.3

Boud, D., & Brew, A. (2013). Reconceptualising academic work as professional practice: Implications for academic development. *International Journal for Academic Development*, *18*(3), 208–221. https://doi.org/10.1080/1360144X.2012.671771

Brown, M., Nomikoudis, M., Bouilheres, F., Nkhoma, C., & Akbari, M. (2022). Implementing a strategic campus-wide approach to authentic assessment: Considerations and recommendations for implementation. *Higher Education, Skills & Work-Based Learning*, *12*(5), 914–927. https://doi.org/10.1108/HESWBL-05-2021-0100

Brush, T., & Saye, J. (2008). The effects of multimedia-supported problem-based inquiry on student engagement, empathy, and assumptions about history. *Interdisciplinary Journal of Problem-Based Learning*, *2*(1), 21–56. https://doi.org/10.7771/1541-5015.1052

Campbell, M., Cooper, B., Rueckert, C., & Smith, J. (2019). Reimagining student employability: A case study of policy and practice transformation. *Journal of Higher Education Policy and Management*, *41*(5), 500–517. https://doi.org/10.1080/1360080X.2019.1646379

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.

Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, *69*(6), 963–976. https://doi.org/10.1007/s10734-014-9816-z

Chalmers, D., & Gardiner, D. (2015). An evaluation framework for identifying the effectiveness and impact of academic teacher development programmes. *Studies in Educational Evaluation*, *46*, 81–91. https://doi.org/10.1016/j.stueduc.2015.02.002

Constantinou, F. (2020). Examination questions as a form of communication between the examiner and the examinee: A sociolinguistic perspective on assessment practice. *Cambridge Journal of Education*, *50*(6), 711–728. https://doi.org/10.1080/0305764x.2020.1761293

Davey, G., de Lian, C., & Higgins, L. (2007). The university entrance examination system in China. *Journal of Further and Higher Education*, *31*(4), 385–396. https://doi.org/10.1080/03098770701625761

Deneen, C., & Boud, D. (2014). Patterns of resistance in managing assessment change. *Assessment & Evaluation in Higher Education*, *39*(5), 577–591. https://doi.org/10.1080/02602938.2013.859654

Dunn, L., & Sambell, K. (2018). Bridging the gap: A review of assessment literacy in higher education. *Teachers & Teaching*, *24*(3), 297–308. https://doi.org/10.1080/13540602.2017.1413031

Endedijk, M., & Vermunt, J. (2013). Relations between student teachers' learning patterns and their concrete learning activities. *Studies in Educational Evaluation*, *39*(1), 56–65. https://doi.org/10.1016/j.stueduc.2012.10.001

Fletcher, A., Kickbusch, S., & Huijser, H. (2022). Authentic learning using mobile applications and contemporary geospatial information requirements related to environmental science. *Journal of Geography in Higher Education*, *46*(2), 185–203. https://doi.org/10.1080/03098265.2020.1850657

Garraway, J., & Christopher, L. (2020). Change laboratories between university and work in emergency medical care. *Studies in Continuing Education*, *42*(2), 196–212. https://doi.org/10.1080/0158037X.2019.1703668

Garraway, J., Wickham, S., & Bronkhorst, J. (2015). Between college and work in the further education and training college sector. *South African Journal of Education*, *35*(1), 1–8. https://doi.org/10.15700/201503070028

Gibbs, G. (2013). Reflections on the changing nature of educational development. *International Journal for Academic Development*, *18*(1), 4–14. https://doi.org/10.1080/1360144X.2013.751691

Hobbins, J., Kerrigan, B., Farjam, N., Fisher, A., Houston, E., & Ritchie, K. (2021). Does a classroom-based curriculum offer authentic assessments? A strategy to uncover their prevalence and incorporate opportunities for authenticity. *Assessment & Evaluation in Higher Education*, *47*(8), 1259–1273. https://doi.org/10.1080/02602938.2021.2009439

Hood, C., & Houston, D. (2016). Evaluation of longitudinal impact from a university teacher preparation program. In M. Davis & A. Goody (Eds.), *Research and development in higher education: The shape of higher education* (Vol. 39, pp. 106–116). Fremantle.

Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., & Boud, D. (2021). The quality of assessment tasks as a determinant of learning. *Assessment and Evaluation in Higher Education*, *46*(6), 943–955. https://doi.org/10.1080/02602938.2020.1828268

Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., Boud, D., Rotsaert, T., Brown, S., Salinas Salazar, M. L., & Rodríguez Gómez, H. M. (2020). The future of assessment in higher education. *Revista Electrónica de Investigación y EvaluaciónEducativa*, *26*(1). https://doi.org/10.7203/relieve.26.1.17323

İlhan-Beyaztaş, D., & Senemoğlu, N. (2015). Learning approaches of successful students and factors affecting their learning approaches. *Education and Science*, *40*(179), 193–216. https://doi.org/10.15390/EB.2015.4214

Jackson, D., Fleming, J., & Rowe, A. (2019). Enabling the transfer of skills and knowledge across classroom and work contexts. *Vocations and Learning*, *12*(3), 459–478. https://doi.org/10.1007/s12186-019-09224-1

Jessop, T., & Tomas, C. (2017). The implications of programme assessment patterns for student learning. *Assessment & Evaluation in Higher Education*, *42*(6), 990–999. https://doi.org/10.1080/02602938.2016.1217501

Karunanayaka, S., & Naidu, S. (2021). Impacts of authentic assessment on the development of graduate attributes. *Distance Education*, *42*(2), 231–252. https://doi.org/10.1080/01587919.2021.1920206

Kaynardağ, A. (2019). Pedagogy in HE: Does it matter? *Studies in Higher Education*, *44*(1), 111–119. https://doi.org/10.1080/03075079.2017.1340444

Kulgemeyer, C., & Riese, J. (2018). From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching*, *55*(10), 1393–1418. https://doi.org/10.1002/tea.21457

Lipnevich, A. A., Guskey, T. R., Murano, D., & Smith, J. K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education Principles, Policy & Practice*, *27*(5), 480–500. https://doi.org/10.1080/0969594X.2020.1799190

MacLellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, *26*(4), 307–318. https://doi.org/10.1080/02602930120063466

Mak, P. (2019). Impact of professional development programme on teachers' competencies in assessment. *Journal of Education for Teaching*, *45*(4), 481–485. https://doi.org/10.1080/02607476.2019.1639266

Matos, D. A. S., Cirino, S. D., & Brown, G. T. L. (2009). Students' conceptions of assessment in higher education in Brazil. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 235–253). Information Age Pub Inc.

Mawa, B., Ali, M., & Haque, M. (2019). Level of learning assessed through written examinations in social science courses in tertiary education: A study from Bangladesh. *Journal of Teacher Education and Research*, *14*(1), 09–14. https://doi.org/10.36268/jter/1413

Meijer, H., Hoekstra, R., Brouwer, J., & Strijbos, J. (2020). Unfolding collaborative learning assessment literacy: A reflection on current assessment methods in higher education. *Assessment & Evaluation in Higher Education*, *45*(8), 1222–1240. https://doi.org/10.1080/02602938.2020.1729696

Montenegro, H., & González, C. (2013). Análisis factorial confirmatorio del cuestionario: Enfoques de Docencia Universitaria (ApproachestoTeachingInventory, ATI). *Estudios pedagógicos (Valdivia)*, *39*(2), 213–230. https://doi.org/10.4067/S0718-07052013000200014

Panadero, E., Fraile, J., Fernández Ruiz, J., Castilla-Estévez, D., & Ruiz, M. A. (2019). Spanish university assessment practices: Examination tradition with diversity by faculty. *Assessment & Evaluation in Higher Education*, *44*(3), 379–397. https://doi.org/10.1080/02602938.2018.1512553

Pereira, D., Flores, M. A., & Niklasson, L. (2016). Assessment revisited. A review of research in assessment and evaluation in higher education. *Assessment and Evaluation in Higher Education*, *41*(7), 1008–1032. https://doi.org/10.1080/02602938.2015.1055233

Postareff, L., Lindblom-Ylänne, S., & Nevgi, A. (2008). A follow-up study of the effect of pedagogical training on teaching in higher education. *Higher Education*, *56*(1), 29–43. https://doi.org/10.1007/s10734-007-9087-z

Prosser, M., & Trigwell, K. (2006). Confirmatory factor analysis of the approaches to teaching inventory. *British Journal of Educational Psychology*, *76*(2), 405–419. https://doi.org/10.1348/000709905X43571

Sambell, K., Brown, S., & Bovill, C. (2020). Assessment for learning in higher education: Challenges and opportunities amidst change. *Higher Education*, *81*(2), 209–227. https://doi.org/10.1007/s10734-020-00504-6

Serrano, M. M., O'Brien, M., Roberts, K., & Whyte, D. (2018). Critical pedagogy and assessment in higher education: The ideal of 'authenticity' in learning. *Active Learning in Higher Education*, *19*(1), 9–21. https://doi.org/10.1177/1469787417723244

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating inter-rater reliability. *Practical Assessment, Research & Evaluation*, *9*(4). https://doi.org/10.7275/96jp-xz07

Stes, A., Coertjens, L., & van Petegem, P. (2013). Instructional development in higher education: Impact on teachers' teaching behaviour as perceived by students. *Instructional Science*, *41*(6), 1103–1126. https://doi.org/10.1007/s11251-013-9267-4

Struyven, K., & Devesa, J. (2016). Students' perceptions of novel forms of assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 129–144). Routledge.

Trigwell, K., & Prosser, M. (2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review*, *16*(4), 409–424. https://doi.org/10.1007/s10648-004-0007-9

Valiandes, S., & Neophytou, L. (2018). Teachers' professional development for differentiated instruction in mixed-ability classrooms: Investigating the impact of a development program on teachers' professional learning and on students' achievement. *Teacher Development*, *22*(1), 123–138. https://doi.org/10.1080/13664530.2017.1338196

van der Pol, J., Mattick, K., & Bloxham, S. (2020). Exploring the impact of assessment for learning workshops on academics' assessment design practices. *Assessment & Evaluation in Higher Education*, *45*(2), 233–247. https://doi.org/10.1080/02602938.2019.1643394

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2017). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, *43*(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396

Villarroel, V., Boud, D., Bloxham, S., Bruna, D., & Bruna, C. (2019). Using principles of authentic assessment to redesign written examinations and tests. *Innovations in Education and Teaching International*, *57*(1), 38–49. https://doi.org/10.1080/14703297.2018.1564882

Villarroel, V., Bruna, D., Brown, G. T. L., & Bustos, C. (2021). Changing the quality of teacher's written tests by implementing an authentic assessment teachers' training program. *International Journal of Instruction*, *14*(2), 987–1000. https://doi.org/10.29333/iji.2021.14256a

Whittle, R., Benson, A., Ullah, S., & Telford, A. (2018). Investigating the influence of question type and cognitive process on academic performance in VCE physical education: A secondary data analysis. *Educational Research & Evaluation*, *24*(8), 504–522. https://doi.org/10.1080/13803611.2019.1612256

Yan, Z., & Boud, D. (2021). Conceptualising assessment-as-learning. In Z. Yan & L. Yang (Eds.), *Assessment as learning: Maximising opportunities for student learning and achievement* (pp. 11–24). Routledge.