Development and implementation of the Assessment for Graduate Teaching instrument: A nationally applicable assessment of pre-service teachers’ performance

SUMMARY REPORT
MARCH 2018
Context and Purpose

Ensuring quality teaching within Australian schools is a continuing priority of the Australian Government. In 2016, following the recommendations of the Teacher Education Ministerial Advisory Group (TEMAG, 2014), the Commonwealth Government requested that the Australian Institute for Teaching and School Leadership (AITS L) assist initial teacher education providers with developing and introducing teaching performance assessment tools with the objective of ensuring:

- a more robust and consistent assessment of students who have undergone initial teacher education; and
- greater quality assurance of initial teacher education.

Following a competitive grant application process administered by AITSL, the University of Melbourne-led Consortium comprising 10 universities, successfully obtained seed funding to develop a Teaching Performance Assessment (TPA). The overall aim of this project was to develop a TPA that could be implemented in Initial Teacher Education (ITE) institutions as a summative assessment of pre-service teachers (PST), providing evidence of their classroom readiness. The resulting instrument, the Assessment for Graduate Teaching (AfGT) was subsequently designed and developed based on the following three principles.

The AfGT should be:

- Pedagogically and psychometrically defensible;
- Aligned with the practice of teaching and the practice of teacher education; and
- Fair in relation to diverse contexts in which teacher education is applied.

In acknowledgement of good design and measurement principles, the AfGT relies on multiple forms of data and evidence. Four elements were designed, trialled and consequently refined, to generate an assessment instrument that represents a robust, comprehensive and authentic series of tasks that reflect teaching. These four elements are:

1. Planning for impact on student learning;
2. Analysing teaching practice;
3. Assessing impact on student learning, and
4. Expanding practice.

Purpose of this Report

The main purpose of this report is to present a summary of the methodology and findings of the trial of the AfGT, namely the design of the instrument for the trial, instructions for its administration and scoring, technical information on its psychometric properties and the process through which all of this was achieved.
Development of the AfGT

Figure 1. shows the overall structure of the project based around three phases: design and development; pilot and test; and implementation. Underlying and influencing across these phases was a set of frameworks and guiding principles, which represented how the Consortium was organised, how communication between the Consortium and other key stakeholders occurred and how the project worked to achieve its deliverables.

Figure 1. Overall structure of the project.
**Consortium Structure**

The AfGT was designed and developed by a representative national Consortium comprising ten ITE providers in three States and two Territories.

In 2017, the Consortium members included:

- Charles Darwin University (CDU),
- Curtin University (CU),
- Federation University Australia (FUA),
- The University of Canberra (UoC),
- The University of Melbourne (UoM) (lead institution),
- The University of Newcastle (UoN),
- The University of Sydney (UoS),
- The University of Technology Sydney (UTS),
- The University of Western Australia (UWA), and
- Victoria University (VU).

The Consortium represented the diversity of ITE in Australia by geography, student population, program types (undergraduate and postgraduate), course offerings and delivery modes (face-to-face, online, and blended). This diversity provided a unique opportunity to design a TPA that had genuine national applicability yet was flexible and adaptable in its design to cater for the specificities of each context.

**Governance**

The governance structure of the AfGT is represented in Figure 2. As the Consortium was large and diverse, governance structures were paramount to ensure that all Consortium institutions had equal voice as well as capturing the voice of the entire profession in the development of the instrument.

The project was led by the UoM, specifically, the Melbourne Graduate School of Education (MGSE). The main Governing Body of the Consortium was the Consortium Committee. This Committee was chaired by the University of Melbourne and comprised one to three lead representatives from each institution. A leadership group and specific working groups were mobilised throughout the project to target aspects of the design, development, monitoring and assessment of the AfGT, and these groups reported to the Consortium Committee.
Figure 2. The governance structure of the AfGT.
Overall Methodology

The overall methodology of this project was developed around the three project phases: design and development; pilot and test; and implementation. In the design phase, systematic collection of an evidence base around ‘what already works’ in PST assessment was conducted to inform the development process. This was achieved through two activities: a systematic literature review and an evaluative audit of Consortium institutions’ current practices and tools. From this base, an initial design workshop established the core elements of the AfGT and further development occurred with the Design & Development Working Group.

For the pilot and testing phase, a mixed methods approach was employed through a concurrent triangulation design where both quantitative and qualitative data were collected and combined to cross-validate and corroborate results and findings.

Finally, the implementation phase for 2018 was informed by a process evaluation. This was a critical component of the project as it allowed for assessment and evaluation of instructional materials as well as implementation processes. Figure 3 shows the development, validation and moderation process for the AfGT. As can be seen, feedback loops occur each time data are collected, which in turn, informs further refinement and revision. This process is intended to continue in 2018 and beyond.

Figure 3. Process flow of instrument development and validation
The Design of the AfGT for the trial

Figure 4 illustrates the final design arising from the literature review, the audit of current practices and the scoping against relevant standards. It is important to acknowledge that whilst the AfGT is comprised of four elements, completed individually by the PST, each element is intricately related to all the other elements. As can be seen in Figure 4, there is a flow of information between the four elements. The overall result is a symbiosis of each element, reflecting teaching in its entirety.

![Diagram of AfGT elements](image)

Figure 4. Design of AfGT including domains of teacher practice and their corresponding AfGT elements.

**Element 1 – Planning for Learning and Teaching**

The aim of Element 1 is for the PST to demonstrate their ability to understand the student cohort, their learning needs and then to evidence this through their planning of a sequence of 5 to 8 lessons.

**Element 2 – Analysing Teaching Practice**

The aim of Element 2 is for the PST to demonstrate their understanding of the impact of their pedagogical practice on student learning. This is achieved by the submission of video-taped Key Pedagogical Segments (KPS) which are self-evaluated by the PST.
Element 3 – Assessing for Impact on Student Learning

This Element is concerned with the analysis of a summative assessment task, administered by the PST at the end of their planned sequence of five to eight lessons, which demonstrates their understanding of the extent to which their students have achieved learning goals.

Element 4 – Expanding Practice

Element 4 assesses PSTs’ capacity to make judgments on typical teaching circumstances that may not be assessed in Elements 1 to 3. This element provides the opportunity to assess the PST’s analytical reasoning in contexts outside their final placement. The scenarios to which PSTs responded were generated randomly from a bank of scenarios developed by the Consortium.

Overall Discussion

The design of the AfGT ensures that mentor teachers are included in the process of providing a PST’s practicum report as well as being actively involved in supporting the PST to complete the AfGT. The PST must do more than simply provide descriptive responses; they must also demonstrate that they are critically reflective, analytical and able to evaluate based on data gathered in relation to their students.

It is important that the AfGT be retained as a whole to ensure the fidelity of the instrument. It is not possible to ‘pick and choose’ which aspects will be incorporated into an institution’s assessment plan.
Assessment

Assessment in this context refers to the way the AfGT is scored for each PST. Assessment properties of the AfGT are conceptually separate to its measurement properties. However, the structure of the assessment process dictates the measurement properties of the tool and its psychometric soundness.

The AfGT was assessed using a purpose-built rubric and grading scale for each element. Each rubric was designed to communicate expectations of performance for the tasks in all the elements. It was the Consortium’s view that the criteria should be public, and the scoring rubric would allow the teacher educator and PST alike to evaluate criteria more precisely, as the assessment of teaching can often be complex and subjective. To that end, the scoring rubrics articulated the Australian Professional Standards for Teachers – Graduate Level on which performance was assessed and rated. They also contained definitions and examples to illustrate the attribute(s) to be measured, and the qualitative rating scale for each dimension, which was consistent across all rubrics. Each rubric clearly articulated the related standard and a progressive criterion to meet for each grade level.

Hierarchical categories (U, G, G-, G+) were designed to reflect and indicate levels of quality (ranging from unsatisfactory to excellent). The Consortium determined that to ensure the assessment could be embedded within the structure and regulation of any ITE program a progressive grading was essential.

Finally, scorer training was provided to teacher educators to ensure a minimum level of consistency. This reduced variability in participant scores due to grading and scoring so that any remaining variability could be attributed to individual performance difference between PSTs.
The Pilot Study

The pilot of the AfGT was conducted in 2017 with 116 PSTs completing the assessment. Of this sample, 86% were female, 89% were completing a Bachelor degree and 85% were completing a Primary specialisation. Reliability analysis was conducted on Elements 1 to 3 with results indicating good reliability on all scales (Element 1 $\alpha=0.82$, Element 2 $\alpha=0.86$, and Element 3 $\alpha=0.75$) and tasks within an element all contributing meaningfully to the structure of the assessment. In addition, a factor analysis confirmed the internal structure of all three elements demonstrating adequate relationships between each element but sufficient conceptual separation to be considered as assessing different but related aspects of teaching.

Element 4 was piloted separately on a larger sample (n=166). Analysis of grade distributions for Element 4 revealed some inconsistencies in grading depending on which set of teaching scenarios a participant received. This has highlighted the need to review each scenario and ensure consistency in the level of difficulty between scenarios, a task planned for 2018. Furthermore, Element 4 demonstrated only a small relationship with the other Elements of the AfGT and this has been further explored in 2018. Despite these results for Element 4, there is no intention to remove Element 4 from the AfGT. Situational Judgement Tests are considered a valid component of assessment into professions requiring analytical ability and the exercise of professional judgment, as they are able to tap into critical reasoning skills whilst also allowing for triangulation of skills assessed in other components of an assessment. Furthermore, Element 4 measures those standards that may be difficult to assess in all practicum contexts. In addition, feedback from PSTs on Element 4 revealed they valued this section of the assessment, hence Element 4 has demonstrated ecological worth and validity.

This was a challenging task that I found to be a valuable addition to the entire project. The scenarios that I was asked to address were multi-layered and required deep consideration. The act of having to choose from various answers and then being required to justify my response, felt very authentic. At times I felt as though some answers had equal value however, their timing and their possible position in appropriate sequences of response, meant that deep consideration of context needed to be made. (Pre-service Teacher)
Setting a Standard, Moderation and Evaluation

A moderation and standard setting workshop, run by the Measurement Working Group was conducted following the pilot study.

Moderation

Teacher educators across institutions were invited to record the grades they used in their evaluations of PSTs’ assessment responses. A Moderation and Assessment Survey was also administered where participants evaluated a selection of Element 1-4 responses, and graded them using both a letter (U, G-, G, G+) and numerical (1-10) scale. They also made explanatory comments about why a particular judgment was made. From this, numerical ratings for each letter grading were developed. Finally, to determine within and across institution scoring consistency, a moderation workshop was conducted. This process involved individual scoring of assessment examples followed by group discussion to reach scoring consensus. From this, and using the modal score for each task, consistency in scoring could be developed.

Standard setting

Determining the quality necessary to achieve a pass grade on the AfGT required a standard setting process, which was conducted as a standard setting workshop with multiple purposes. The workshop was designed to collaboratively engage the whole Consortium in the standard setting process, while at the same time determining any revisions that might need to be made to the AfGT for implementation in 2018.

Using a modified Body of Work process, mean scores on each element, score distribution and expert knowledge was triangulated to determine a cut score (pass grade) for the AfGT.

Process Evaluation

A process evaluation survey was conducted with stakeholder groups to understand the validity of the AfGT in situ and to determine potential implementation issues for 2018 and beyond. The survey asked respondents to reflect on the Elements of AfGT and estimate how long each element took to complete, then rate the elements across three dimensions: Clarity, Appropriateness and Difficulty. The results of this survey indicate that Completion Time and Task Appropriateness are adequate. Clarity of the tasks demonstrated some concerns for Elements 2 and 3; this will be addressed in 2018. The process evaluation was critical to gathering the voice of the end users of the AfGT and has provided invaluable feedback regarding implementation.
There is so much more to quality teaching practice than is 'measured' by the AfGT. However, it seems to do a reasonably good job of gathering evidence on the particular aspects of teaching it concerns itself with. It is of course, just a snapshot of academic progress in a single curriculum area. This is probably enough for beginning teachers, but should not be mistaken for telling the whole story of 'classroom readiness' - which a gradual and continuous process during the first few years of teaching. (Teacher Educator)

Implementation

In order to arrive at a vision for the future, Consortium members at the Moderation & Standard-Setting Workshop in November 2017, participated in a critical reflection to ensure that 11 months into the implementation, the objectives remained aligned and true to the original intent and spirit of the collaboration.

As a consequence, the following actions for implementation 2018 were agreed:

1. The AfGT, once refinements are concluded, will be suitable to all modes of delivery of ITE courses
2. The AfGT must become embedded in PSTs’ courses of study (including in the final practicum experience)
3. Fidelity within and across institutions must be maintained
4. All institutions will continue to have access to support materials (documents and assessor training video) that will provide guidance for all assessors
5. The AfGT must continue to be scored by initial teacher educators
6. Cross-institutional moderation (via Zoom) of assessment tasks and clarification of arrangements for assessment
7. Cross-institutional moderation and standard-setting will continue with initial teacher educators at the end of Semester 1 and again at the end of Semester 2 in 2018. Whole-of-Consortium moderation, must include detailed consideration of the extent to which the AfGT is suited for all specialisations
8. Information about the AfGT will be communicated to all Australian ITE providers directly and via the AfGT website, with an invitation for them to sight the full instrument.
The AfGT (Re)Design

In light of the recommendations that arose during the Moderation and Standard-Setting activities, the instrument has been re-designed, as described here.

Overview of the (Re)Designed AfGT

The AfGT is a multi-faceted assessment that comprises four elements:

- **Element 1: Planning for Teaching and Learning**
- **Element 2: Analysing Teaching Practice.**
- **Element 3: Assessing for Impact on Student Learning**
- **Element 4: Expanding Practice**

These elements, when taken together, enable Pre-Service Teachers to demonstrate the impact of their teaching on student learning across the elements of planning, teaching, assessing and making situational judgements.

Things to note about the re-designed AfGT:

1. Not all of the 37 *Australian Professional Standards for Teachers - Graduate level* – are assessed in the AfGT. This means that institutions will need to ensure that the 11 standards that are not assessed in the AfGT will need to be taught, practised and assessed elsewhere in the PST’s course of study.
2. Elements 1 to 3 build on each other, so it is important that they are part of the summative assessment in your final unit/subject/course, and are completed in order, from 1 to 3.
3. Whilst the AfGT is an integrated task, it is possible that the University may schedule Element 4 as the first piece of assessment because it can be completed independently of the other three Elements.
4. To ensure the fidelity or trustworthiness of the AfGT as a valid assessment, it is not possible to change or modify the AfGT. All items must be implemented as they appear in this document.
5. The University might require submission of additional assessment tasks in the final subject of the course of study. These tasks will not replace Elements 1 to 4 of the AfGT, but will be additional to the AfGT. The University will explain any additional requirements to their PSTs.
6. The AfGT does not take the place of the mentor or supervising teacher’s assessment of teaching in the final unit/subject/course, however it is a requirement that a PST passes both the AfGT and the professional experience placement.
7. Teacher educators from the University will be assessing the AfGT, however because the AfGT is an assessment task common to all universities in the Consortium, for purposes of moderation, it is possible that a Consortium member’s AfGT might be also be accessed and assessed by teacher educators from across the Consortium.

(Re)-designed assessment items at a glance

There are inter-related assessment activities for each of the Elements in the AfGT. Each element enables demonstration of meeting the Australian Professional Standards for Teachers – Graduate level. This is a summary of the assessment task instructions that are provided to PSTs:

**Element 1: Planning for Teaching and Learning** (approx. 750 words)

a) Submit Table 1\(^1\), which will demonstrate your ability to draw on your professional knowledge and skills, as well as your understanding of the context, to plan a learning sequence and associated assessment tasks, address outcomes and students’ needs and abilities.

**Element 2: Analysing Teaching Practice** (approx. 750 words)

a) Provide two unedited video-recordings – each of 6 to 10 minutes in length – of your teaching.

b) Provide a copy of your mentor teacher’s written feedback to the sections of lessons upon which your video-recordings are based.

c) Submit Table 2, which will demonstrate your ability to plan, teach, critically appraise your pedagogical approaches in two Key Pedagogical Segments, and incorporate feedback into your teaching.

**Element 3: Assessing for Impact on Student Learning** (approx. 900 words)

a) Provide a data display (table, chart and/or graph) that represents the class’ responses to an outcome-related summative assessment task

b) Provide de-identified samples of the summative assessment task, including your feedback, for three students who represent the diversity of the class or group.

c) Submit Table 3, which will demonstrate your ability to implement a summative assessment task, analyse data arising from this task, including de-identified, annotated samples from three students, and to nominate the next steps for teaching the whole class in general and the three students, in particular.

---

\(^1\) The tables referred to in this report are those that have been developed by the Consortium to scaffold the PSTs’ responses by providing prompts and an indicative word length for each response.
Element 4: Expanding Practice (approx. 600 words)

Links to a URL will be provided for you to demonstrate your capacity to make situational judgments about four typical situations that you may face during their teaching career, but which may not have been assessed elsewhere in Elements 1 to 3.

Final word

The AfGT is a work in progress as described. The AfGT is the result of collegial and collaborative actions of initial teacher educators from across Australia, who worked together to address the recommendations of the Teacher Education Ministerial Advisory Group that underpins the development of the National Program Standards.

AITSL’s Expert Panel found that the AfGT is a valid method for assessing whether a teacher’s performance meets the Australian Profession Standards for Teachers at the Graduate Level. The panel noted that this is a very well designed and executed project. The principles-based development of the AfGT has led to strong reliability and validity, even at this early phase. The Expert Panel believes that further reliability and validity data, and analysis of cross-institutional similarities and differences will strengthen this TPA as time goes on. The Consortium recognises the AfGT will benefit from continued adaptation and evidence of its predicative value, however is confident in its merit, worth and significance.