Title: Collaborative Assessments of Learning Outcomes: Generating Positive Institutional Change

Authors: Jacob Pearce and Daniel Edwards

Institution: Australian Council for Educational Research, Australia

Abstract

Insight into the development of higher education assessment collaborations in two fields (medical education and engineering education) is presented. These collaborations aim to improve and share assessment and assessment practice in order to enhance educational outcomes of students. This model is efficient and effective, can be applied nationally or internationally, and can generate positive institutional change around the globe.

Introduction

This paper provides an overview of some recent higher education assessment collaborations in two fields of study. Each collaboration was managed by the Australian Council for Educational Research (ACER). These collaborations, in the fields of medical education and engineering education aimed to improve the quality and facilitate the sharing of assessment items and assessment practice, in order to enhance the educational learning outcomes of students nearing the end of their degrees. Each collaboration involved the development of assessment frameworks and assessment instruments, the engagement of higher education institutions and students, the implementation of assessment instruments, and the reporting of outcomes.

These collaborative endeavours were conceived and deployed in an era of a growing higher education system both nationally and internationally, coupled with increasing calls for accountability in the sector, and the desire for quantitative metrics to help monitor and improve quality. While different approaches were used in the different projects, and the overall outcomes were distinctly different, each project was built on the ethic of collaboration between interested parties, and focussed on measuring learning outcomes in a useful and meaningful way. The different objectives, governance, implementation and outcomes are discussed, and highlight the different ways in which positive institutional change can be achieved by collaboratively working on the measurement of learning outcomes. Further, the different advantages of the collaborations illustrates that different approaches may be needed in different contexts, and a simplistic model of collaboration for any discipline would be naïve. Still, insights into the benefits of these successful collaborative approaches is offered in the hope that the work can be replicated more widely, and generate positive educational improvement in the higher education sector across the globe.

The paper begins by providing a brief context to the drivers in the higher education systems, to which the collaborative efforts were responding. It then explores the collaborations separately, starting with the medical education community in Australia and New Zealand, and finishing with the international engineering education community. The paper concludes by reflecting on the different collaborative approaches to the assessment of learning outcomes, highlighting some lessons learnt, and offering suggestions for future work in the international

higher education landscape.

Overall, the collaborative model of assessment has much to offer in the search for valuable metrics to measure the learning outcomes of graduates. However, there is much more that these models can offer aside from being tools to generate data for accountability and assurance measures. Collaborative models are also an efficient and effective way of generating a greater understanding of the educational attainment of students, and the means for improving learning and teaching within institutions. Collaborations in assessment can instigate positive institutional change if deployed appropriately.

Context

The projects presented respond to many international pressures in the higher education sector. As the sector continues to grow in significance and scale, there is a clear rationale for ensuring quality outcomes. There is a push for increased levels of regulation, with more regulatory bodies requiring more quantitative evidence that university graduates are reaching minimum standards.

High-quality university education is vital to a society. Of course, the usefulness of higher education for society can only exist if the educational processes are of optimal quality. Unfortunately, quality assurance of educational processes often comes only from within an institution and there are few chances to compare the quality of one's education to that of others. Accreditation processes seek to address this problem by repeated audits of an institution's educational processes and its supporting organisation. That in itself is an important process but it is only half of the equation, as this is predicated on the assumption that institutional—or departmental—level practices automatically lead to good graduate outcomes, which may not always be the case.

What is generally missing in this quality-assurance equation is actual evidence of the learning outcomes, competencies and educational gains that students have developed during their qualification. Coupled with the other information collected on graduate destinations, institutional processes, statements of graduate capabilities and other accreditation requirements, the use of empirical data that offers a measure of the learning outcomes of students can be highly effective in reviewing and evaluating the processes, teaching methods, curriculum used in individual courses and disciplines.

As such, the examples presented in this paper provide important insight into the potential benefits of developing learning outcomes assessments, especially if this development is done in a collaborative way that spans across a number of institutions or even across higher education systems.

There have been significant initiatives in the past decade that have helped to identify core competencies in a range of disciplines. Internationally, the Tuning Project sponsored by the European Commission has been an extremely important vehicle in achieving this outcome. Nationally within Australia, strong projects to identify key competencies in a range of disciplines have been facilitated by the Australian Learning and Teaching Council (now known as the Office for Learning and Teaching).

These projects have worked hard to identify areas in which graduates from certain disciplines should know and be competent in, and the next logical step is the development of tools to

begin to objectively assess whether these competencies have been achieved. However, there is always danger in these kind of assessment projects that a 'dumbing-down' could occur, or that developments of this kind may turn into regulatory 'batons' used as tools for de-registration of institutions or courses. The prospect of 'top-down' approaches, whereby certain practices are enforced on institutions is unpalatable.

The emphasis of this paper is on ensuring that these kind of projects emerge from collaborative ventures with institutions, that begin with the aim of improving—through a better understanding of student outcomes, and the use of outcomes assessments as one tool in what should be a large 'tool-box' for educational monitoring—quality assurance and most importantly, continuous improvement of student experiences and learning.

Collaborative efforts can potentially aim to provide a valid and reliable evidence base for institutional quality monitoring, while also maintaining ownership of the processes within the institutions themselves. With this 'bottom-up' model, data which is useful at the regulatory level can be generated through assessment initiatives, with the added benefit of improving the way that assessment is both developed and reported. The examples in this paper offer insight into how this has been happening and how it might continue to develop over the coming years both within national systems and across borders.

Collaborations in medical education

The main case study for this section, is the recent collaborative effort between 16 of the 20 medical schools in Australia and New Zealand. The Australian Medical Assessment Collaboration (AMAC) was instigated as a way of improving the quality of medical education through the recognition of the need for tools for comparison and evaluation of learning outcomes, and acknowledgement of the need for high quality assessment, and to share expertise in these areas (Edwards, Wilkinson, Canny, Pearce, & Coates, 2014; Wilkinson, Canny, Pearce, Coates, & Edwards, 2013).

In Australia, there is a growing acceptance and adoption of collaboration amongst medical schools. This collaborative culture has produced the Australian Medical Schools Assessment Collaboration (AMSAC), a group of seven medical schools which focus on developing biomedical sciences items which are embedded in university examinations around the mid-point of the medical degree, marking the transition from campus-based to clinic-based learning environments. There is also the Australian Collaboration for Clinical Assessment in Medicine (ACCLAiM); a group of four medical schools who collaborate on developing and using common clinical assessment instruments such as Objective Structured Clinical Examinations (OSCEs). Results are used for benchmarking graduate outcomes across schools. On top of this, some schools participate in the International Database for Enhanced Assessments and Learning (IDEAL); another tool used by numerous Australian medical schools for obtaining and sharing assessment items. IDEAL has 29 partner schools around the world that work towards generating databanks of summative and formative assessment items. Schools contribute to the databank and also generate assessments based on the items submitted by other schools.

Internationally, a well-known example of collaboration in medicine is progress testing in the Netherlands, where a number of schools have banded together to produce massive amounts of items for regular and continuous testing of students throughout their degree. Another example of sharing expertise in medical schools items is the UK's Medical School Council Assessment

Alliance (MSCAA) which is an alliance developed to create and generate shared assessment items as an alternative to having a national examination process imposed.

When AMAC begin in late 2010, it involved two medical schools in partnership with ACER. By 2012, AMAC had expanded to include almost every medical school in Australia and New Zealand. The schools were enticed by the possibility of working together to develop best practice in the development of common items for the assessment of learning outcomes. Medical schools were driven by a common pressing need to acquire high quality assessment material. The central tenet of the collaboration was the idea that cooperation between schools, such as 'pooling' resources, would improve the quality of assessment processes. The secondary upshot was the creation of an additional tool for monitoring the quality of graduating students at a time there was a lack of quantitative evidence. The schools saw an opportunity to create benchmarking data which could truly improve the learning and teaching within the universities, and become a means for ongoing self-regulation and leadership in the sector.

The model for developing the AMAC assessment framework (AMAC, 2012) and instrument was built on the idea of iterative development. Assessment items which were mapped to the framework were submitted to the consortium by participating schools. Test developers at ACER edited the structure of the items, performed some basic overall quality assurance, and prepared items for review in workshops. Workshops were conducted where clinicians from schools scrutinised, reviewed and revised the items. Medical educators were consulted before the items were prepared for pilot testing. ACER then collected item statistics and performed a range of validity and reliability analyses based on psychometric models to determine whether items were performing according to item response theories. Finally, items were calibrated and re-edited before being placed in the AMAC item bank. Since 2011, more than 170 items have gone through this quality assurance process, and over 1,400 students have taken AMAC items in both a formative online platform and as embedded items in summative in-school assessments.

AMAC offered powerful benchmarking potential between medical schools through de-identified reporting. Schools were given a valid and reliable evidence base to investigate whether they were meeting certain standards in the quality of their graduates, without this processes becoming one of naming and shaming. Each participating institution received a report which indicated the performance of their graduates as a whole on the assessment, benchmarked against other institutions with identifying information removed. Each school had a picture of whether they were meeting minimum standards and achieving quality in graduates either above or below others in the country, without being able to find out who the other institutions in the graphs were.

Further, by breaking down the reporting into disciplinary areas, schools could also determine whether there were gaps in their programs relative to the other schools. For instance, a school could determine if their students were performing below others at the national level in one sub-discipline, but not others. While this data was not based on representative samples of each entire sub-discipline, the data still proved to be a powerful diagnostic tool when used in combination with other information. There are many potential benefits of this type of data, which in this case allowed medical deans to diagnose aspects of their programs which required support and change.

At the same time, the results from the assessment were reported to individual students across

a wide range of categorisations, allowing students nearing the end of their degree some valuable feedback to aid their learning. These reports were confidential and were not delivered to institutions. Each student could see whether they were performing poorly across a range of sub-disciplines and medical contexts. Again, although this data was not definitive, it was a useful diagnostic tool which has the potential to become more useful for students in the future. Feedback from these reports indicated that students found the information incredibly helpful as formative feedback while they were preparing for their final examinations.

The AMAC structured 'bottom-up' approach of building a multi-institutional collaborative assessment for the evaluation of learning outcomes was founded on more general principles which are transferrable across a range of disciplines in higher education. More recently, the consortium have been developing a suite of resources for reference by future collaborative endeavours. The AMAC team wanted to share the experiences and lessons learnt in areas of quality, governance and dissemination, and implementation with others higher education sectors. These 'manuals' detail the approaches and processes for establishing a functioning and productive collaboration in higher education, while outlining major issues and considerations for such undertakings.

The AMAC consortium is now working towards establishing an International Medical Assessment Collaboration (IMAC), by extending the AMAC collaborative model into the international sphere. Preliminary discussions are happening with other medical schools (and assessment alliances) around the world.

Collaborations in engineering education

2010 – 2013 saw the one of the biggest worldwide collaborative projects in engineering education. The Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study was funded through the Organisation for Economic Cooperation and Development (OECD). The study was the first of its kind in the higher education sector. It involved the development of assessment frameworks and the testing of final year higher education students in three strands – Generic Skills, Economics and Civil Engineering. The explicit purpose of the study was "to see if it is practically and scientifically feasible to assess what students in higher education **know and can do** upon graduation' (OECD, 2011). In the Civil Engineering strand, 6000 students at 89 institutions in 9 countries participated in implementation. The study demonstrated that it was technically and practically feasible to do so across linguistic, cultural and institutional diversity. In particular, the Feasibility Study demonstrated the potential to provide institutions with valuable information about the relative performance of their students *without* the need for rankings or league tables.

The AHELO Feasibility Study was facilitated by a consortium of organisations lead by ACER. The consortium was given guidance by the OECD secretariat and through the OECD's Education Policy Committee. However, this collaboration was managed in a different structure than the projects described in the medical education section. In practice, AHELO ran in a top-down manner, with the overall management and information flowing from the OECD to the consortium, to the participating countries, to institutions and then to students. AHELO was designed and deployed analogously to the OECD Program for International Student Assessment (PISA), and many would argue that although the model is built around the notion of countries working together, the model is arguably less collaborative than the other projects listed in this paper.

As described in detail elsewhere (Coates & Richardson, 2011; OECD, 2013), the assessment instruments for the Feasibility Study were developed by the AHELO Consortium, primarily through the guidance of an Expert Group. In the case of the engineering instrument, this group included a number of international experts in engineering education, most with a specific focus on civil engineering. While much of the initial assessment development for the project was undertaken outside of the involvement of participating countries, the drafted instruments were provided to National Project Managers for consultation with experts within each country. Rather than beginning with the institutions, and a common goal and need for both assuring and improving the quality of graduates, processes and outcomes, AHELO began with the OECD and education policy representatives from OECD countries. Elsewhere, we have discussed the differences between the 'bottom-up' collaborative model which is the focus of this paper, and 'top-down' collaborations such as AHELO (Edwards & Pearce, Forthcoming).

One example of a collaborative initiative between engineering faculties across a number countries is the more recent Cross-Border Learning Outcomes Collaboration in Engineering (CLOCE). Building on the collegiality and momentum of AHELO, this project beginning in 2014 aims to develop a collaboration in the assessment of learning outcomes, but built from the 'bottom-up'. In part using the methodology realised in AHELO, along with the collaborative lessons learnt in AMAC, international partners will come together to collaborate on assessment materials. The ultimate goal will be to develop a new assessment instrument which can be used to benchmark engineering learning outcomes across a range of contexts. The initial focus in CLOCE will be on the domain of mechanical engineering. Other fields of engineering and levels of study will be considered as the project develops and processes and partnerships are established.

CLOCE will focus on assessing the ability of students to apply their skills and knowledge to solve real-world engineering problems. It will consider what students know and can do at the end of their bachelor level (first cycle) degrees, and their readiness to enter the engineering workforce. Currently educators in engineering from Australia, Japan and Canada are working on this collaboration, and while in its early stages, the potential for further expansion, while maintaining the collaborative emphasis, is significant.

Reflections on generating positive institutional change

There are many ways that a collaborative assessment of learning outcomes can be developed and deployed in a variety of national and international contexts. The model has much to offer as the international community continues to search for valid and reliable metrics and an evidence base of the attaining of learning outcomes of graduates. However, the more appealing (and almost secondary) outcomes of establishing an assessment collaboration are the benefits to higher education institutions which are above and beyond the simply search for data. We conclude by reflecting on several aspects of collaborative assessment with respect to the positive change that can be generated at the institutional level.

Accountability and assurances

Certainly, one of the main drivers for projects such as the ones mentioned in this paper is the lack of valid and reliable information on the attainment of learning outcomes of graduates at higher education institutions worldwide. The benchmarking data which has been generated from AHELO and AMAC for instance gives (or at least shows the potential to give)

institutions an evidence base on which to judge the quality of their programs in connection with similar programs both within and across educational systems.

One of the key facets of this data is the de-identified nature of the reporting. Institutions receive authentic, valid and reliable data on how they are performing against comparable institutions, but the reporting does not degenerate into a public 'naming and shaming' of the best and worst institutions through league tables. Furthermore, the data cannot end up being reported in this way, as all the raw data is housed in an independent research organisation, rather than one of the potentially commercially competing institutions.

However, the upshot of providing an evidence base on which to make judgments concerning accountability and quality assurance is just one aspect of collaborative assessment. There are other drivers and benefits of these projects which go beyond monitoring. Specifically, as will now be discussed, the outcomes of collaborative assessments are a more holistic tool.

Efficiencies

This model of collaboration has proven to be highly efficient. By working together, institutions could pool resources and achieve a large number of quality assessment items with minimum financial and time input. Institutions across the world have a need for quality items which have been scientifically validated and calibrated. These items are even more expensive to develop. The long process of iterative development, which includes the continual revision and re-working of items according to feedback mechanisms, such as item statistics and psychometric analysis, is a complex and time consuming process.

Assessment is a costly business. Significant amounts of time are spent within institutions developing assessment items. In AMAC, internal time spent on assessment was reduced in some institutions when AMAC items were embedded in summative assessment instruments. It is conceivable that in the future, a proportion of final year assessments may include shared items. While these items offer powerful benchmarking potential, they also reduce the amount of time, and thus investment, that institutions need to spend on developing their own internal items.

An additional benefit of the AMAC model is that the process of collaborative development acts as a form of professional development, improving the understanding of what constitutes a good assessment item, which in turn can lead to an increase in the quality of items developed within institutions.

Student benefits

In the projects described, student feedback has been overwhelmingly positive. When the AMAC items were delivered formatively, students preparing for their final examinations were given a new insight into their performance across a range of categories and measured against their peers. This allowed them to diagnose gaps in their knowledge and understanding and prepare more effectively for their examinations. In the AHELO assessment, students were presented with authentic engineering contexts and scenarios and asked to solve real-world problems, as they would do once they entered the workforces. Students appreciated being given an opportunity to test the applicability of their acquired skills and found the feedback useful.

Benefits to educators

As mentioned above, there are many efficiencies for institutions which can be achieved through collaborative assessment. But there are other benefits to educators in the sector. The most relevant here is the detailed reporting by sub-discipline which some of these projects have produced. When an institution can view some valid and reliable data highlighting aspects of their programs which are not meeting minimum standards, this provides the basis for a conversation to begin. That is, particular aspects of a program can be reviewed against data and new initiatives can be instigated. If educational improvement is to be continuously achieved, positive change at the institutional level can be initiated when problematic aspects are identified.

Assessment in the name of educational improvement

Finally, by changing the way assessment is conceive, deployed, and reported, the power of high-quality assessment is starting to be realized by institutions. By working together to achieve a common goal, and putting commercial interests and pressures aside, higher education institutions can improve the collective quality of their teaching, the student learning and, in turn, the quality of their graduates,

Some initial resistance from institutions offered an important lesson for future collaborations, as collaborating on the assessment of graduates does not always align with cultural or ideological constraints at the institution level. However, when institutions began working together, they found a great deal of commonality in the problems they face every day. By taking ownership of the collaborations, and by showing academic leadership, these higher education assessment collaborations have been a resounding success. The methods and outcomes presented have the potential to generate positive institutional change around the globe.

References

AMAC. (2012). *Australian Medical Assessment Collaboration: Assessment Framework*. Melbourne: Australian Medical Assessment Collaboration.

Coates, H., & Richardson, S. (2011). An internatinoal assessment of bachelor degree graduates' learning outcomes. *Higher Education Management and Policy, 23*(3).

Edwards, D., & Pearce, J. (Forthcoming). Outcomes Assessment in practice - Reflections on implementation of two different models. In H. Coates (Ed.), *Learning Outcomes Assessments*. New York: Peter Lang.

Edwards, D., Wilkinson, D., Canny, B., Pearce, J., & Coates, H. (2014). Developing outcomes assessments for collaborative, cross-institutional benchmarking : progress of the Australian Medical Assessment Collaboration. *Medical Teacher, 36*(2), 139-147.

OECD. (2011). AHELO Feasibility Study Website.   Retrieved March 25, 2014, from http://www.oecd.org/education/skills-beyond-school/testingstudentanduniversityperformancegloballyoecdsahelo.htm

OECD. (2013). *AHELO Feasibility Study Volume 1 - Design and implementation*. Paris: Organisation for Economic Cooperation and Development.

Wilkinson, D., Canny, B., Pearce, J., Coates, H., & Edwards, D. (2013). Assessment of medical students' learning outcomes in Australia : current practice, future possibilities. *Medical Journal of Australia, 199*(9), 578-580.