Using assessment of student learning outcomes to measure university performance: towards a viable model

Linley Margaret Martin

ORCID identifier 0000-0003-0279-3609

Thesis submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

May 2016

Melbourne Centre for the Study of Higher Education

Melbourne Graduate School of Education

University of Melbourne

Produced on archival quality paper

Abstract

This study investigates the possibility of developing a suite of performance indicators which could measure differences in universities' performance in attainment by their students of specified institutional or course-based learning outcomes. The measurement of learning outcomes has been the subject of active interest in higher education for over 20 years but to date there is no approach which has led to a sustainable generalised solution to this problem.

A four staged measurement model is proposed which explores the learning outcomes specified by universities, establishes a set of standards against which such outcomes could be assessed, and examines local assessment of students' learning for these outcomes to identify what graduates have learned and can do by the end of their study. Data on the grades achieved by individual students in local assessment tasks are then considered for use in a suite of institutional indicators which are designed to differentiate between universities in terms of the knowledge and skills demonstrated by their students. The focus of the study was to investigate whether the model could be applied to measure learning outcomes and institutional performance for Australian university undergraduate degrees.

The study showed that it was possible to derive a generalisable set of learning outcomes relevant to Australian universities and also a set of standards relating to each of these outcomes which could be used to grade assessments in a quantitative way for individual learning outcomes measurement. It was also possible to define a suite of quantitative performance indicators which appear to be valid for measuring differences in achievement for a subset of the specified learning outcomes. However it was discovered that Australian universities' current practice in describing and testing learning outcomes for subjects rather than courses or for the institution is different to the approaches commonly used internationally, requiring an adjustment to the model. Universities' practice in this is also different to the approach they espouse on their websites and in their assessment policies. The Australian approach requires a bottom-up model for measurement rather than the top-down model originally identified from international practice.

Various options are presented for types of local achievement assessment that are likely to produce the greatest consistency of learning outcome results between different

universities. The favoured option is a set of newly devised signature assessments to test achievement of cognitive learning outcomes which could be framed in a discipline context, but this is a contentious solution. The bottom-up model has face validity based on detailed analysis of the expected outputs from each of its stages, but it could not be fully tested because assessment data held in universities' repositories is not held at the level required.

Implementation of such a model, while appearing feasible, would have implications for policy, pedagogy, scholarship and practice within universities, and it would require a strong commitment from government and the sector for implementation to be successful. The benefits to students, staff, employers and the government would be substantial and appear to outweigh the costs associated with implementation.

Declaration

This is to certify that

(1) the thesis comprises only my original work towards the PhD,

(2) due acknowledgement has been made in the text to all other material used, and

(3) the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed … .....

Date ……23 May 2016…….

Preface

Publications from this thesis

Martin, L.M. (2014) *Assessing student learning outcomes: research trajectories* in Higher Education Learning Outcomes Assessment - International Perspectives

(H. Coates Ed. Vol. 6). Frankfurt, Germany: Peter Lang.

Related publications in higher education policy

Wells, J. & Martin, L.M. (2013) *Regulation* in The Dawkins Revolution: 25 years on (Edited by G. Croucher, S. Marginson, A. Norton and J. Wells). Melbourne, Australia: Melbourne University Press.

Martin, L.M. (2016) *Framing the framework: the origins of A Fair Chance for All* in Student Equity in Australian Higher Education: Twenty-five years of A Fair Chance for All (Edited by A. Harvey, C. Burnheim and M. Brett). Melbourne, Australia: Springer.

Acknowledgements

I would like to thank my primary supervisor Professor Richard James for his excellent guidance and sound advice during this study. He has stimulated me to think more broadly about this topic and been endlessly patient in getting me to challenge myself to think, analyse and write like a social scientist rather than the mathematician or bureaucrat that I once was. I am not sure I have succeeded in this but I hope I can now recognise the difference.

I also wish to thank Professor Hamish Coates for his supervision from 2013 onwards, for believing that it may be possible to measure learning outcomes using a method similar to that outlined in this thesis, and for giving me opportunities to publish some of the material from it.

Thanks are also due to Professor Simon Marginson who was my co-supervisor with Richard James for the first year of my candidature until he left the University of Melbourne in 2013 to take up a new post in the United Kingdom.

I consider myself very fortunate to have been guided by three supervisors who are so well respected in the higher education sector, and who have offered me encouragement, support and wise counsel during the course of this research.

At a personal level I have been supported by several good friends and PhD colleagues. I wish to acknowledge the friendship and encouragement shown to me by Marian Mahat and Joan McMeeken, who were fellow travellers on the PhD journey, and who shared their experiences. I also particularly wish to thank my friend and colleague Janet Beard for her attendance at seminars I gave, for her willingness to listen and discuss ideas that I had about my study, and the practical assistance she gave me with some of my IT challenges, which greatly improved the appearance of the thesis.

Finally I wish to dedicate this thesis to my father, David Martin, who believed in the transforming power of education, when he had little himself.

Contents

Tables and Figures

Glossary of terms and acronyms

Glossary

Absolute knowledge	Knowledge and skills attained at a particular time point irrespective of background of the student.
Artefacts	Assessment achievement elements stored in portfolios to demonstrate acquisition of learning outcomes
Assessment	Any group of methods used to appraise the knowledge, skills and attitudes of an individual or a group of students at the completion of a course of study.
Authentic assessment	Assessment tasks related to real problems the graduate is likely to encounter in the workplace after completion of their course of study.
Calibration	A process to ensure common understanding of standards between assessors of student work.
Competence	A satisfactory level of performance or achievement of a standard associated with a learning outcome.
Course	A collection of subjects which when completed satisfactorily constitutes a coherent program of study.
Evaluation	The drawing of conclusions and making judgments about performance of individuals or institutions based on the results of measurement and/or reporting.
Learning outcomes	A short description of knowledge, skill or a personal attribute resulting from a course of study containing a criterion for achievement which provides insight into how the outcome can be measured
Local achievement assessment	Internal classroom based assessment tasks conducted as part of the teaching and learning activity in a subject.
Measurement	Assignment of numerical scores or quantifying the classification of observations about students' performance in assessment tasks.
Moderation	Independent marking and grading of an assessment task to ensure standards are appropriately adhered to and there is consistency in results.

Real world problems	Tests or problems which are representative of issues which would be encountered by graduates in employment in their professional discipline
Reliability of assessment	Assessment for which outcomes are able to be determined consistently over time or in different contexts.
Reporting	Provision of information for stakeholders at the group or institution level on knowledge, skills and attitudes of students resulting from a course of study.
Subject	A unit of study with its own learning outcomes, which is a component of a course.
Signature assessments	Generalisable assessment tasks specifically designed for use in testing achievement of course or institutional learning outcomes by individuals or institutions which address the criteria associated with standards used to score such achievement.
Standard	An agreed specification or other criterion used as a rule, guideline or definition of a level of performance or achievement.
Value added	A measure of student achievement at the end of a course, as indicated by exit knowledge and skills, which takes into account differential inputs (ie knowledge and skills on entry).
Validity of assessment	Assessment which truly measures what it purports to measure – the stated learning outcome is actually being measured.

Acronyms

AACU	American Association of Colleges and Universities
AAGLO	Assuring Assessment of Graduate Learning Outcomes project
AHELO	Assessment of Higher Education Learning Outcomes project
ALTC	Australian Learning and Teaching Committee

AMAC	Australian Medical Assessment Collaboration
AQF	Australian Qualifications Framework
AUQA	Australian Universities Quality Agency
CAAP	College Assessment of Academic Proficiency
CEQ	Course Experience Questionnaire
CLA	Collegiate Learning Assessment
CQAHE	Committee for Quality Assurance in Higher Education
DQP	Degree Qualifications Profile
EFTSL	Equivalent Full-Time Student Load
ELO	Essential Learning Outcomes
ETS	Educational Testing Service
GAP	Australian national Graduate Attributes Project
GPA	Grade Point Average
GSA	Graduate Skills Assessment
HEC	Higher Education Council
HEQCO	Higher Education Quality Council of Ontario
LEAP	Liberal Education and America's Promise
MAPP	Measure of Academic Proficiency and Progress
NBEET	National Board of Employment, Education and Training
NILOA	National Institute for Learning Outcomes Assessment

OLT	Office of Learning and Teaching
PCFC	Polytechnics and Colleges Funding Committee
QAA	UK Quality Assurance Agency
QILT	Quality Indicators for Learning and Teaching
SPR	Student Progress Rate
TEQSA	Tertiary Education Quality and Standards Agency
UES	University Experience Survey
VALUE	Valid Assessment of Learning in Undergraduate Education
VSA	Voluntary System of Accountability, US

PART 1: DEFICIENCIES AND STRENGTHS OF CURRENT PERFORMANCE MEASURES FOR HIGHER EDUCATION TEACHING AND LEARNING

Chapter 1 – Context and the need to measure learning outcomes in higher education

1.1 Aim and conceptual framework of the research

The aim of this research is to develop a suite of performance indicators which will measure differences in universities' performance in attainment by their students of specified learning outcomes. This is an ambitious goal because the measurement of university performance is often a vexed issue for educational policy and practice even in less complex areas than teaching and learning. Secondly there is considerable debate about what learning outcomes are for higher education courses and then whether achievement of these outcomes can be measured in any valid way at an institutional or sector level.

But the task is a worthwhile and timely pursuit because answers about what constitutes high quality teaching and learning in higher education, while influenced by the quality of the student, resource inputs and the teaching process, are determined by what outputs and outcomes are achieved as a result of these influences. Using a production model for higher education, outputs are measures of what is produced using a set of inputs and are indicative of efficiency of the system (2015). Outcomes are what is actually achieved in terms of benefits to students, the government and society as a result of the higher education learning experience and are indicative of the effectiveness of the higher education system.

The objectives of modern higher education sectors in most developed countries, as documented in government legislation and policy statements, are generally related to one or more of the following:

- undertaking high quality teaching and research;
- supplying a labour market with highly skilled graduates and thereby improving their country's productivity;
- providing equality of opportunity; and
- encouraging lifelong learning.

This can broadly be summarised as being concerned with quality, equity, effectiveness and efficiency. In Australia these characteristics were enshrined in legislation in the

Higher Education Support Act (2005) and were re-affirmed by the Bradley Review of Higher Education in 2008 (Bradley, Scales, Noonan, & Nugent, 2008). The expenditure on higher education is large from a range of sources ($15.4bn in Australia in 2014) and Governments and stakeholders such as the general public, students and their parents, graduates and employers wish to ensure that this expenditure is well spent and deployed to achieving the best possible outcomes for students and for the community for the lowest possible cost.

The performance of universities in teaching and learning is currently measured using indicators which are mostly proxies for learning outcomes. The focus has been on the quality of the inputs or processes (course content, structures, the way the content is taught) with little attention being paid to the measurement of what the student actually learns (learning outcomes) at a practical level. Learning outcomes are easier to describe than to measure and yet there has been little agreement on what higher education learning outcomes are at a sector level. Nevertheless over the last 15 years in most systems of higher education, increasingly there are broad descriptors of expected outcomes from particular levels of awards in various qualifications frameworks, and universities have individually developed statements of graduate attributes, which can be regarded as aspirational learning outcomes. It appears that these graduate attributes, which are often specified for whole institutions, particular course levels, and individual courses have not generally been measured except for some professional areas as part of an accreditation process, and it is not clear how institutions demonstrate the attainment of these expected learning outcomes by their students. Most links that have been formed to these graduate outcomes have involved relating inputs, such as curriculum content or the type of teaching undertaken, to the attributes or the student achievement has been represented through surrogates such as student self-perception of achievement. The only valid way to measure what students have learned is through assessment that is closely aligned with the specified learning outcomes.

Conclusions about the effectiveness and quality of higher education are drawn by comparing achievements with planned objectives. A measurable set of learning outcomes and related assessment and reporting would therefore help demonstrate whether the objectives of quality and efficiency of the sector are achieved for all courses and disciplines. If this could be done, and attainment by the individual learner combined in some way to assess the performance of the institution where the study took place, it

would lead to more useful and accurate real teaching and learning outcome indicators for the sector. This would enable an accurate view of the quality of the higher education sector to be formed by stakeholders, demonstrate that the education dollar is well spent, and also open the possibility for development of new strategies in teaching and learning, including improvement in assessment.

1.2 Focus of the research

This study was undertaken to explore the feasibility of using assessment of student learning outcomes to measure institutional performance in teaching and learning and to attempt to improve the current range of indicators which have been in use in the Australian higher education sector in the last ten to fifteen years.

Gaining a comprehensive understanding of the history of performance indicator development and use internationally in teaching and learning led to a conclusion that the measurement of learning outcomes both at an individual level and for a course, discipline or university first requires an investigation of the nature of learning outcomes themselves and their applicability in general across the higher education sector. This then raises the question of how students are being or could be assessed to ascertain whether or not they have attained the learning outcomes specified by their university or more generally. Assessment is commonly done at the subject level and varies between courses, disciplines and institutions, raising the question of whether such diversity of approach could be harnessed to measure in a consistent way the attainment of an individual's course based learning outcomes. If it could, what pre-conditions might apply to ensure reliability of such a measurement. Finally there is the issue of whether an individual level of achievement could be used in some way to derive quantitative indicators which might be of use in evaluating the performance and quality of an institution or the sector in delivering learning outcomes.

Hence this research study explores the relationship between the specification of learning outcomes, the assessment and reporting used for individual student learning, and whether this can be combined and reported at an institutional level, and a potential set of performance indicators be defined which could measure the extent of student learning in undergraduate higher education. In developing statements of learning outcomes and exploring forms of assessment in use in universities the study focuses on the Australian higher education sector. It argues from a premise that, given international examples, it

should be possible to specify learning outcomes adequately for the sector and using a combination of assessment processes currently applied as part of local assessment regimes, measure the level of individual student achievement of these outcomes. The primary output from the study is a model for the development of institutional performance indicators which measure, at a university or sector level, student learning for a range of disciplines against a set of generalisable graduate learning outcomes using local assessment of the individual student. The strength of the performance indicators is judged in terms of their feasibility and reliability at measuring institutional or discipline effectiveness at enabling students to attain the generalisable learning outcomes. The end focus of the work is therefore on performance indicators and their applicability to the learning environment of universities. The study is therefore an attempt to develop a broad end to end model for learning outcomes measurement across disciplines and not a deep exploration of measurement or assessment within a particular course of field of study.

It is expected that the research will impact on the public and institutional agendas of teaching quality evaluation and standards and the political environment which governs higher education.

1.3 The challenges

This is a courageous study because while considerable work has been undertaken over the last twenty years on the specification of student learning outcomes at the course or institutional level, little progress has been made in measuring them across courses, disciplines or universities. Descriptions of generalised learning outcomes are sometimes embedded in various qualifications frameworks in different countries but the measurement of student achievement and institutional performance in relation to these learning outcomes remains an area of considerable unresolved interest.

There are several challenges in developing such a model for institutional measurement of learning outcomes. These challenges occur at every point in the sequence outlined above leading to the identification and application of the performance indicators. Some relate to pedagogical issues, some to the general imperfection of assessment in determining students' achievement, and others to the politics surrounding the use of any performance indicators developed.

1.3.1 The specification of learning outcomes

Apart from difficulties in getting agreement from the key stakeholders in teaching and learning in universities – the students, staff, employers of graduates and the Government which partially funds the enterprise - over which learning outcomes are appropriate and their relative importance, many of those that have been specified in the past may not be measurable or may not be influenced by teaching. For example, university graduate outcomes, which are described either at the university or course levels often include changes in students' personal characteristics or behaviour which may not be related to the teaching process or the education environment in which they study – that is, to things within institutional control.

There is also debate about whether graduate attributes, such as those defined by all Australian universities, are sufficient descriptors of learning outcomes. While it is shown that the statements which do exist for different universities are very similar, graduate learning outcomes are usually more specific than the broad characteristics included in statements of graduate attributes. The latter are considered to be aspirational and acknowledged to be stretch targets in terms of achievement for all students. Learning outcomes on the other hand, coupled with a set of associated standards statements, aim to reflect what all students should be able to achieve but to varying degrees.

In addition there are necessary differences in statements of learning outcomes between disciplines. For example professional disciplines are much more likely to include statements about competencies required for practices relevant only to the particular profession. A challenge for producing useful statements of generic learning outcomes is the bridging of the gap between sector-wide learning outcome specification and those of different disciplines. This is particularly the case between general disciplines such as science, humanities and social sciences and the career-oriented disciplines of medicine, engineering and education for example.

How to develop a generic set of learning outcomes which are assessable in a comparable way between disciplines and institutions and which are still meaningful nationally and internationally is a key question for this component of the research.

1.3.2 Challenges relating to assessment

Once learning outcomes have been defined, the next group of challenges relates to how attainment of them could be assessed. At present assessment approaches are often not well aligned to the statements of learning outcomes or to graduate attributes.

Currently in Australia few assessment approaches directly ascertain students' success at achieving *course* level learning outcomes although attempts are often made of this for various professional disciplines such as medicine, accounting and engineering as part of course accreditation.

There is a tension between the teachability of subjects and courses and their assessability. Greater emphasis is more frequently placed on the nature and objectives of the curriculum taught rather than the way the curriculum is assessed to evaluate the attainment of learning outcomes. Further, while learning outcomes are specified at the course level, assessment is commonly designed and carried out for subjects. Assessment for the latter more commonly involves one or two types of assessment within the subject, standardised tests and normalised outcomes of results. For courses there are many different assessment types needed depending on the range of subjects taught; and there are multiple assessors. A complex, holistic perspective on student performance must be gained and a composite suite of measures of achievement is necessary. Research on assessment used in higher education shows that frequently academics are conservative about the forms of assessment they use and there may be a need to overcome resistance from this community to be more innovative about assessment which accurately measures student learning, if the objective of this study is to be realised.

This variety in the approaches to assessment and grading and the need for a composite suite of assessment tasks which would be applicable to multiple universities are barriers to the construction of generalised forms of assessment of learning outcomes. In this study relevance to local contexts would need to be recognised while ensuring that the description of the form of assessment and its operationalisation are sufficiently applicable across disciplines and institutions to reap benefits to the sector in terms of consistency of results against common learning outcomes.

The type of assessments which may be needed to evaluate attainment of learning outcomes are likely to go beyond testing of core knowledge and skills and on to

exploring students' capacity to reason and apply skills across different disciplines in different ways (i.e. generic skills). And if assessment is to be employed in constructing quantitative institutional performance indicators, complex cognitive responses must also be measured. All types of assessment – formative and summative –should be used to measure progress of a student towards the attainment of learning outcomes. A classification of competency and achievement is also required which is able to differentiate sufficiently so that discipline or institutional performance at achieving a suite of learning outcomes can be evaluated. Hence specification of appropriate standards and a form of grading of achievement of them is a pre-condition to the development of quantitative institutional performance indicators in this study.

At a practical level, a significant challenge is the ability to aggregate or combine individual assessment data to get a meaningful outcome across all students in different courses. Measurement theory (Kleinberg, 2003) suggests that there are limitations to getting reliable measurement at course or institutional level from subject level information and so identification of the means of assessing, at the course level, each learning outcome will need to be explored. This will involve addressing the feasibility of defining new assessment instruments for some of the concepts in learning outcome statements.

1.3.3 Performance indicator development

While a suite of performance indicators might be expected to include both qualitative and quantitative measures, in order to move towards a more viable approach towards measuring collective learning outcomes the derivation of the quantitative indicators must as far as possible avoid the current use of proxies, which do not describe what students know or can do. For the indicators to be reliable they should also be framed so that their values cannot be easily manipulated by institutions and the system of measurement 'gamed' to present more positive outcomes than are actually achieved.

There would need to be multiple measures across the whole set of learning outcomes to reflect the complexity of representing overall performance, but the tendency to define indicators which are based on factors that can be measured rather than what needs to be measured should be avoided. Both of these aspects present considerable challenges for the research study. If the indicators are to be reliable in differentiating performance between universities, disciplines or sectors in different countries, the research will need to determine a classification scheme and benchmarks which can be used for rating

different standards of performance. The indicator set should acknowledge the diversity of the higher education environment but the study should also identify ways to control for quality of students on entry, resources, differences in approaches and other factors that might affect measurement results.

The indicators would need to be able to be operationalised to be of practical use to the higher education sector and its stakeholders, consistent with the aims of this study.

1.4 Approach to the research

These issues are explored by first reviewing the history of performance measurement in higher education, the use of such indicators in ascertaining the quality of teaching and learning, and attempts made to measure learning outcomes.

The model proposed will attempt to bring together four strands of investigation in a coherent way:

- The specification of learning outcomes;
- The definition and use of a set of standards for assessment;
- Assessment of individual student learning; and
- Institutional performance measurement in terms of groups of students' achievement of the learning outcomes; and

Together these elements should describe a conceptual framework for the measurement of learning outcomes at institutional, discipline and sector levels.

1.4.1 The scope of the study

The study addresses the following five research questions which are related to the above steps of the model.

- What are the deficiencies and strengths of current performance measures for higher education teaching and learning, and what are the characteristics of a robust set of measures for learning outcomes?
- In terms of measurability, what does a minimal set of generic course-based learning outcomes, which are representative of the range of graduate attributes specified by Australian universities, look like and how should they be classified?
- What types of assessment approaches validly measure individual students' achievement of some or all of these learning outcomes at the course level which can be used to inform performance indicator development?

- How can these assessments of individual learning outcomes be combined and used to produce a measure of an institution's or discipline's capacity to deliver the specified learning outcomes for its students?
- Which performance indicators constructed in this way comprise a representative suite derived from these aggregated individual absolute assessment outcomes, directly measure institutional achievement of learning outcomes, and allow stakeholders to differentiate between the performance of different disciplines or universities?

What is discovered from these investigations should be able to inform the constructs of a model for the measurement of higher education student learning outcomes at an institutional and sector level.

1.4.2 Methodology and organisation of the study

The study uses a mixed-mode methodology of comparative review and analysis of past practices in performance measurement of teaching and learning, formulation of a set of potential generalised learning outcomes applicable to the Australian higher education environment, capture of new assessment information using a data collection proforma applied to Australian universities, and the construction of possible performance indicators based on proposed assessment regimes.

The study is organised into four parts reflecting the above research questions.

Part 1 – Deficiencies and strengths of current performance measures for higher education teaching and learning

The first part examines the history of the development and use of higher education performance indicators in a range of countries including Europe, the United States and South East Asia, and tracks the use of such indicators to measure the quality of teaching and learning in higher education. This enables conclusions to be drawn generally about the characteristics of good performance indicators, the political and practical influences in the construction and use of such indicators, and the current use of institutional level indicators for the measurement of teaching and learning outcomes.

Part 2 – On defining and classifying a minimal set of generalisable course-based learning outcomes

The second part of the study analyses the learning outcomes specified at sector and degree level in a range of countries as descriptors of what students are expected to learn and be able to do on graduation from undergraduate study. It compares similarities and differences of these approaches, and the nomenclature and the taxonomies used to classify the outcomes, and identifies strengths and weaknesses from the perspective of what could be used to gain a shared understanding across jurisdictions of what students are expected to learn from their higher education experience. This international comparison highlights differences in approaches taken, with the Australian work on describing the outcomes tending to be within individual universities with less emphasis than in Europe and the United States on course- or sector-wide outcomes. A detailed analysis of the graduate attributes and learning outcomes specified by all Australian universities is undertaken and a conceptual framework built for the specification of a generalised set of learning outcomes in Australian higher education. This involves a review of work done as part of both the Graduate Attributes and the Standards Projects conducted under the aegis of the Australian Teaching and Learning Council, and represents the first step in applying the model for measuring generalisable learning outcomes in this country using a newly developed set of standards which might be able to be applied across all undergraduate courses.

Part 3 – Use of assessment approaches to validly measure individual students' achievement of learning outcomes by graduation

This step addresses the question of whether it is possible to assess individual students reliably to gauge achievement of a range of learning outcomes similar to those identified in Part 2. The study reviews the characteristics of assessment methods currently in use to test undergraduate student learning in the higher education sector and explores issues relating to the alignment of these assessment practices with the specification of learning outcomes for a range of disciplines. Apart from an analysis of the literature on this topic and a high level examination of the relationship between assessment and learning outcomes as documented on Australian university websites, qualitative research is undertaken to obtain views of key staff in all Australian

universities on the appropriateness of the set of generalised learning outcomes developed in Part 2 and how these are currently assessed within universities. This is achieved through a survey and follow up with some senior academic staff responsible for teaching and learning policy at the institutions on how individual student assessment is carried out for their own learning outcomes and how the institution believes it is able to demonstrate the achievement of them. It is expected that the output from this part of the study would contribute to the alignment of assessment practices which could be used to test a subset of the generalised set of cognitive and non-cognitive learning outcomes derived in the specified set. Any caveats or conditions on combination of such assessment outcomes to allow measurement of performance at the institutional or sector level should also be identified. From the analysis of assessment tasks used internationally, types of assessment regimes are identified which have potential to measure achievement of the specified learning outcomes in a reliable and valid way. The study develops some options and explores their strengths and weaknesses in terms of their use in institutional performance measurement.

Part 4 – Indicators with potential to measure institutional performance derived from these aggregated individual assessment outcomes

From this information, an approach to using results of individual student assessment to provide an indication of the overall success of the institution in enabling the achievement of learning outcomes is explored. The proposed set of indicators is examined against a set of good practice criteria for performance measurement developed in Part 1 of the study.

The proposed model and the suggested performance indicators have not been trialled due to lack of availability of assessment outcomes in the form proposed in the study in institutional data bases. However the apparent validity of the performance measures is examined in terms of expected outputs from this stage of the model is discussed, and issues impacting on the quality and comprehensives of the indicator set are explored.

Chapter 2 - History of performance indicator development and its relevance to measuring learning outcomes

This chapter sets the scene for the research study on trying to use assessment of student learning in a viable way to measure university performance. It shows how the performance measurement agenda in higher education has changed over the last twenty years from resource considerations to the desire to demonstrate the quality of the teaching and learning, and how identification and the measurement of learning outcomes is central to the ongoing development of this agenda. The study commences by reviewing the history of the use of performance indicators internationally and then describes the Australian context in more detail as a lead-in to the later specification of learning outcomes for undergraduate study in that environment.

2.1 The policy context of performance indicator development and use

The emergence of higher education performance indicators was stimulated by the transition from an elite to a mass system of education. Enrolment growth accelerated internationally from the 1970s, flattened during the late 1970s in the industrialised nations, and then gained further momentum in the late 1980s. Martin Trow (Trow, 1974, 1976) classified the higher education system into three education states - elite; mass; and universal - based on the proportion of the appropriate age group in the population participating. Trow's categorisation was based on the following participation boundaries:

Elite	Less than 15% of the relevant age cohort;
Mass	Between 15% and 35% of the cohort; and
Universal	More than 35% participation of the population cohort.

These boundaries were later challenged but the concept of these participation-based categories is still used to classify higher education sectors today.

The two-tiered US system of higher education had always been market driven and was ahead of European and United Kingdom universities in the transition to a mass system. In fact by the 1980s the US system was, on Trow's categories, already moving towards universal participation. The shift to mass higher education as defined by Trow escalated in the United Kingdom in the early 1980s, and in Australia following the White Paper

on higher education (Dawkins, 1988). The Dawkins reforms led to ambitious planning and achievement of an accelerated enrolment growth pattern compared to the steady increases experienced between the early and late 1980s. These growth trajectories were a universal pattern in OECD countries and their desirability from Government perspectives was based on a belief that universities could supply an educated workforce and assist in delivering improved national productivity and economic circumstances as a result of the change in profile. This massification of higher education was generally accompanied by increased revenue for higher education institutions provided by students, government and industry in varying proportions and hence a heightened need for accountability. In Australia the expansion in funding was primarily made possible by requiring students to contribute towards the cost of their education for the first time since 1974, through an income contingent loan scheme (Wran, 1988).

The increase in scale of institutional operations also led to changes in the way universities and colleges were managed and in the perceptions of institutions as part of a sector in which government and other stakeholders, through increased funding, had a strong interest. Trow claimed that there were significant differences between elite and mass higher education institutions partly necessitated by the size of their student populations. He stated that:

> *The essential characteristics of universities which offered elite higher education before the Second World War were their relatively small size and their function in the selection and preparation of the political elites and the elites of the learned professions. By contrast, forms of mass higher education were created to prepare young persons for careers in a great variety of new or expanding technical, semi-professional and managerial occupations.* (Trow, 1976)(p51).

The mass higher education sector needed to appeal to a larger audience and public perceptions about accessibility and student outcomes on completion of study programs were considerably changed with an increased focus on societal and individual objectives. Increasingly in the public debate there was mention of the concept of lifelong learning as one of the outcomes of a higher education. For example, Cerych *et al* (Cerych, Furth, & Papadopoulos, 1974; Department of Employment Education and Training & Organisation for Economic and Cultural Development, 1993) noted that in addition to their traditional mission of transmission and extension of knowledge, universities were now required:

- *to play an important role in the general social objective of achieving greater equality of opportunity;*
- *to provide education adapted to a great diversity of individual qualifications, motivations, expectations and career aspirations;*
- *to facilitate the process of lifelong learning;*
- *to assume a 'public service function' ie to make a contribution to the solution of major problems faced by the community surrounding the higher education institution and by society at large; and*
- *to participate directly in the process of social change.*

The different perception of institutional role was also accompanied by a change in government and institutional relationships with greater responsibility placed on institutions for planning, quality and the efficiency of their operations, and a required change in focus to be more concerned with the relationship between higher education and employment. There was therefore an increased emphasis on the student outcomes of higher education learning, even though it was mainly concerned only with employability of graduates. It became important that graduates had particular skills and knowledge which could be applied in the pursuit of a broader range of professions and the concept of 'generic skills' or 'core competences' emerged.

However, while there was a significant increase in government subsidy of higher education during the 1980s in Europe, Australia and Canada, in most of these jurisdictions the increase in funds did not keep pace with the growth in student enrolments. As well, in the UK (Bourke, 1986; Cave, Hanney, & Kogan, 1991) there were pressures through cost cutting imposed by Government, which led to a lowering of the average per capita contributions to higher education institutions. The shift in the policy debate therefore also moved to a greater focus on the efficiency and effectiveness of higher education institutions.

Teichler (1988) confirms this and states that there were several issues which gained prominence in higher education policy in the late 1980s, including:

- the emphasis on a need for quality improvement;
- increased support for diversity in higher education;
- increasing attention to more efficient management of higher education institutions;

- growth in the use of performance indicators to measure quality and efficiency in the allocation of resources;
- greater interest in the utility of higher education to the private sector; and
- the internationalisation of the higher education agenda.

The change in the policy focus in Australian higher education in the early 1990s reflected these issues following the restructuring of the sector in the wake of the Dawkins White Paper, particularly in relation to the use of performance indicators and the need to demonstrate quality improvement.

Using a production model (van Dooren et al., 2015), performance indicators were first applied to higher education in the mid-1980s largely as a result of a need for higher education institutions to demonstrate efficient and effective use of funds provided by governments or others. In both the United Kingdom and in Australia sets of performance measures were identified by peak bodies and the institutions themselves and were intended to be used in formulae for performance based funding, thus forging the link between accountability and the financing of the sector. Most of the indicators were resource-related such as unit costs, student/staff ratios, class sizes, staff workloads and access to library and information technology resources, and were not directly related to the quality of university operations.

2.2 The implementation of performance indicators in higher education

Bourke and Cave *et al* (Bourke, 1986; Cave et al., 1991) outline the experience in developing and using performance indicators in relation to funding of higher education in the United Kingdom, the United States and Australia. While this focus on funding issues was the genesis of the articulation of performance indicators in higher education they rapidly became the means of governments and the institutions to inform the student and employer market through use of key indicators in ranking systems and in reviews of the sector, and to drive efficiencies in institutions' operations.

A concern voiced by the sector in Australia was whether the indicators proposed actually measured performance in the academic areas of research and teaching and learning. A report from the OECD published at that time (*Performance Indicators in higher education : A study of their development and use in 15 OECD countries)* (Organisation for Economic Cooperation and Development, 1988) stated that

performance indicators were only an adjunct to inform expert judgment on the state of higher education and rejected proposals that the indicators be used in funding formulae. In spite of this development, work on performance indicators proceeded in all developed countries to some degree, in some cases as the basis of determining funding arrangements.

The United Kingdom

Cave *et al* (1991) provide a detailed chronology of the activities undertaken by various government bodies and committees over the period from 1984 to 1990 concerning performance indicators. The first systematic use of performance measurement in the UK was by the University Grants Committee (UGC) in 1981 when it was asked to respond to a directive of the Thatcher Government that funds were to be cut across the sector by 14-15% but that the cuts would not be uniform (Bourke, 1986)(p4). To address this problem, it was necessary for the UGC to develop a funding rationale based on performance and quality of teaching to ascertain where to recommend cuts in funding in the sector. The indicators used for this purpose were mostly qualitative but also included some quantitative information on graduate employment, staffing resources, student demand and efficiencies in the use of resources. The use of qualitative information and a review of the data was consistent with the view that performance indicators were only a guide rather than a determinant of funding allocations. In effect the reductions made were not based on any public criteria and were the result of a form of peer review.

In 1984, the Jarrett Committee was established in the UK by the Committee of Vice Chancellors and Principals (CVCP) to report to the University Grants Committee (UGC) on the efficiency of UK universities. In this investigation, according to Bourke, the Committee commented on the lack of an objective and comprehensive set of performance indicators on which to base judgments about funding even though at that time quantitative parameters were used in the allocation of funds between and within universities. As part of this process a range of performance indicators in common use in individual universities were identified by the Jarrett Committee and were grouped into *internal; external; and operating* indicators (see also (Cave et al., 1991)). There was no attempt to develop indicators which were involved in the measurement of academic teaching performance but the list did include some broad indicators of learning outputs and outcomes such as graduation rates, success rates and employment outcomes.

These events in the UK and the early use of performance indicators are also discussed at length by Goedegebuure and by UK contributors to the book he and colleagues edited (Goedegebuure, Maassen, & Westerheidjen, 1990).

The Jarrett Committee was followed by a Government Department of Education and Science White Paper in 1987 (*Higher Education -Meeting the Challenge*) in which a target participation rate of 18-24 year olds was set and the introduction of performance indicators into higher education policy was recommended. These were modelled on the input-process-output (production) model of higher education operations, were focussed mainly on efficiency and effectiveness, and did not address the issue of quality of outputs or outcomes (Goedegebuure et al., 1990)

Subsequently in the UK in 1989 the Polytechnics and Colleges Funding Committee (PCFC), which was established through the split of the original University Grants Committee into two funding bodies (the Universities Funding Committee and the PCFC), formed a performance indicators committee to recommend a range of measures, with an emphasis on measuring efficiency and effectiveness of institutions. Specifically the Committee was to:

- review existing practice and knowledge relating to institutional performance as part of its funding mechanism;
- suggest those aspects of institutional performance which ought to be evaluated in terms of their contribution to efficiency and effectiveness of each institution;
- advise on how performance indicators might be used by the PCFC institutions in their policy and planning; and
- suggest a range of Performance Indicators.

(Council for National Academic Awards, 1990)

The list of performance indicators for ‘education’ developed in the UK in response to this initiative were focussed on graduate outputs and outcomes and included per capita support service expenditure:

- Careers services expenditure per FTE student
- Student unions expenditure per FTE student;
- Destinations of degree graduates after 6 months;

- Total graduates with known destinations;
- Destinations of unemployed or short-term employment;
- Expected numbers of unemployed students or those in short-term employment;
- Difference between expected and actual unemployed graduates;
- Number of successful completions (leavers);
- % leaving their studies (attrition);
- Terms of attendance per success (how long it took students to complete);
- Years taken to complete as a % of course length; and
- A variety of UK specific entry qualifications details.

(Goedegebuure et al., 1990)(p 162) and (Cave et al., 1991)(pp50-51)

These indicators appeared to be developed independently of any 'education' goals of institutions or the sector and were more measures of productivity rather than quality of learning achieved. They were therefore management statistics rather than measures of teaching performance.

The United States

In the United States (US) at this time there was considerable use of quantitative measures and analyses at intra-institutional level in higher education but no systemic use of performance indicators. The higher education sector in the US, being a diverse mixture of public and privately funded institutions, focused on ranked lists of institutions based on perceived reputation, as most of the formal international ranking systems do now. These rankings were derived from peer judgments and research publications. In the early 1980s, according to Bourke, effort was put into taking into account outcomes of undergraduate and postgraduate education in the institutional rankings. Bourke refers to work done by Astin in 1982 to 'construct ways to discover variations between institutions in terms of the actual benefits which higher education academic experience confers on its recipients', in effect attempting to measure learning outcomes. It is sobering to think that this task framed by Astin in 1982 is still not fully addressed even though in higher education the use of performance indicators is widespread internationally and progress has been made on identifying the nature of the student experience.

In the market driven sector in the US, indicators of available opportunities and quality of degrees were also needed to inform prospective students of their likelihood of entry to and the desirability of particular institutions. It is asserted by Cave *et al* that such indicators were not performance indicators in the sense being developed in the UK or Europe because they were 'used to inform peer review and assist internal management and formative self-assessment rather than related to government goals and published in league tables that might influence funding decisions' (Cave et al., 1991)(p58). In the context of this study, this focus on funding is a narrow interpretation of what performance indicators might be used for and in fact there has been relatively little use of performance indicators in funding allocations to institutions.

Most of the indicators developed in the US were related to research but scores of students on standardised tests such as the Scholastic Aptitude Test and the American College Test and students' achievements in later life were discussed as possible teaching performance indicators (Bourke, 1986). In the 1980s the US focus shifted to finding ways to measure the 'outcomes of undergraduate education' (Cave et al., 1991)(p61) and seeking the opinions of various stakeholder groups was central to this change. This approach of assessing satisfaction of students and gaining opinions of staff about the quality of what was produced was quite different to the efficiency/ funding base behind the UK and European indicators of the time.

This focus on outcomes led to an attempt to quantitatively assess the 'value-added' or the benefits that the institution confers on its students. Bourke points out that measurement of value-added involves pre-and post-testing of students on a number of scales related to generic skills and personality development (Bourke, 1986) but the use of particular tests is only one approach to evaluating whether students have attained desired learning outcomes. Value-added measurement through the use of standardised tests remains a primary component of evaluation of learning outcomes for individual students in the US as discussed in Part 3 of this dissertation.

Other European countries

In the 1980s the development of higher education performance indicators varied greatly in different countries (Cave et al., 1991) (pp65-66). Decisions of whether to proceed with such a system of performance evaluation are influenced not only by technical and

measurement concerns but also the political influences on the higher education system in place in those countries at the time.

The situation in the Netherlands is discussed at length by Goedegebuure, with an outcome that peer review of educational and research activity was preferred to the implementation of a general set of performance measures in the 1990s although some bibliographic measures were developed to evaluate research outcomes and to assist in funding allocations. The Netherlands moved earlier towards a system of quality assurance as a means of ensuring high standards in the higher education system. The situation in Sweden and Finland was similar. Further comparisons between this approach and the UK and Australian performance indicator driven systems is made on the section on quality assurance.

Australia

As in the UK, the Australian Government became interested in formally measuring the performance of universities in research and teaching and learning in the 1980s, and this interest pervaded the sector reforms which occurred in 1988. There had always been a greater interest in research performance with much of the funding based on competitive bidding through the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC). On the teaching side, individual universities had monitored graduations, retention and subject pass rates but the results of this analysis were not available outside of their jurisdictions. There were no comparable national performance measures for teaching and learning and funding was based on government-agreed planned levels of enrolment of equivalent full-time students.

The Green and White Papers (Dawkins, 1987, 1988) heralded a desire by the Government to move to a mass system of higher education (as defined by Trow) and to develop performance indicators to measure the achievement of core goals of universities and possibly to provide funding based on performance. The Green Paper (p42) stated that:

> *The Government supports the development of a funding system that responds to institutional performance and the achievement of agreed goals…. While it may be difficult at present to categorise and compare the outputs of the diverse activities of higher education institutions and their staff, the Government considers that an*

effective resource allocation system must have close regard to performance. At the simplest level, it is possible to measure system outcomes and efficiencies through student and staff data collections and unit cost comparisons.

This position became policy in 1988 in the White Paper. It was reported that a partial range of performance indicators had been compiled as part of an earlier national Review of the Discipline of Engineering and measures such as 'student satisfaction and completion rates, relative staffing levels and research publication and consultancy rates' (pp85-86) were mentioned. The Government also stated that it would support a joint approach to developing a set of performance measures by the then two sectors of higher education which would be used nationally to assess institutional performance. A Research Group on Performance Indicators was established in 1989 to develop a more systematic approach to performance evaluation across the higher education sector, to identify a set of indicators, and to evaluate their usefulness in differentiating this performance. This is very similar to the remit given to the PCFC performance indicator committee in the UK at about the same time.

The Research Group reported on the outcomes of its deliberations in 1991 (Linke et al, 1991). A set of indicators had been developed and tested and they were grouped into three broad categories:

- Institutional context indicators;
- Institutional performance indicators; and
- Social equity indicators.

These categories were quite different in intent to the classification used in the UK as the context indicators defined in Australia contained many which were included as internal and external indicators in the UK system. Linke's performance indicators did try to measure teaching performance of the institution and the sector by using the US approach of seeking the opinion of students and staff about teaching performance. Linke's definition of what constituted a performance indicator was that it should 'measure the relative achievement of institutions and their constituent organisational units *against their respective aims'.* (Linke et al, Vol 1, p17). He argued that performance should be judged in terms of the objectives set by the institution and the context indicators which would identify the characteristics of the institution. This

approach was more nuanced than that reported for the UK and resembles the later approach to the use of performance indicators in quality evaluation.

Many of the measures which are still described as performance indicators for teaching and learning (eg student/staff ratios; average student entry score; preference to application ratios etc) were classified by Linke as context indicators and the only true performance measures identified were:

- Perceived teaching quality through the Course Experience Questionnaire;
- Student progress rate;
- Program completion rate;
- Mean completion time;
- Research higher degree productivity rate; and
- Graduate Employment Status.

Four of these indicators are related to efficiency and productivity of the teaching process. The exceptions were perceived teaching quality and graduate employment status. The most significant immediate outcome of the Linke Research Group was the use of the Course Experience Questionnaire (CEQ) as a key measure of performance of universities in teaching and learning. The CEQ was developed by Paul Ramsden under commission to the Research Group and was trialled for the Research Group. It was based on work undertaken at Lancaster University by Ramsden and Entwhistle ((Entwhistle & Ramsden, 1983) on student learning in higher education. The use of student perceptions to measure teaching performance assumed that there were 'aggregate level associations between the quality of student learning and students' perceptions of teaching' (Linke, 1991, Vol 2, p1). The original survey scales used in the CEQ included good teaching; appropriateness of student assessment; clear goals and standards; emphasis on student independence; and appropriate workload. Ramsden states (Linke 1991, Vol 2, p2) that 'when academic departments were perceived to provide these characteristics, students were more likely to learn effectively from courses run within them'. These two findings represented an advance in measuring the quality of teaching in Australia through the use of student perception as a measurement tool. But the indicators based on the CEQ are surrogates for student learning outcomes – they ask the students how satisfied they are with their learning and using Ramsden's findings

allow conclusions to be drawn about the quality of their learning. They provide no measures of what students have actually learnt.

The Government subsequently supported the development and promulgation of the CEQ nationally as part of an annual graduate careers survey, which also captured information about employment outcomes of graduates. A similar questionnaire is in use in the United Kingdom to assess quality of teaching (the National Student Survey) and provides some broad comparative international data but there the survey is administered to all students enrolled rather than graduates only as in Australia.

The performance indicators developed and trialled by the Linke Research Group for Australian higher education have never been used to allocate core teaching resources between universities in spite of that being the main aim of the Government in 1988. However, graduate employment outcomes and CEQ ratings were published annually from 2001 onwards for the sector by the Department of Education, Employment and Youth Affairs. This was the first systemic use of the Linke indicators of teaching performance. A more comprehensive application of a set of measures of teaching and learning student outcomes for the sector was undertaken in 2006 as part of the national teaching quality agenda, which is discussed further in the next section.

2.3 The shift from measurement of resource usage to measuring educational quality

Application of quality assurance techniques and the need for continuous improvement in higher education gained prominence in most countries during the 1980s and early 1990s, and there was a change in focus to using performance indicators to measure educational quality. Concerns arose that because the per capita funding rates were decreasing and student/staff ratios increasing in most higher education systems, there might be an issue about the quality of the teaching delivered to students. If the quality of higher education declined and stakeholders became dissatisfied with their experience, they would be less likely to participate in the sector. Without quality, higher education sectors would not assist economic recovery, generate growth and improve productivity, which were the goals of most countries in the shift from elite to mass higher education.

There was therefore a need to introduce quality assurance practices into the sector. Quality is a relative concept and can be defined operationally only in relation to achievement of a set of goals. Performance indicators contribute to quality assurance by

quantifying outputs and outcomes and allowing alignment with objectives so that future improvements can be identified as resulting from the quality processes. And so to demonstrate the existence of high quality and the achievement of sector goals, performance indicators which measured progress towards those goals were needed in all core areas of universities' operations.

2.3.1 Quality systems in other countries

During the 1990s quantitative indicators were developed and used in many countries to measure the performance of higher education systems in providing high quality teaching and learning. Systems such as the Dutch, US and Canadian sectors have used other mechanisms such as peer review by groups or a combination of peer opinion and rankings based on perceived quality and prestige. In the Australian and UK systems there has been a combination of peer review in the assessment of quality guided in some cases by sets of performance indicators and self-reviews which set the context for the peer review. In the UK the national Academic Audit Unit assessed higher education institutions in relation to quality assurance processes in use but also made comparative quality judgments between the institutions. Goedegebuure (Goedegebuure et al., 1990) discusses quality assessment in British and Dutch higher education in detail, comparing the approaches of using performance indicators and peer review.

2.3.2 Quality in the Australian higher education sector and the emergence of graduate attributes

The impetus in Australia in relation to the quality was a policy paper released in 1991 by the Commonwealth Minister for Higher Education and Employment Services (Baldwin, 1991). This policy statement took stock of the changes that had occurred since the implementation of the Dawkins White Paper of 1988, including the work done by Linke et al on performance measures, and mapped out the path that Australia would follow to assess the quality of its higher education sector. The Minister stressed the importance of providing a 'degree of quality assurance at both the institutional level and for the higher education system as a whole' and noted that 'this is particularly important now that a majority of OECD nations have or are establishing such arrangements and the Australian higher education system is becoming increasingly internationalised' (p3). The National Board of Employment, Education and Training (NBEET), through its Higher Education Council (HEC), was asked to provide advice on 'the characteristics of

quality and its diversity in higher education, the strategies that may be developed by Government and the higher education system to encourage, maintain and improve the quality of higher education; and …the means by which changes in quality over time may be monitored and evaluated' (p59). The Minister claimed that the approach to quality being pursued was innovative and appropriate to Australia and that it balanced institutional autonomy with public accountability, unlike some of the systems in use overseas (p34). Given the comments above about quality systems in some other countries being a combination of peer review and performance measurement, this was rather a bold statement, particularly as advice had not at that time been received from the HEC. In spite of this, Baldwin stated that 'the Government will support the further development of quantitative performance indicators, while being aware of their limitations' (p4).

The HEC and NBEET subsequently released a discussion paper which sought advice from stakeholders on how quality should be defined. The HEC finally settled on the notion of 'fitness for purpose' and defined quality in terms of 'describing the attributes that graduates should acquire when exposed to quality education' (Higher Education Council, 1992b)(1). The Council proposed that 'the attributes acquired by graduates provide the ultimate test of the quality of the system to which they have been exposed' and that 'the major criterion to be applied to the judgment of the quality of the individual elements of learning programs should be linked to the contribution that it makes to the staged development of students' (p7). They then went on to describe the graduate attributes they were considering:

- Generic skills (skills which are independent of discipline studied);
- A body of knowledge (both knowledge of the particular discipline and its theoretical base at a level of detail appropriate to the level of study); and
- Professional/technical or job related skills (applied immediately to their employment).

This is a critical moment in the development of the concept of summative learning outcomes of graduates in Australia and articulation of the need to be able to measure student progress towards these outcomes as a key component of quality assessment.

Following a framework similar to that discussed by Cave *et al*, the final advice provided to the Minister by the Higher Education Council and NBEET (Higher Education Council, 1992a)(2) included a specification of the goals of the higher education system in Australia at that time, thus allowing quality of the sector to be evaluated. There were fifteen goals listed but three were directly relevant to the concept of quality and good learning outcomes. These were that the sector aimed:

> *To serve the community by :*
>
> - *Ensuring high quality programs of teaching and supervision;*
> - *Ensuring that graduates of Australian universities are enabled to operate anywhere in the Australia or overseas at standards consistent with best practice; and*
> - *Ensuring that students are encouraged to achieve beyond their own expectations*. (p12).

These goals represented the broad objectives for teaching and learning against which the quality or fitness for purpose of the sector was to be measured. In its final advice to the Minister the Council returned to the concept of student learning outcomes as a key component of the quality of the higher education sector and its institutions. It stated that

> *'one important focus on the 'characteristics of quality' should therefore be on outcomes of the higher education system: research, involvement with the community and critically, the aspirations we have for our graduates; a description of the attributes that graduates should acquire if exposed to a high quality higher education system – including all its processes. This does not imply that all the described attributes are quantifiable, and certainly does not mean that they must be related to immediate employment but it does suggest that if the objectives are known and explicit all the stakeholders – students, staff employers the community more generally – are in a better position to judge whether the processes are suitable – what the universities seek to do is achievable'.* (p19)

This statement makes it clear that the Council believed graduate attributes were able to be defined for universities and for the sector and that at least some of these outcomes should be able to be measured in a quantifiable way. It also hints that some attributes might be fairly abstract but even by just stating what the aspirational outcomes were and working to a mutual understanding of their nature amongst stakeholders of the sector

the Council believed that this would go part way to demonstrating achievement of quality.

The HEC reported that the views of the relevant stakeholders converged in the description of graduate attributes but that they diverged when it came to evaluating whether the graduates actually had the agreed characteristics. This is the crux of the limited progress made to date in measuring course level learning outcomes in any sort of generalised way. The HEC then went on to define graduate attributes for the higher education sector in the same three categories as in the Discussion Paper but spelt out what might be included in each group of attributes and how and at what levels they might be dealt with in the curriculum, rather than addressing how the learning might be assessed and the outcomes measured.

In relation to its term of reference on measuring changes in quality as defined over time (quality improvement) and the evaluation of quality, the Council concluded that the problem of measurement was a major issue in the assessment of quality, regardless of national context or the particular approach taken. Again they reflected that desired learning outcomes for graduates were not all easy to measure and identifying whether students attained those attributes such as generic skills and abilities was more 'a case of inference than observation' (p72). There was also discussion about the feasibility of using performance indicators generally to denote quality. The use of performance indicators to monitor changes in quality over time at the institutional level was acknowledged but the point was made that most indicators merely reflect some measures which might be correlated with quality rather than the achievement of quality itself. Hence the problem of measurement of learning outcomes as a key element of demonstrating the quality of the higher education system remained unresolved.

However, as a result of these recommendations of the HEC on the structure of a quality assurance process for Australian higher education, in 1993 the Commonwealth Government established the Committee for Quality Assurance in Higher Education (CQAHE). This Committee conducted three rounds of quality review for each university in the sector over the period 1993-1995 and the second of these was focused on teaching and learning. While metrics such as graduation and progress rates (as defined by Linke *et al* in 1991) were used in individual submissions and attempts were made by some to examine the extent to which universities were adding value to

students' base level knowledge and skills, there was little evidence of innovation in measuring achievement of graduate attributes or students' progress towards their attainment.

2.3.3 Attempts at measuring learning outcomes

In March 2000 the Australian Universities Quality Agency (AUQA) replaced the CQAHE with a responsibility for auditing universities and State Accrediting Bodies on the quality of their higher education offerings. This was part of an overall quality framework which consisted of a focus on National Protocols for universities and the funded development and piloting by the Australian Council for Educational Research (ACER) of a Graduate Skills Assessment (GSA) test as a way of measuring students' achievement of graduate attributes. The test was similar to those used in the US to assess generic skills obtained by students in the course of their higher education study. It was proposed that the GSA be applied on students' commencement and completion to assist universities in assessing students for graduate entry and measuring their own performance in adding value. The Australian government published some data on the pilot study in its triennial funding report for 2001-2003 (Department of Education Training and Youth Affairs, 2001) which showed variation in results by discipline (p38). In the event, the take-up of the GSA in the Australian higher education sector was low and so the publication of data similar to that of the trial results was not continued.

In 2006 following a detailed analysis by Access Economics for the Department of Education, Science and Training on the statistical reliability of what were described as teaching and learning output and *outcome* performance indicators (Access Economics Pty Ltd, 2005), the Government established the Learning and Teaching Performance Fund (LTPF) and commenced the first comprehensive use of a set of measures of teaching and learning student outcomes for the sector. The development of the process and the set of indicators was iterative and a new set of eight performance measures were used in 2007 (Department of Education Science and Training, 2007) clustered into three groups:

Student satisfaction

- Overall graduate satisfaction (the proportion of graduates who were 'broadly satisfied' with the overall quality of their courses);

- Graduate good teaching rating (the proportion of graduates who were 'broadly satisfied' with the quality of the teaching they have received); and
- Graduate generic skills (the level of 'broad satisfaction' with the generic skills acquired).

Attrition and retention

- Student progress rates (the proportion of subject load passed by a student);
- Student attrition rates (the proportion of students in any year who neither graduate nor continue studying in an award course at the same institution in the following year).

Graduate outcomes

- Graduate full-time employment (the number of Australian graduates in full-time employment as a proportion of all Australian graduates available for full-time work for that institution);
- Graduate full-time study (the proportion of Australian graduates proceeding to further full-time study); and
- Graduate starting salaries (graduates' mean nominal starting salaries).

This set certainly reflected good practice in that it represented a limited group of performance indicators but it still used proxies and was focused more on the quality and efficiency of the teaching process and output measures than exploring what graduates have learned or can do by the end of their studies (learning outcomes). It also measured what was easy to measure rather than what needed to be measured.

Universities were then assessed based on the values of the above indicators moderated by a range of characteristics of the university such as full-time/part-time student mix, socio-economic profile of students; tertiary education ranks of the commencing cohort etc. The 2007 moderation was complex and appeared to give perverse results reducing the faith of the institutions in the use of these learning indicators. Teaching and Learning performance for each university was then published by 4 broad discipline categories (Humanities, Arts and Education; Science, Computing, Engineering, Architecture and Agriculture; Health; and Business, Law and Economics) and graded to various performance categories (A1, A2, B and C).

This revised scheme remained in place for one year and was then abandoned, largely because of its complexity and lack of transparency. There was much debate in the

Australian higher education sector about the appropriateness and usefulness of the measures and a belief that some could be subject to manipulation or interference, highlighting one of the negatives relating to the use of performance indicators for resource allocation identified earlier. In addition the graduate employment indicators were affected by external economic factors beyond the sector's control.

Most of the indicators used in 2007 were ones that had been canvassed and trialled by the Performance Indicators Research Group in 1989, and none really addressed the issue raised by the HEC on the measurement of graduate attributes of a less utilitarian nature. And so, in spite of considerable effort and independent advice to the Government, the sector was no further advanced in the development of measures for true learning outcomes.

The above indicators of student satisfaction were based on the Course Experience Questionnaire (CEQ) which has been administered to graduates of undergraduate courses in Australian universities for over 20 years and asks student opinion about their acquisition of generic skills as well as their rating of the quality of teaching received and their overall satisfaction with their studies. These indicators of teaching quality rather than learning outcomes have shown great longevity in performance measurement in Australian higher education having been in continuous use since 1992. Later developments included a separate survey about student engagement which addressed the ways students interacted with their colleagues and their teachers. This was based on the National Survey of Student Engagement (NSSE) used in the US and introduced in Australia in 2008. It was administered to a sample of all students enrolled in undergraduate courses rather than graduates. Neither of these surveys directly addressed achievement of learning outcomes but attempted to link teaching quality to the responses to the surveys through some proxy performance indicators based on the scales used in the surveys.

In order to provide prospective students with more information about the quality of the higher education offered by providers in the sector, from 2014 the Australian government re-focussed attention on a new set of performance indicators labelled QILT – quality indicators of teaching and learning - by introducing an interactive website containing data. A report has also been published on work done by an expert working group established to develop the set of indicators of quality for teaching and learning

which used a set of Government endorsed surveys of enrolled students in Australian universities.

(see http://www.qilt.edu.au/docs/default-source/ues-national-report/report-on-the-development-of-the-university-experience-survey/ues2011developmentreporte12d8591b1e86477b58fff00006709da.pdf?sfvrsn=4)

The QILT indicators are based on the following three surveys of higher education students and graduates:

- the University Experience Survey (UES), measuring satisfaction and experience of current students;
- the Graduate Outcomes Survey examining employment outcomes of higher education graduates such as the nature of the employment gained post-graduation and the salary level achieved six months after graduation; and
- a new Employer Satisfaction Survey to assess the generic skills, technical skills and work readiness of graduates.

The University Experience Survey replaced the Course Experience Questionnaire in 2014 but is much broader and addresses other issues such as learner engagement, learning resources provided, student support provided by the institutions, and skills development as well as student satisfaction and the overall quality of the student experience. The last of these surveys has not yet been developed but could possibly address more directly some of the generic skills such as participation in teamwork, ability to learn independently, and learning collaboratively. In the meantime data on employment outcomes from the Graduate Destinations Survey, which has been conducted since 1972 is provided on the QILT website (www.qilt.edu.au) (also see http://www.graduatecareers.com.au/research/start/agsoverview/ctags/had/ for the history). These outcomes are what have in the past been described as positive graduate outcomes: the percentage of respondents either gaining full-time employment or in full-time study by the relevant census date of 31 October or 30 April in the year following graduation. The median salary earned by the respondents at that reference date is also included.

The data provided by QILT could potentially assist in the measurement of some of the learning outcomes frequently cited by Australian universities in their graduate attribute statements and will be considered later in this dissertation.

2.4 Linkages between learning outcomes, assessment and standards

2.4.1 Quality of teaching and learning and learning outcomes

In this way the path from performance indicators being developed and used to assess effectiveness and efficiency of higher education, through considerations of quality of teaching and learning, leads to an imperative for the sector to understand the extent of student learning and to measure learning outcomes.

To date in all systems where performance indicators have been used in a formal way to measure institutional or sector-wide student learning they have related to teaching quality and have focused on the process of teaching and student perceptions rather than the learning process itself (Chalmers, 2008). The indicators usually fit the production model of input-process-output measurement with little emphasis on outcomes, no doubt because these are much harder to measure. As shown by the previous historical summary there are no direct measures of learning for Australia at present and the proxy indicators in use are neither directly aligned with statements about learning outcomes nor goals for teaching and learning. So in order to derive some valid institutional indicators which actually measure students' achievement of learning outcomes a different approach should be taken, starting with detailed specification of the expected learning outcomes themselves.

Considerable work has been undertaken over the last two decades in various countries in specifying learning outcomes for courses at a range of levels. Most of these attempts are at discipline level and relate to particular groups of courses, especially those in professional disciplines. For example the Higher Education Quality Council of Ontario (HEQCO) is shepherding work done internationally in establishing what students should know and be able to do within a discipline through application of the Tuning Process. More recently, attempts have been made to generalise these discipline based outcome statements at the sector level (eg (Higher Education Quality Council of Ontario, 2013c) and (Association of American Colleges and Universities, 2010)). The various approaches are documented in detail in Part 3.

The main emphasis on specifying learning outcomes in Australia has been on developing graduate attributes as part of the quality agenda for the sector, in line with the original HEC recommendations. Each public university in Australia has specified a set of graduate attributes for its undergraduate programs and some work has been done by the Australian Learning and Teaching Council on analysing these statements and classifying them into a new typology they have defined (Barrie, Hughes, Smith, & Thomson, 2009). Graduate attributes could be considered to be statements of learning outcomes but there is some debate about this in the sector. They are often defined at the university level with no distinction between levels of course. In a few institutions the broad statements are customised for broad discipline groups. But critically there is little evidence of any approach to measurement. Rather, graduate attributes are used by universities in prospectuses or as part of marketing campaigns to show to future students and other stakeholders what a higher education at a particular institution will confer on a graduate on completion of a course – rather like a parenthood statement and not operationalised in any way. Learning outcomes can differ from these statements in that they are usually designed to specify in greater detail what is expected to be learned at the individual student level. However in reality when statements of learning outcomes are available and compared with graduate attributes on the surface there appears to be little difference in the semantics.

In any case, based on the evidence so far, it is possible to define learning outcomes at some level for institutions and the next critical step is to work out how to measure an individual student's learning against these learning outcome statements. In Australia higher education students' learning is tested in many ways through a range of assessment tasks during their courses of study, so the possibility of harnessing the information available through this individual student assessment and bringing it to bear on the formal measurement of learning outcomes should be explored. This requires strong links between the learning outcome statements and the nature of the assessment tasks set for the students. As stated earlier this task is not without its challenges.

2.4.2 Assessment and standards

Astin and Antonio (2012) present a conceptual framework of 'learning-oriented assessment' which provides such a link between learning outcomes, assessment and standards. Their model identifies three purposes of assessment: the judgment of student

achievement, the maintenance of standards and quality, and to promote learning. It is through this last component that assessment tasks are presented as learning tasks directly linked to specified learning outcomes for a student's course of study.

Hence the ability to measure attainment of learning outcomes even for the individual student is dependent on the type of assessment tasks set by teachers of the course and a minimum or threshold set of standards to enable the quality of learning to be assured. The examples of the alignment between learning outcome statements and standards undertaken in recent years in overseas educational systems generally stop short of describing the nature of appropriate assessment which will demonstrate achievement of the learning outcomes. Such frameworks present a useful starting point for this study and could be adapted to the Australian environment. For professional degrees these statements of outcome and standards expectation are frequently developed by the relevant professional body and in other generalist disciplines like arts, humanities, social sciences and science by discipline communities.

2.5 Conclusions, costs and benefits

This review of the past and current status of use of performance indicators in higher education highlights a gap in the measurement of teaching and learning outcomes and the difficulties which have prevented this being achieved.

The challenges outlined at the beginning of this chapter will need to be overcome but there seems to be an increasing interest internationally in being able to specify learning outcomes for higher education across national boundaries and at a generalised level both within disciplines and at the sector level.

From this work it appears that it may be feasible to describe a coherent set of learning outcomes together with a set of standards against which to measure individual learning outcome achievement. This study therefore builds on the gains made in recent times in Australia and overseas in the alignment of statements of learning outcomes and standards by adding the measurement dimension through exploring assessment approaches as the basis of outcome measurement of achievement by individual students. If this could be done and the individual results combined to provide an institutional measure, it would complete a conceptual framework for valid performance indicator development.

2.5.1 What makes a good set of indicators

In order for performance indicators developed as part of such a model to be useful and valid measures of learning outcomes they need to have certain properties. Many authors in this field (eg (Linke et al, 1991)(p7), (Dochy, Segers, & Wijnen, 1990), (Cave et al., 1991) and (Chalmers, 2008)) have identified problems with sets of indicators but these can be transformed into positive characteristics by taking a different perspective.

Based on such an analysis and this historical review, a good set of performance indicators should:

- be clearly related to the prime functions and objectives of the institution
- form a coherent set but *be relatively small in number*;
- be valid, objective, reliable and verifiable, and be defined and collected in a uniform and agreed fashion;
- Have a connection between the theoretical concept to be measured and the empirical variable used in the indicator (ie have *content validity*);
- be relative and not absolute in value;
- be calculated over time to demonstrate progress towards the achievement of the agreed target;
- not be ‘gameable’ – ie their values cannot be manipulated by the institution they are measuring;
- focus on what needs to be measured not on what is easy to measure;
- not encourage behaviours that 'score' rather than behaviour geared towards goal achievement; and
- not be driven by a desired political outcome.

This list of characteristics can provide a benchmark against which to evaluate any set of indicators for learning outcomes developed as part of this study.

2.5.2 Benefits of performance indicator development in this area

Benefits which might be expected to flow to the sector from the definition and implementation of a suite of performance indicators with the above characteristics include being able to:

- demonstrate the value of higher education to stakeholders such as students, governments, employers and the public by producing evidence of what students learn as a result of their study;
- improve the quality of teaching by involving teachers and other university staff in the specification and measurement process;
- assist students to become internationally competitive in their achievements; and
- demonstrate the necessary alignment between learning outcomes and the way they are assessed.

The insights provided into what students actually learn in higher education would also have benefits for the institution as a whole, and for students and teachers. These are likely to include:

- improvement in institutional productivity in the teaching of students through the identification of sets of principles for assessment and performance indicator development;
- identification of best practice examples of assessment which would assist in improving the quality of teaching and learning;
- a better understanding by students of what the aims of their institution are for their learning and how it thinks about learning outcomes; and
- provision of publishable information on teaching and learning performance for use in institutional ranking systems which parallels that used for research.

2.5.3 Costs

Depending on the approach used to address the absence of sector-wide information on assessment and grading in relation to the measurement of learning outcomes there may be a considerable cost in terms of establishment of an appropriate infrastructure to collect and process such data on a large scale. In some cases new assessment instruments similar to those used in the US and the Assessment of Higher Education Learning Outcomes (AHELO) project (Coates, 2008) may be required and research undertaken on how best to use these results to inform performance indicator development.

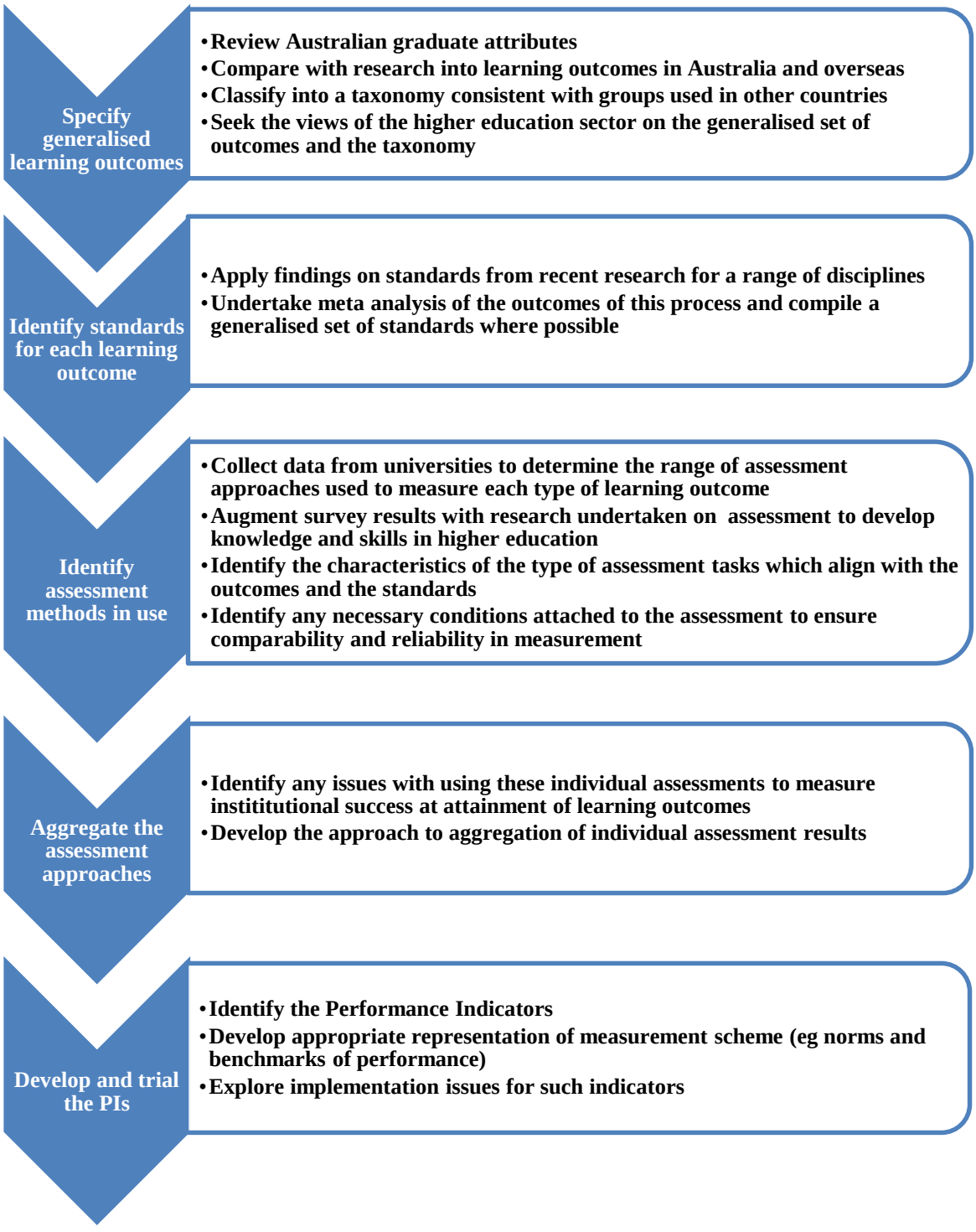
If this study is successful in defining appropriate institutional performance indicators, the application of the principles identified may require a reshaping of curriculum and

modes of assessment which could be more labour intensive than at present. The documented attempt of this by Carless *et al* (Carless, Joughin, & Liu, 2006) at Alveno College in the US identifies the extent of the reorganisation of teaching and assessment that was required to focus on learning outcomes and the associated costs.

2.5.4 Next steps

This review of the history of performance indicator development and use and its focus on the Australian higher education environment has identified gaps in the specification of a set of learning outcomes for the sector and in how achievement of such outcomes might be measured. This study plans to address these aspects through the development of a measurement model and the strands of investigation identified in Chapter 1 and shown below in greater detail.

Figure 2.1 – Summary of steps to be taken towards the development of performance indicators for measurement of higher education learning outcomes



PART 2: DEFINING AND CLASSIFYING LEARNING OUTCOMES

Chapter 3 – Specification of learning outcomes

This chapter addresses the question - What is it that students should learn in their study for a higher education degree?

The aim of this part of the study is to specify a set of generalised course-based learning outcomes which is applicable across disciplines and consistent with the range of graduate attributes or competencies specified by Australian universities. Because of the aim of developing performance indicators which measure institutional effectiveness at the achievement by its students of the learning outcomes, it is important that those developed have a sound pedagogical basis but also have good potential for measurement.

To do this several issues should be addressed: how large the set of outcomes should be; whether it should have clusters and different levels of descriptors; and at what level of detail the learning outcomes should be described to facilitate measurement. The tension between a smaller number of more generalised outcomes which would be manageable practically but may be too broad to distil meaning, and a large number of very detailed ones which may not be comparable across disciplines or institutions requires resolution.

3.1 Origins and feasibility of specifying learning outcomes in higher education

There are various views about the origins of the use of learning outcomes in higher education. Nusche (Nusche & Organisation for Economic Cooperation and Development, 2008) states that ‘the term learning outcomes has its origin in outcomes-based education’ (p8) in which curriculum and assessment are both organised around statements of anticipated learning outcomes. The link between assessment and learning outcome description is a key factor in the feasibility of the measurement of learning outcomes at the centre of this research study. Nusche’s view is that provided the alignment between the statement of learning outcomes and the modes of assessment is achieved comparative measurement of learning outcomes becomes feasible (Nusche, 2008) (p8).

Melton (1996) argues that the development of learning outcomes in higher education arose from the specification of competencies in applied and vocational courses. He considers that when higher education expanded and changed with the emergence of

broader curriculum and much less obvious career paths for university graduates on qualification (as discussed in Chapter 2), it became necessary to think about the more general knowledge and skills needed for university graduates to enter any of a wide range of professional areas as employees. This led to the desire in universities to ensure graduates developed what Melton calls 'core skills' of 'problem solving, communication, learning to learn, working with others, numeracy, information technology, competency in a foreign language'. In Australia these are usually referred to as generic skills.

More recently Ewell (2010) alludes to the growing competition of a global higher education market for students and funding as drivers for the ongoing need to specify learning outcomes accurately. He argues that this competition results in greater mobility of graduates in search of employment and further study options in which they will seek credit for work undertaken in different jurisdictions. In his view this has led to various mechanisms which are 'manifestations of the global standards movement'. Ewell quotes quality assurance networks such as the Bologna process, the increasing prevalence of international university rankings, alignment efforts for standards such as the Tuning profile and approaches to international comparative assessment such as the AHELO project sponsored by the OECD, as evidence of an increasing need to specify graduate learning outcomes effectively and measure them accurately.

The ongoing focus on learning outcomes as an approach to quality assurance of teaching and learning in higher education has a number of positives and negatives. Ewell believes that being able to compare outcomes and measure them to allow comparisons across populations and different settings is a significant positive, against which he sets a number of pitfalls which include:

- Conceptual and operational difficulties in defining the outcomes;
- Questions from traditional academic staff about the legitimacy of the exercise;
- Potential to lose important distinctions between the programs offered by different institutions;
- Potential to lose the sense of the whole by defining a list of abilities and knowledge rather than looking at the overall interactions of the parts; and
- The problem of not being able to specify all important outcomes in advance.

(Ewell, 2010)(p7)

In this study it will be necessary to consider these issues if progress in measuring institutions' performance is to be made in a valid way.

Achievement of a set of course learning outcomes is the result of complex interactions between expected knowledge and skills executed through a range of individually taught subjects, many of which are discipline specific. Sets of learning outcomes also usually include a range of personal attributes expected of graduates, as identified by Melton and others, but in some cases these may not be directly the result of the higher education experience. This presents one of the difficulties in the precise specification of such outcomes and their measurement which must differentiate between what has been learned at a higher education institution and what other skills or personal characteristics arise solely from individual development or maturation.

Hence while it may be possible to describe a comprehensive set of learning outcomes which includes knowledge, skills and personal attributes, only some may be related to the education process, and even fewer will satisfy the requirement of alignment with assessment. This study is concerned with the smaller subset of a group of learning outcomes which can be attributed to the learning process, can potentially be directly assessed, and whose results can be aggregated to a level which allows comparison between institutions.

3.2 Sharing values about learning – specifying what students should learn

The approach used is to examine the strengths and weaknesses of various international and Australian attempts to specify learning outcomes and to draw conclusions on their relevance and applicability for the development of an Australian national set of outcomes and standards at the undergraduate level. These attempts have all used inputs from a range of groups which include academic staff within the universities, State and federal governments, students and industry. Hence gaining a balance between top-down and bottom-up approaches to the specification of learning outcomes has largely been addressed in each of the examples discussed.

Considerable work has been done on describing learning outcomes during the last 20 years. The end results differ in purpose and usefulness in the evaluation of the achievement of these outcomes which is the matter of interest to this study. Similar terms are used in various taxonomies and descriptions but their meanings often differ.

Because of this it is sometimes difficult to conclude that the learning outcomes proposed in various jurisdictions are similar. Consideration of the semantics of the terms and how they are used in practice is needed to determine whether a particular approach could be applied to the Australian higher education sector.

It therefore seems a good idea to start with some definitions and criteria for evaluating the appropriateness of any group of outcomes and standards.

3.2.1 Definitions

A common definition of a higher education learning outcome is *the personal and societal changes or benefits that follow from the experience of learning* (eg (Chalmers, 2008) (p15), and (Nusche & Organisation for Economic Cooperation and Development, 2008)(p7)) but often it is assumed that the meaning of the term being used is obvious. The above definition recognises the impact of student learning in higher education both on the individual and on society as a whole. Similarly, in the more recent Degree Qualifications Profile (Lumina, 2011) developed by the National Institute for Learning Outcomes Assessment (NILOA), learning outcomes also include attributes relating to external impacts and are defined as 'information and skills that graduates need for work, citizenship, global participation and life' which are focussed on life after study.

Others have defined outcomes in more personal or operational ways which include an aspect relating to assessment. Shavelson (2009)(p9) focuses on personal growth and defines learning and by imputation a learning outcome as a 'permanent change in observable behaviour over time'. This is a behaviourist definition from his perspective as a psychometric analyst. It underpins the conceptual framework he has devised which describes the nature and relationship between types of learning outcomes. In contrast work done in the UK by Otter uses an operational definition (1992) that a learning outcome is a description of a course objective with a clear statement of assessment requirements is far more pragmatic. The more recent European Tuning process also links the learning outcome to a standard of achievement and describes it as the extent and the level of standard of competence including knowledge that a student will develop by graduation (Lennon et al., 2014)(p 19). This suggests the need for distinction between the concepts of standards, competence, achievement and outcomes as these terms are often blurred with 'competences', sometimes used interchangeably with learning outcomes. For example the application of the Tuning profile approach across

different disciplines and institutions undertaken by Higher Education Quality Council of Ontario (HEQCO) uses the label *competence* as part of a taxonomy which is a higher level classifier of detailed learning outcome statements, while the Tuning Process approach uses the term *competence* in lieu of generic skills.

The Australian Qualifications Framework (AQF) states that learning outcomes are 'constructed as a taxonomy of what graduates are expected to know, understand and be able to do as a result of learning' (Australian Qualifications Framework Council, 2013)(p11). This Framework is focussed on providing advice to universities on the characteristics graduates from particular degree levels are expected to have and is used in Australia primarily as a reference document for regulation and accreditation purposes.

From the examination of approaches which follows it is apparent that the definition of the term 'learning outcome' is influenced by the purpose of the particular framework in which it appears and it is clear that there is no single ideological position developed to date by practitioners. All approaches make an assumption that it is possible for students to achieve the learning outcomes as specified and that there is a valid and reproducible way to assess their achievements, but the hope of obtaining a shared understanding across national, institutional and disciplinary boundaries remains elusive.

Given the emphasis of this study and the lack of consensus in the field, the following definitions of terms are used in the comparative analysis of the various approaches, and in the development of an appropriate set of outcomes and taxonomy for Australian higher education:

> **Learning Outcome** – a short description of knowledge, skill or a personal attribute resulting from a course of study containing a criterion for achievement which provide insight into how the outcome can be measured.

The term standard in this study is not used as a general description of quality of a cohort of students or the appropriateness of a particular course of study as in 'Standards are falling'. Instead it is used to describe a level of individual achievement in relation to a learning outcome. The Australian Universities Quality Agency (AUQA) discusses in detail what it understands to be an academic achievement standard to be applied to individual or group learning which appears to be suitable for the purpose of this study.

Sadler (1987)(p194) defines a 'standard' as 'a definite level of excellence or attainment, or a definite degree of any quality viewed as a prescribed object of endeavour or as the recognised measure of what is adequate for some purpose, so established by authority, custom, or consensus'.

This definition is generalised to address any situation but implies the existence of an agreed criterion and an indication of level of achievement. This is also included in the AUQA definition of an academic standard which will be used for the purpose of this study.

> **Standard** – an agreed specification or other criterion used as a rule, guideline or definition of a level of performance or achievement.
>
> (Australian Universities Quality Agency, 2006)(p8)

The term *competence* as a descriptor of a learning outcome is not used in the framework developed later for learning outcomes for the Australian higher education system, but may be a term used in the description of the level of achievement of a standard. The term used in this study, in situations where competences are a subset of learning outcomes in other approaches, is generic skills. Hence,

> **Competence** is a satisfactory level of performance or achievement of a standard associated with a learning outcome.

3.2.2 Criteria for assessing appropriateness and measurability

The approaches to specifying learning outcomes also vary in the levels of their taxonomies, the scope of the set of learning outcomes included, and in their interest in standards and assessment. In one sense setting of standards strays into the field of assessment, which is the next step of the measurement model proposed in this study, but thinking about how the learning outcomes relate to assessment is important if the measurement of the achievement of learning outcomes is at all feasible. The evaluation of the various approaches benefits from a clear statement of the criteria against which they will be assessed for applicability to the Australian environment and whether they have the necessary potential for assessment and measurement.

Melton (1996) (p416) suggests a number of issues of principle that should be considered when devising sets of learning outcomes. A primary concern is whether a 'one size fits all' approach to defining learning outcomes and standards at the institutional or

discipline level is possible or appropriate. His other issues include whether the focus is on individual learning or the needs of others such as Government and the professions; the need for a classification scheme or taxonomy for the set of learning outcomes to allow direct comparison between different entities; and the need for national standards for measurement of the learning outcomes to allow for comparability of attainment across institutions and disciplines.

In the model presented in this study the emphasis is on measuring both individual student learning and considering the interests of the government and funding bodies in wishing to differentiate performance at the institutional level. The last two of Melton's points of principle are therefore endorsed and taken into account in the criteria enunciated below for the evaluation of the proposed learning outcomes framework explored later for Australian undergraduate education.

Criteria for evaluation of approaches to the specification of learning outcomes

From the analysis of attempts to specify higher education learning outcomes, a set of criteria has been developed which could be used to form a judgment on whether a set of such outcomes would be useful in the evaluation of student learning. For the purposes of this study, the characteristics of a set of learning outcomes, appropriate to Australian higher education, might be:

- Comprehensive covering knowledge, skills and personal attributes;
- Aligned with the expectations of the higher education sector in relation to graduate attributes;
- Relatively small in number but sufficient to cover the breadth of learning undertaken;
- Applicable to both professional and generalist education;
- Generalisable across disciplines;
- Clear concise statements which are devoid of jargon so they are easily understood by all stakeholders;
- Framed as a set of actions which can be taken rather than a passive statement of what needs to be learned;
- Inclusive of appropriate standards specific to each outcome in the set;
- Able to be assessed directly against the standards relevant to the outcomes;

- Part of a conceptual framework or taxonomy which allows different levels of aggregation of the achievements of the outcomes to allow institutional comparisons to be made; and
- Developed through top-down specification from the policy perspective but affirmed by practitioners in a bottom-up analysis.

3.3 Feasibility of specifying learning outcomes

Attempts to specify and measure higher education learning outcomes at a state or national level commenced in the 1990s and are still occurring today. Those examined in this study include an early approach in the United Kingdom (Otter, 1992) and a series of more recent examples in various countries:

- The Tuning Process and degree profiles (Europe, UK, in 2000);
- The American Association of Colleges and Universities (AACU) LEAP program and VALUE rubrics (US, 2005);
- Shavelson's work on developing a conceptual framework for student learning outcomes (US, 2010);
- The Degree Qualifications Profile (DQP) of the Lumina Foundation (US, 2011);
- Generalisation of the Tuning process across disciplines by the Higher Education Quality Council of Ontario (Canada, 2013); and
- The Australian Qualifications Framework (latest version 2013).

The Otter project was a practical attempt to examine the feasibility of describing degrees in terms of their learning outcomes**.** The methodology involved a consultative approach with staff and students from a large number of UK universities and polytechnics and a group of graduate employers. Otter's work was ground-breaking in several ways. She developed a conceptual model for learning outcomes which forced alignment between course objectives and assessment; proposed a practical staged approach to specifying learning outcomes which related them to graduate attributes; favoured the use of classroom-based assessment approaches in ascertaining whether the outcomes had been achieved; and demonstrated that it was possible to gain a degree of consensus about what students were required to learn in some categories of learning across several disciplines. This was done by generalising work done by a group of four discipline panels, each of which undertook its work in a different way.

Otter's five stage methodology for the development of learning outcomes was as follows:

- **Stage 1** – Write a key statement which contains the notion of a good graduate, and the underlying ethos and values which apply to the courses to which the outcomes apply.
- **Stage 2** – Identify which of these attributes fall into three broad categories: skills, knowledge and personal qualities.
- **Stage 3** – Identify the major activities in the course which relate to the Stage 2 categories.
- **Stage 4** – Develop learning outcome statements consistent with the Stage 3 activities.
- **Stage 5** – Develop statements which describe what the student has to do to demonstrate the learning outcome.

This is an ordered approach which may prove useful in the development of learning outcomes for undergraduate Australian learning outcomes and related standards undertaken later.

Otter encountered several challenges in her work and in the end did not succeed in describing a single set of learning outcomes across the four disciplines. However she developed a classification of the outcomes which allowed some comparison of them and their assessment results across a category called *core outcomes* which resemble generic skills. Her classification of learning outcomes was:

- **Core outcomes** –These were outcomes that should be achieved by all graduates, regardless of subject (for example, ability to communicate complex information or ability to solve problems).
- **Subject specific outcomes** – These outcomes reflect the values and cultures of the academic disciplines involved in the course and are closely related to the subject matter of the degree.
- **General outcomes** –These reflect the underpinning skills needed in employment and are most common in the technical or professional disciplines (for example,

safe handling of materials in fieldwork, knowledge of a range of teaching techniques).

- **Other outcomes** – Outcomes which are specific to a particular institution or discipline (for example, proficiency in a second language or a particular type of information technology).

Using this taxonomy Otter concluded that it was possible to specify higher education learning outcomes explicitly at a course level but that those descriptions would vary over time in a complex and changing educational environment. This is an interesting point as most current approaches assume that graduate attributes or the more specific learning outcomes are fixed. In Australia they appear to have changed little from their inception in the sector in the 1990s and up until recently do not appear to have been regularly reviewed.

Otter also devised guidelines for the development of learning outcome statements which she believed could be applied independently of the discipline and the approach taken to their formulation. These stated that a set of learning outcomes should:

- include subject specific knowledge as well as skills and methods;
- address qualities sought by employers;
- be short statements involving a subject, object and action verb;
- have related short standards statements which qualify the outcome in a way and give insight into how the outcome could be assessed; and
- be intelligible to all stakeholders.

All of these have been taken up in the proposed criteria listed earlier for evaluation of different approaches to learning outcome specification.

3.4 The Tuning process and degree profiles– Europe and the United Kingdom

The Tuning project commenced in 2000 with the aim of contributing to the Bologna process by encouraging implementation of that process in universities. It concentrated on ‘development of a common language in the description of higher education programmes’ (Lokhoff et al., 2010) (p11). It was a collaborative project originating in Europe, and by 2005 the project began focussing on what students learned in academic

programs rather than on those programs' content and credit transfer arrangements. The emphasis of the project therefore shifted from the course to the student and to an outcomes rather than inputs perspective. The intention was that the Tuning project would produce a practical and consistent approach to documenting these outcomes which could be included in Diploma Supplements or Graduate Statements on graduation. However, the project discovered that 'the ways in which competences and learning outcomes were described by universities were so diverse, ranging from short lists of very general statements to lengthy descriptions of several pages, that they could not be used as a coherent and balanced source of information by the target group' (Lokhoff et al., 2010)(p12). This diversity of description had to be addressed to achieve the goal of a common language envisioned by the Tuning process.

A second project was then set up to develop a guide which would provide a consistent way of describing learning outcomes and skills across institutions and countries for different levels of degree qualifications and disciplines. The Tuning Profile which resulted describes 'in terms of learning outcomes what graduates will know, understand and be able to do *by the time they have successfully completed the programme'* (p 15), underlining its application in the awarding of credit as originally intended in the Bologna process. The project team wished to demonstrate that qualifications awarded on graduation to students were 'fit *in* purpose as well as fit *for* purpose' (p19), thereby ensuring that learning outcomes met the expectations of the students themselves as well as those of society and employers. Demonstration of fitness for purpose of graduate attributes has been used in many quality assurance regimes including in Australia but fit *in* purpose with a focus on the student's needs is less common.

Learning outcomes are distinguished from *competences* (or generic skills) in the Tuning analysis. In this interpretation, a learning outcome contains a concept of measurability and is described as 'a **measurable** result of a learning experience which allows us to ascertain to which extent/level/standard a *competence* has been formed or enhanced' (p21). Hence, by this definition, learning outcomes, which may encompass development of generic skills, must be assessable to allow measurement of achievement, and are related to courses of study.

Tuning profiles have been developed for nine disciplines. These profiles contain very detailed information on the nature of qualifications with explicit lists of competences (or

generic skills), knowledge and other learning outcomes described in the context of each discipline.

The Tuning process also led to a set of essential components of a Bologna degree known as Dublin Descriptors to classify learning outcomes. These provide a taxonomy for the classification of the learning outcomes.

- Acquiring knowledge and understanding;
- Applying knowledge and understanding;
- Making informed judgements and choices;
- Communicating knowledge and understanding; and
- Capacities to continue learning.

The establishment of this qualifications framework and these broad learning descriptors mirror similar frameworks established in Australia, Canada and the United States but in the Tuning case they have been used to identify desired learning outcomes aligned across disciplines and which are relevant to a course level.

Tuning addresses several of the philosophical issues identified at the start of this chapter on size and scope of the set of course-based learning outcomes and suggests a mapping approach to ensure coherence. The Tuning Guide (Lokhoff et al., 2010) advocates fewer than 20 outcomes and use of succinct descriptive statements. The expected outcomes are first mapped for each subject of the particular degree under consideration (ie curriculum mapping). Through assessment of the subjects mapped in this way, this approach is one way of addressing the challenge of assessability of learning outcomes at the course level.

The criteria outlined in Section 3.2.2 about succinct, clear statements with active components of achievement and the necessary alignment between outcome and standards are demonstrated in the good practice model provided (pp44-45) in the Tuning Guide. Hence the Tuning model appears to have many of the characteristics which seem to be necessary to enable comparability of learning outcomes across different institutions, but only within a particular discipline. It does provide a common language for the development of curricula while allowing institutional flexibility and autonomy in their application (Tuning process, 2010). This capacity addresses concerns raised by the higher education sector that specification of learning outcomes results in homogeneity of course curriculum and stifles innovation.

The Tuning process is being developed further and has been applied successfully in many countries of the European Union. Its strengths are its flexibility while providing a clear methodology for institutional practitioners. However it remains discipline-centric and accreditation/compliance focused rather than performance or quality based.

3.5 Tuning across disciplines

The generalisation of the discipline-based Tuning work has been taken up by the Higher Education Quality Council of Ontario (HEQCO). Since 2011 HEQCO has been working in partnership with Ontario's colleges, universities and external organisations (Higher Education Quality Council of Ontario, 2013a) to identify a set of generalised learning outcomes for the same levels of degrees as the original Tuning Profile. A Tuning Advisory Panel has been examining similarities and differences between learning outcomes specified under Tuning across different disciplines in Ontario higher education institutions, and in 2014 published a set of learning outcomes appropriate to the Ontario higher education sector. This document bridges the gap between the broad 'degree qualification' learning outcome frameworks and the narrower 'program based' learning outcomes' seen in the original European Tuning work (Higher Education Quality Council of Ontario, 2013c).

This cross disciplinary Tuning approach uses different terminology to the original discipline based profiles. It almost reverses the use of the terminology in the single discipline approach with competences as the higher level classification of learning outcomes rather than a subset of them. The learning outcomes defined by HEQCO satisfy the test criteria relating to concise descriptions which are in plain language and they are framed using action verbs. However they give little insight into standards for measurement or levels of achievement expected and the scheme is therefore less useful for the current study with its focus on assessment and measurement.

The classification structure has six categories of learning outcomes (called competences), and five of these each have between five and six sub-categories all of which are independent of discipline of study. The sixth learning outcome below is discipline dependent (Life and Health Science, Physical Science and Social Science):

- Knowledge
- Critical and creative thinking
- Communications

- Social responsibility
- Personal and interpersonal, and
- Practice and methods.

The learning outcomes are specified for two-and three-year diplomas, bachelor and masters degrees.

The HEQCO system does contain a full range of learning outcomes including generic skills and knowledge and in spite of the three different groups in the *Practice and Methods* group there is a lot of overlap. The main difference between the discipline specifications is that the Physical Science group is more oriented towards experimentation and modelling. This suggests that the specification by discipline might be politically based rather than pedagogical and it appears that a single model might be possible if general terms were used rather than the specific language for the discipline (eg. expertise in practical skills relevant to the discipline).

The HEQCO approach is not favoured as the basis of an Australian framework for learning outcomes because of the lack of standards, the confusing use of terminology, and difficulty in aligning the learning outcomes with assessment tasks. In this context HEQCO's Research Report 2013-14 points out that 'both in Ontario and around the world, groups have become very skilled at writing learning outcomes' and 'there seems to be little added value in further projects to write more, or even refine current, descriptions of learning outcomes. Rather, the key impediment is now assessment' (Higher Education Quality Council of Ontario, 2013b). The Council's advice is that learning outcomes 'are best measured as part of the normal and on-going evaluations of students as they proceed through their course of study', and it warns that there are some significant impediments if measurement of these learning skills is pursued as an 'add-on' (Higher Education Quality Council of Ontario, 2013b)(p3). Such an approach is consistent with the proposal to consider use of classroom based assessment later in this study.

3.6 Shavelson's learning outcomes conceptual framework

A valuable contribution to the research on the measurement of learning outcomes in higher education is from Richard Shavelson. He describes a framework for specifying and assessing learning outcomes in college education in the US (Shavelson, 2009). Shavelson is an educational psychologist who has made many contributions to

knowledge in the field of assessment, aptitude testing, educational evaluation, and in the application of generic skills testing instruments such as the Collegiate Learning Assessment. His conceptual framework for learning outcomes is based on his research on cognition and cognitive abilities (Shavelson & Huang, 2003).

He considers that a generalised set of learning outcomes for higher education should be comprehensive, as indicated in the criteria identified for this study, stating that they should 'go beyond cognitive to individual and social responsibility outcomes including, for example, the development of one's identity, emotional competence, perspective taking and resilience' (p6).

Based on his definition of learning as 'a permanent change in observable behavior over time' his framework covers the following high level tiers of learning outcomes:

- Knowledge and reasoning in the majors;
- Broad abilities such as critical thinking, the ability to write critically and reason analytically;
- Ability to adapt or learn in novel situations; and
- Demonstration of individual and social responsibility. (p9)

These four broad categories are similar to those used by Otter and in the 'competences' of the Tuning work, but Shavelson makes the point that the learning outcomes must be attributable to the learning experience. Because his definition of learning involves change over time, his approach to measurement is one of comparing achievement in learning outcomes at both the start and end of a period of study. Shavelson also identifies the issue of 'recursive complexity' which he believes is at the centre of specification and the measurement of the outcomes. This is a circular model of learning where capability is a product of the learning process and the natural attributes of the students. Shavelson states that 'what is learned and how it translates to new situations depends on the students' natural endowments, aptitudes and abilities, which are a product of their education' (pp 13-14). This suggests that accurate measurement of achievement must be controlled for the impact of these characteristics, and this may need to be taken into account when attempting comparative performance measurement between institutions.

Shavelson refers to the Spellings Commission's work (Spellings Commission, 2006) which highlighted the need to measure institutional success in terms of student

achievement against a clearly enunciated set of learning objectives, and to report this success in an 'accessible, understandable way' so that 'the relative effectiveness of different colleges and universities' is made clear' (p4). In Shavelson's view, specific higher education learning outcomes should be based on a set of characteristics which graduates can be expected to have on completion of their college studies (ie graduate attributes). In fact he based his taxonomy of learning outcomes on research undertaken in 2000 (Immerwahr, 2000) which involved surveying residents in New York on what characteristics, skills and knowledge were desired for higher education graduates. The attributes or traits of US college graduates he used were:

- 'A sense of maturity and ability to manage on their own;
- The ability to get along with people who are different to themselves;
- Improved problem solving and thinking ability;
- High technology skills;
- Specific knowledge and expertise in their chosen career;
- Top notch writing and communication skills; and
- Responsibilities of citizenship'.

(Table 3)

Hence, like Otter's recommended methodology, he started from the perspective of what being a graduate means rather than specifying in detail the actual outcomes expected from an analysis of content of a course of study or a usual set of generic skills.

Shavelson's work has the strength of being based on research in cognition and cognitive abilities. It therefore presents a theoretical basis for the nature and classification of appropriate learning outcomes – that is, it goes to the heart of how students learn and how this impacts on the type of learning outcomes that should be considered. His framework also highlights the non-linear nature of building up learning through a course of study and how this would need to be assessed to get valid performance information. Shavelson's work is therefore a significant influence in the approach used later in this study for formulating learning outcomes based on graduate attributes in the Australian environment.

3.7 AACU LEAP and VALUE rubrics

In 2005, the Association of American Colleges and Universities (AACU) launched the Liberal Education and America's Promise (LEAP) initiative, 'a long term effort to promote the value of liberal education' for individuals and for the US economy (Association of American Colleges and Universities, 2013b). The AACU states that LEAP embraces a '21st century definition of liberal education' and promotes an integrated approach to higher education through:

- Essential Learning Outcomes (ELOs);
- High Impact Educational Practices;
- Authentic Assessments; and
- Inclusive Excellence.

The Essential Learning Outcomes (ELOs)

The LEAP project involved extensive consultation with key stakeholder groups, including educators and employers. The set of ELOs is cross-institutional and/or discipline-independent and outlines what knowledge and skills students should attain as a result of their education, namely:

- Knowledge of Human Cultures and the Physical and Natural World;
- Intellectual and Practical Skills;
- Personal and Social Responsibility; and
- Integrative and Applied Learning.

Under each of these areas, are several broad learning outcomes which are short statements similar to generic skills plus 'study in the sciences and mathematics, social sciences, humanities, histories, languages and the arts' reflecting the LEAP's origin in liberal education. These descriptors are at a level of detail similar to the lists of graduate attributes in many Australian universities. The AACU reports (Association of American Colleges and Universities, 2013a)(p1) that few valid assessment approaches exist for the last three ELOs in spite of these attributes being the primary skills gained in a liberal education. Consequently it has put effort into the development of a set of Rubrics (benchmarks for achievement) which define a new approach to assessing and grading the broad learning outcomes associated with the ELOs. These rubrics define what is intended by the broad ELO descriptors.

The VALUE rubrics

The rubrics for the Valid Assessment of Learning in Undergraduate Education (VALUE) project were released in their current form in 2009 (Association of American Colleges and Universities, 2010). It is reported that the rubrics for the last three ELOs listed above have been tested and validated by staff in over 100 institutions in the US. From the perspective of this study, of interest is the VALUE project's focus on the need to apply direct forms of assessment to the second tier LEAP learning outcomes in order to 'define, document, assess and strengthen student achievement of the Essential Learning Outcomes in undergraduate education for student success' (Association of American Colleges and Universities, 2013c). Hence the rubrics aim to articulate the type of assessment which is needed to allow measurement of achievement of the agreed learning outcomes. The learning outcomes statements apply to all disciplines and prescribe four levels of attainment to allow differentiation between individual students' performance. The VALUE Rubrics and the statements of learning outcomes therefore might be able to be adapted to the Australian environment and provide base data on achievement for comparison of institutions. This view is supported by the AACU's claims that the current version is now in use in 'thousands of campuses and individuals in the United States and around the world' and by 'whole higher education systems and multi-State consortia' as a guide for measuring student learning (Association of American Colleges and Universities, 2013b). A report on using the rubrics has been published which documents the experiences of twelve US institutions in gathering information on student performance and using the findings to adjust assessment tasks (Association of American Colleges and Universities, 2013b). These experiences will be explored further in the next part of this study relating to assessment.

In spite of its potential the VALUE approach does not meet several of the criteria for evaluation set earlier and would not be appropriate for non-generalist undergraduate higher education in Australia. The ELOs are not a comprehensive set and exclude the discipline related knowledge and skills considered important by Otter, Tuning and Shavelson. For example, the 16 first tier learning outcomes (i.e. the sections of the VALUE taxonomy) are short phrases and the second tier statements are slight expansions of these. The real descriptions of what is to be learned, which would be comparable in detail to the HEQCO Tuning learning outcome statements, lie in the very detailed four Rubrics for each of the 88 second tier short outcomes (i.e. equivalent to the

criteria of the standards). The rubrics are framed as a set of student actions and provide detailed statements of benchmarks for assessing achievement but they are specialised and complex and more suited to the liberal US College environment. At a pragmatic level there are too many of them and they are too detailed for comparative analysis of a practical kind leading to the formulation of a small set of institutional performance indicators.

And so on balance the VALUE Rubrics would not be easily adaptable for undergraduate education in Australia for use in this study. In addition the first and second tier 'learning outcomes' are not a good match the range of characteristics, knowledge and skills sought by Australian universities for their graduates. For example, there is much more emphasis on civic responsibilities in the VALUE system and no emphasis on sustainability, equity, innovation and research skills which are common graduate attributes in Australian higher education.

3.8 The Degree Qualifications Profile (DQP) of the Lumina Foundation

The Degree Qualifications Profile was developed under the auspices of the Lumina Foundation of Indianapolis, a private foundation which has a goal 'to increase the proportion of Americans who hold high-quality degrees and credentials to 60% by 2025' similar to the current Australian higher education attainment goal adopted by the government in 2009. In 2011 the Foundation developed a Degree Qualifications Profile (Ewell, Gaston, & Schneider, 2011) which outlined 'what students should be expected to know and do once they earn their degrees' (p1) for a range of different program (course) levels – associate, bachelor and masters degrees. The DQP consists of a set of learning outcomes which are independent of discipline and which were drawn from 'more than a decade of widespread debate and effort' in US higher education, 'to define expected learning outcomes that graduates need for work, citizenship, global participation and life' (p1).

The DQP was developed by experts from the Institute for Higher Education Policy, the National Center for Higher Education Management Systems (NCHEMS), Kent State University, and the Association of American Colleges and Universities (Ewell et al., 2011). Similar to the Tuning process with its focus on accreditation and credit, it provides a taxonomy for classifying learning outcomes for each level of study by five

basic areas of learning, which do not necessarily have the same emphasis for all higher education providers:

- Broad integrative knowledge – 'learning that is broad enough to support inquiry into the relationships among subject areas and the integration of related realms of knowledge';
- Specialised knowledge – 'deep knowledge which assures mastery of strategically chosen subject areas';
- Intellectual skills – 'manifestations of well-defined cognitive capacities and operations' spanning both the above knowledge and applied learning dimensions;
- Applied learning – a set of skills which demonstrate 'what graduates can do with what they know'; and
- Civic learning – requires the integration of knowledge and skills as intellectual skills and in the specialised discipline study.

(pp8-9)

The authors of the DQP stress that in practice there is considerable overlap and integration between these five dimensions in the actual learning experience. For example 'students gain conceptual understanding and sophistication both by exercising their intellectual skills and by applying their learning to complex questions and challenges in academic and non-college settings' (p5). This is similar to Shavelson's model of 'recursive complexity' of cumulative learning which does not necessarily occur in a linear sequence.

The DQP itself does not specify outcomes but rather provides some examples of what statements of learning outcome might look like for various course levels. These exemplar learning outcomes are summative for each type of degree in the profile, and imply threshold levels of required achievement or basic levels of knowledge or skill. The DQP authors use spider web diagrams with the five dimensions representing the areas of learning, to allow institutions applying the Profile to determine whether they are out of kilter with state averages. It is therefore a useful tool for comparative course design and accreditation but does not really focus on expressed learning outcomes per se.

At present it is reported that there are over 260 universities and colleges in the US working with the DQP as an approach to learning outcomes and quality assurance, with a little over half of these institutions supported by Lumina funding for this work (Schneider & Ewell, 2013).

Because of its different purpose, the DQP approach does not satisfy many of the criteria for specifying learning outcomes for measurement identified for this study. For example it does not contain a comprehensive list of outcomes and does not address the issues of assessability and standards.

3.9 The current Australian environment and the AQF

Since 1992 when the Australian Government accepted the advice of the Higher Education Council in Achieving Quality (Higher Education Council, 1992a) and its view of the centrality of graduate attributes to the quality of the higher education system, each university has moved to define a set of graduate attributes. Many of these sets have changed little since they were formulated in the 1990s and given information published since, little has been done on evaluating whether the attributes have actually been achieved. The attribute statements available on university websites are generally very broad, not distinguished by the level of course or discipline, and a great distance from the detailed statements of learning outcomes discussed above which are in place in some other countries. The Australian Universities Quality Agency (AUQA), which was responsible for assuring the quality of all aspects of university operations in the Australian higher education sector between 2000 and 2009, reported (Stella & Woodhouse, 2007) that there was a need particularly to strengthen 'academic standards in areas such as standards for final year undergraduate assessment and graduate outcomes and satisfaction' (p29).

A government policy focus on learning outcomes in higher education in Australia occurred with the publication of the report of the Review of Higher Education (Bradley et al., 2008) which argued that for Australian universities to remain internationally competitive, the nation 'must enhance its capacity to demonstrate outcomes and appropriate standards'. The Bradley Panel argued for 'more systematic processes ... at both institutional and individual discipline level to provide stronger assurance of academic and organisational standards' (p128).

In response the Australian Government set up a new commission, the Tertiary Education Quality and Standards Agency (TEQSA), and stated that future quality assurance arrangements should include development of 'a set of indicators and instruments to directly assess and compare learning outcomes; and a set of formal statements of academic standards by discipline along with processes for applying those standards'(Commonwealth of Australia, 2009) (p61). Funding was provided by the Government to the Australian Learning and Teaching Council (ALTC) to support research work on these learning outcomes and standards issues. This resulted in a number of projects related to graduate attributes and the development of standards for a range of disciplines, which are discussed later.

The Australian Qualifications Framework, which had existed since 1995, was a primary protocol document for ensuring that courses offered by different institutions at the same level were comparable in requirements and learning outcomes, but it was mostly ignored by universities except when there were boundary issues about degree courses between the vocational education and the higher education sectors. It was primarily used as benchmarks for the accreditation of courses by non-university higher education providers. At the same time as the release of the Bradley Report a new Australian Qualifications Framework was finalised, following a separate review by John Dawkins, the architect of the current sector in Australia, overseen by the Australian Qualifications Framework Council.

The most recent version of the Australian Qualifications Framework (Australian Qualifications Framework Council, 2013) covers all levels of education from school to vocational and higher education. Its objectives include providing 'a contemporary and flexible qualifications framework that …enables the alignment of the AQF with international qualifications frameworks' such as the Bologna related Tuning Profiles, and those of the Degree Qualifications Profile of the Lumina Foundation, though it differs from these in significant ways.

The AQF specifies three broad learning outcomes for each course level– Knowledge (defined as what students know); skills (what students can do) and application of knowledge and skills (interpreted as the context in which the student may need to work). The *criteria* for the learning outcomes (called learning outcomes themselves in the Tuning and DQP examples) contain detail of the type of knowledge, range of generic and personal skills expected to be developed and demonstrated through the

application of knowledge and skills during the course. The AQF is not discipline based nor does it classify types of knowledge or skills into a tiered taxonomy.

The AQF learning outcomes criteria are very general. For example, the summary criterion for an ordinary bachelor degree (Level 7) qualification is 'Graduates at this level will have broad and coherent knowledge and skills for professional work and/or further learning' (p13). While more detailed criteria support this general statement they are themselves broad and insufficient for comparison with the schemes used in other countries. This type of detail is found in the work on graduate attributes and learning outcomes commissioned through the Australian Learning and Teaching Council.

3.10 What this review has shown

Appendix 3.1 summarises the key features of the various approaches outlined above which have been taken to the specification of learning outcomes and associated standards and compares their main characteristics. The origins of these approaches differ with some arising from accreditation and credit transfer sources and others from a desire to improve teaching and learning quality and this influences the nature of their findings. Each of the schemes has a multi-level hierarchical taxonomy for the classification of the outcomes and the standards or rubrics prescribed. At the second tier of these hierarchies, there is apparent commonality in the use of terminology but the detailed expression of the learning outcomes show that the meanings of the terms differ between the schemes. Most of the taxonomies discussed above do not include standards as defined earlier or learning outcomes which readily lend themselves to assessment. The schemes are idiosyncratic to their jurisdictions and the scope of undergraduate education addressed, and it would not be a good fit to translate one of these options directly into the Australian higher education environment and its model of undergraduate education. That said, of the approaches considered, the VALUE Rubrics matrix seems closest to how a set of learning outcomes and standards for Australian undergraduates might possibly be specified given the objectives of this study, because of the multiple levels of achievement it spells out and its greater focus on assessability.

This review also shows that at present Australia is not as far advanced as other countries in formulating learning outcomes for higher education in a generalised and structured way for the whole sector and in engaging a range of the relevant stakeholders in the formulation of a hierarchy and taxonomy for learning. This is not for lack of endeavour

or lack of government funding but rather due to a different focus and a resistance to the concept of generalisable learning outcomes. Australia has not to date adopted a nation- or state-wide set of graduate attributes or learning outcomes as there has been for the Tuning process, LEAP or the DQP in the US and Europe. While some detailed work on generalised learning outcomes and standards for a range of disciplines has been undertaken under the auspices of the Australian Learning and Teaching Council, this material has not yet been synthesised or used across disciplines or for the sector as a whole.

What this international review does show is that it has been possible through various methodologies to specify a comprehensive set of generalised learning outcomes for different course study levels at a sector level within an educational jurisdiction. This finding addresses to some degree the question: What is it that students should learn in their study for a higher education degree? And yet in spite of all of these attempts developed independently over 20 years a single model of outcomes, standards, assessment and measurement remains elusive and in fact nothing done to date goes as far as the aim of this study – that is, trying to use local individual assessment results in an accumulated or quantitative form to measure institutional performance in student achievement of learning outcomes. All of the approaches summarised above assume that students will be able to acquire the learning outcomes described in their particular system but to date there is little evidence of success in measurement of whether this is the case.

In determining assessment approaches it is also important to remember Shavelson's conceptual framework which emphasises the recursive relationship between generic skills and basic knowledge and how this is built up to increase intellectual capacity and depth of learning over time. This raises the inter-related nature of knowledge based learning outcomes and generic skills which are applicable to that learning. This concept and later research (Jones, 2006), which has suggested that generic skills should be assessed in the context of the knowledge of the discipline, is pertinent to the formulation of new learning outcome statements and standards and more importantly to the nature of appropriate assessment which might be used to evaluate individual and institutional success across the higher education sector as a whole.

At a practical level, suggestions by Otter and more recently by Shavelson that learning outcomes should be based on a concept of what it means to be a university graduate,

and the use of a staged process for the specification of learning outcomes and their assessment through standards, provide the basis of a rigorous approach to the development of a framework which could apply to Australian undergraduate education. Otter's use of well-known educational taxonomies such as those of Bloom and Carter (Bloom, 1956; Carter, 1985) for classifying levels of learning and understanding assists in describing appropriate and assessable outcomes, and puts educational learning theory at the heart of what is necessarily a pragmatic approach to the alignment of educational outcomes and assessment.

As indicated earlier, the capacity to frame descriptions of learning outcomes in a way that enables feasibility of assessment and performance evaluation of institutions has been identified as the next big challenge for higher education teaching and learning by the Australian government and several of the groups behind the development of learning outcomes in other countries. This challenge is now an important component of their research agendas and a critical issue for this study with its similar aim.

For a conceptual framework for the measurement of learning outcomes to work in Australia the outcomes need to be appropriate to the goals and aspirations of the universities, students, employers of graduates, and the Government stakeholders in this higher education system. Hence progressing this study for the Australian higher education environment will require a mixture of a theoretical position and a degree of pragmatism about what the set of learning outcomes should look like for this environment, and what sort of standards or rubrics will provide meaningful information about levels of achievement.

Chapter 4 - A generalisable set of undergraduate learning outcomes for the Australian sector

4.1 A methodology and a conceptual framework

The approach used below to develop Australian undergraduate learning outcomes and standards, which draws on Otter's method, consists of the following steps:

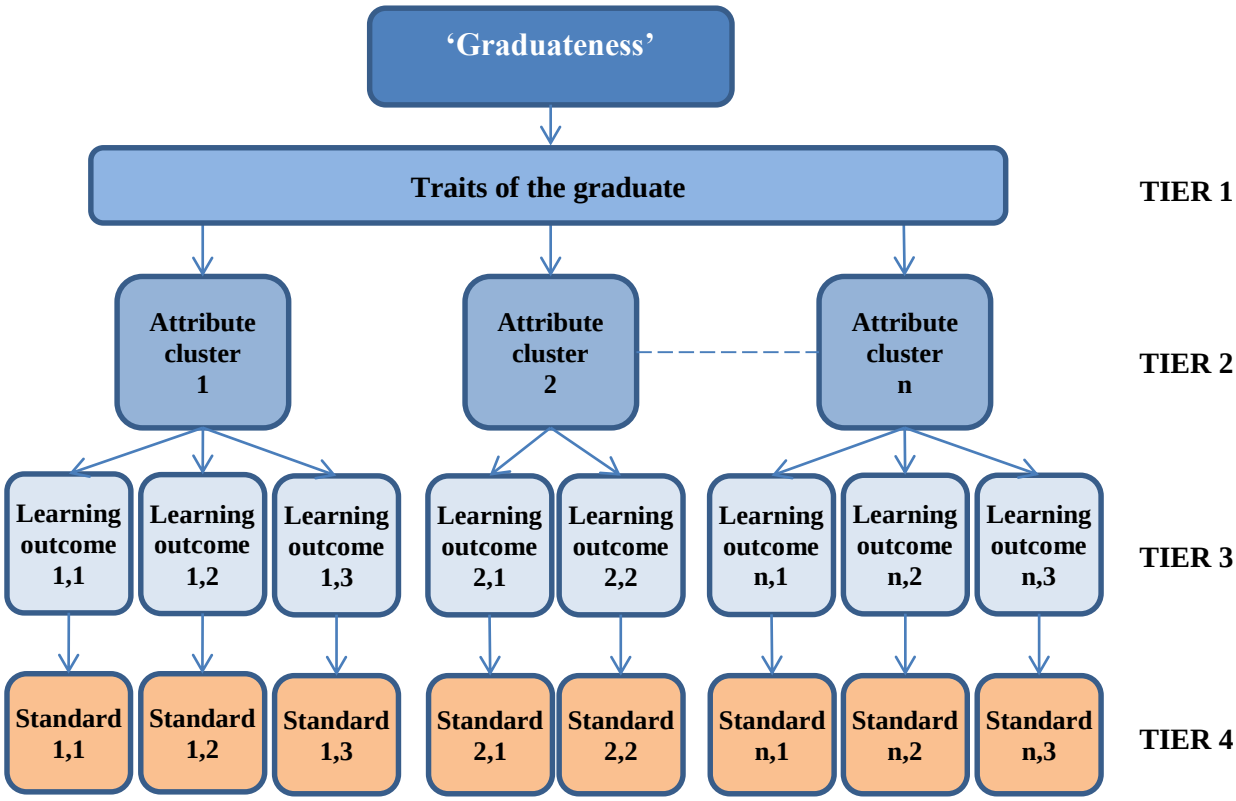
- Describe what it means to be a graduate of an undergraduate degree course in Australia (i.e. what 'graduateness' means);
- Identify a set of traits from this description;
- Analyse the statements of graduate attributes formulated by Australian universities for undergraduate degrees, and group them into clusters based on these traits;
- Analyse findings on standards and learning outcomes by disciplines in the ALTC projects and align these types of descriptors with the clusters of graduate attributes;
- Formulate generalised statements of learning outcomes to apply across all disciplines; and
- Devise a set of standards for each learning outcome, which define levels of achievement of the outcomes, similar in style to the rubrics of the VALUE system.

Such an approach involves both 'top down' specification of the traits of a graduate and 'bottom up' identification of graduate attributes by the universities, and learning outcomes and standards from the various ALTC projects. Describing community expectations of a graduate from a layperson's perspective is an attempt to think about the Australian system afresh and develop a conceptual framework for learning outcomes from this base. The cornerstone of this approach is to think about the characteristics that the community and employers might expect graduates of universities to have. An attempt is made in this study to develop a description of the traits a graduate might ideally develop during their period of study – this is described as a state of *graduateness*. Several universities are now using this approach (e.g. The University of Melbourne in general, and The University of Sydney at a discipline level) rather than specifying a list of general graduate attributes which is frequently the same as those of other universities. This potentially differentiates each university from their competitor

institutions in terms of their graduates' individual skills and characteristics, particularly those gained though their educational experience**.**

A four-tier conceptual framework, as shown in Figure 4.1, for the specification of learning outcomes and standards is proposed. The approach described above aligns well with this and conforms to the definitions of learning outcomes and standards described earlier. The criteria for assessing the learning outcome schemes are also applied to ensure that those in Tier 3 of the conceptual framework satisfy the conditions for a superior set of outcomes.

Figure 4.1 – A proposed conceptual framework for Australian undergraduate learning outcomes and standards



The set of traits at Tier 1 derived from the statement of *graduateness* includes personal qualities, as well as learned skills and knowledge. These are the basis of the classification and development of the Tier 2 clusters of graduate attributes, which are summary descriptors of the actual graduate attributes of all Australian universities for undergraduate programs.

The Tier 3 learning outcomes associated with each Tier 2 attribute cluster summarise the range of knowledge or skills identified, and are based on findings of the ALTC

standards projects for a range of ten disciplines. Tier 4 of the framework represents the standards, using the earlier definition, which provide the benchmarks against which achievement of the learning outcomes might be assessed. There are several standards statements associated with each Learning Outcome. This is necessary to differentiate between individual students' achievement and for use in any quantitative performance indicators which might measure institutional performance.

4.2 The traits of a bachelor degree graduate – the idea of 'graduateness'

An Australian university graduate is expected to have a range of knowledge, skills and personal characteristics that distinguishes them from others in the population with lower levels of education.

When thinking about what it means to be an Australian graduate, the first characteristic that usually comes to mind is that they should have an **in-depth knowledge** of their chosen field of study in at least one discipline. They also should have a set of skills which enable them to **apply this knowledge** to examine issues in other disciplines or in society more generally. It is expected that a graduate will be able to engage in **scholarship or enquiry** in their discipline and apply these **research skills** more broadly to solve problems in future employment or further study. They should be **good communicators** both orally and in writing and be able to explain complex issues to a variety of different audiences. They should be able to **work in collaboration** with other people, often taking a leadership role in teams of their student peers and future work colleagues, but also be **able to work alone** and be a self-starter.

Some of the personal characteristics that might also be expected are that the graduate should be **intelligent, ambitious, hard-working**, disciplined and confident without being arrogant; have a strong **ethical framework** through which they operate professionally and personally; and be **reflective about their own experiences** and their role in society. They should be **critical thinkers** and able to use their knowledge and skills to **analyse complex situations** and devise **innovative** but logical solutions to problems which are based on data and analysis. At a practical level they should be able to use the knowledge and skills learned during their studies **productively in the workforce** and in the solution of problems more generally in the community. By the time they graduate they should have acquired a broad **understanding of different cultures** and groups in society, a strong **sense of fairness**, an understanding of their

privileged and advantaged position in society and a good understanding of **Indigenous culture, history and values** so that they can work and interact effectively with these different groups who are all part of multi-cultural Australia.

With the increasing importance of the global economy and the need for graduates to be able to operate across this economy, there is a need for them to have knowledge and an **understanding of how to operate in different economic and social environments**. From their previous learning experiences, they should be interested and motivated to **continue to learn** and motivated to regularly **update their skills** and to build a career in the longer term after graduation. They should have the capacity to become **future leaders** in their communities, particularly in relation to **social change and environmental sustainability**.

If the above statement is analysed to pick out key words (shown in bold in the above paragraphs), the graduate characteristics relate to the following twelve Tier 1 traits of a proposed conceptual framework to which graduate attributes will be linked:

- Knowledge of discipline;
- Research and scholarship;
- Communication;
- Collaboration;
- Self-awareness and self-discipline;
- Ethics;
- Thinking and analysis;
- Equity/social justice;
- Civic and social understanding;
- Leadership, and skills and their application in employment;
- Global understanding; and
- Social and environmental sustainability.

4.3 Clusters of graduate attributes

The next step is to explore work already undertaken on specifying and clustering graduate attributes. There is considerable public information available about graduate attributes as they have been specified by Australian universities. These graduate attributes are general aspirational statements and are commonly not specific to level of course or discipline. For example at present only 8 of the 39 universities in Australia

have sets of graduate attributes for different levels of course - and these are usually at two levels only (undergraduate and postgraduate) - and only 10 of them have discipline-based attributes created to align with a parent generic set (based on an analysis of websites of universities, August 2013).

Some universities have linked their own attributes back to aspects of their curriculum through a mapping process but few have enunciated a set of learning outcomes corresponding to their general graduate attributes, and even fewer appear to have considered the nature of assessment required to measure whether students achieve these attributes or not.

4.3.1 National Graduate Attributes Project

Following the recommendations of the report on the Review of Australian Higher Education (Bradley et al., 2008) , the Australian Learning and Teaching Council funded a National Graduate Attributes Project (GAP) which compiled and published a database of these attributes for all Australian universities (Barrie, 2009; Barrie et al., 2009). The project retrieved statements of graduate attributes in 2008 from all Australian universities' websites and grouped them into a taxonomy, *Conceptions of Generic Attributes* (COGA), a framework for generic skills which was derived from previous work (Barrie, 2007). This approach excluded graduate attributes related to the acquisition of knowledge and focused on generic skills.

The COGA taxonomy used by Barrie consisted of two different types of attributes: Enabling and Translation. The first of these covered scholarship, global citizenship and lifelong learning and the latter research and enquiry, information literacy, personal and intellectual autonomy, ethical and social understanding and communication. Consistent with Shavelson's view, Barrie makes the point that 'the boundaries between these categories are artificial constructs and some university attributes are related to more than one category' (Barrie et al., 2009) (p1) but he reported that all of the attributes at that time for Australian universities could be allocated to these groups. This appears to be rather at odds with current graduate attributes for Australia where subject-knowledge, collaboration, communication, employment skills, social justice and sustainability are among the most frequently mentioned graduate attributes in 2013. These are missing from the COGA structure.

The GAP project did show that there is considerable similarity between the attributes stated by the universities with many of the broad statements used being common to over 30 of the 39 universities for which data was collected. While the report was the first comprehensive analysis of Australian graduate attributes it went no further with any analysis of the attributes themselves beyond stating them and counting their frequency within each category. This study updates this type of analysis to reflect the current graduate attributes on university websites and places the findings into the proposed conceptual framework.

Oliver (2011a) summarised the findings of all of the 54 ALTC projects funded on assuring graduate outcomes and she comments on the inter-relationship between discipline outcomes, generic skills outcomes and associated professional practice. She also conducted a review of learning outcomes in Australia by accessing policy statements and statements of graduate attributes. Apart from knowledge or subject-based outcomes, she found that Australian graduate attributes were all clustered in seven broad areas. These groups differ from Barrie's with less emphasis on research and inquiry and stronger emphasis on the specific skills and teamwork identified as missing from Barrie's set. Oliver is interested in quality assurance rather than performance measures per se and she points out that few universities specify levels of achievement in their statements of graduate attributes. She also observes that universities 'should be more specific about which of the generic outcomes are 'developed' or 'fostered' and which are assessed (or 'warranted'). This is an important point to resolve, if the learning outcomes are to be reliably assessed and used in cross-institutional comparisons of teaching and learning performance. Oliver recommended that future work on graduate outcomes 'should focus on assuring standards and investigating ways of providing evidence of the achievement of standards in generic outcomes'. These issues are covered in the analysis of learning outcomes and their assessment which follows.

4.3.2 Analysis of graduate attributes for undergraduate degrees in Australian universities

Given the differences between these earlier analyses, the extraction and classification of undergraduate graduate attributes was repeated in this study (as at August 2013). A search of all Australian universities' websites was undertaken to identify their undergraduate degree graduate attributes. All attributes found were recorded in a data

base with details about whether the attributes were general for the institution, specific to a bachelor degree, or were discipline specific. Those relevant to undergraduate study were then grouped by the Tier 1 Traits of the conceptual framework and set of finer attribute clusters (Tier 2 level) was identified based on the similarity of key words in the long-form descriptions of the attributes. The number of attributes and the detail included in their description varies greatly across universities as shown in the summary of each institution's data in Appendix 4.1. The 12 Tier 1 Traits and the 25 Tier 2 Attribute Clusters derived in this way and their concordance are shown in the following table, including the percentage the attributes represent of the total group identified.

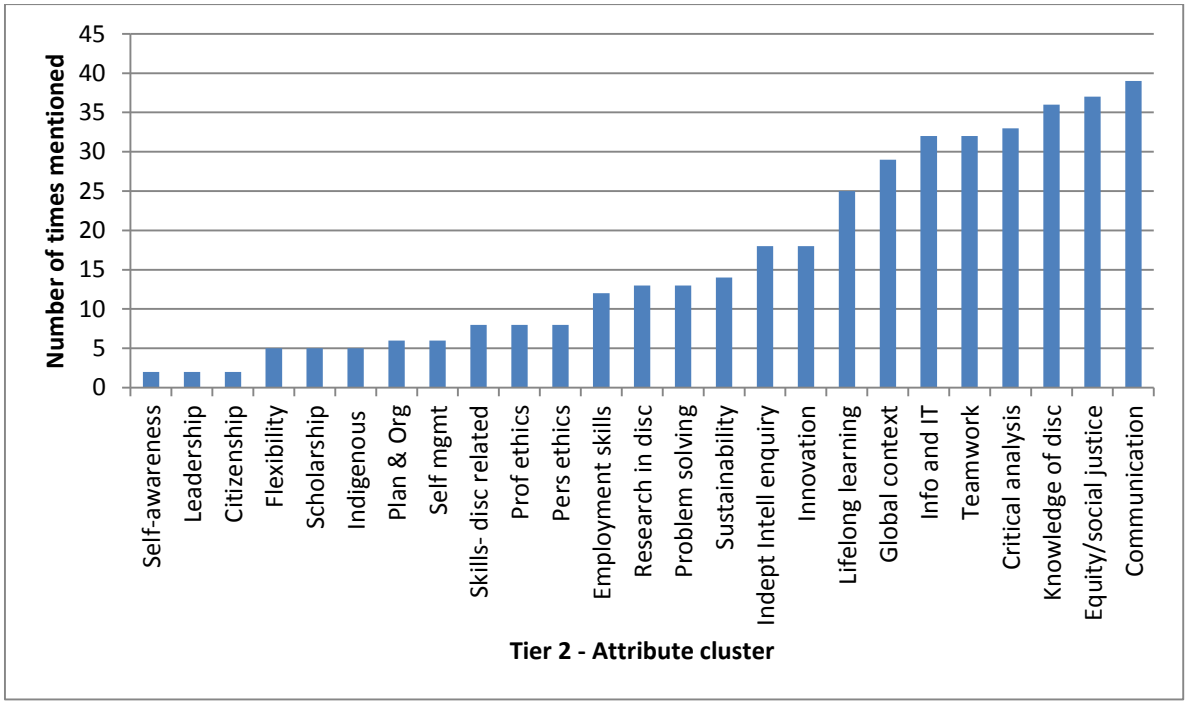
Table 4.1 – Tier 1 and Tier 2 elements of the model derived from graduate attribute statements

Tier 1 Traits (% of all attributes)	Tier 2 Attribute Clusters
Civic and social understanding (1%)	Citizenship
Collaboration (9%)	Teamwork Leadership
Communication (10%)	Communication (oral, written, graphical)
Equity/social justice (10%)	Equity/social justice Indigenous cultural understanding
Ethics (4%)	Professional ethics related to discipline Personal ethics
Global understanding (7%)	Global context
Knowledge of discipline (12%)	Discipline knowledge Discipline –related skills
Research and scholarship (9%)	Scholarship Research in discipline Innovation
Self-awareness and self-discipline (13%)	Independent intellectual enquiry Lifelong learning Self-management Self-awareness
Thinking and analysis (11%)	Critical analysis Problem solving
Skills and their application in employment (11%)	Information and IT skills Planning and organisation Employment skills Flexibility
Social and environmental sustainability (3%)	Sustainability

Figure 4.2 shows the frequency with which each of these Tier 2 attribute clusters is mentioned across the university sector and highlights the most common clusters of attributes derived from the current analysis.

Many universities have more than one attribute per cluster, and there are 8 categories above which are mentioned by at least 15 of the 39 universities in Australia. The most commonly mentioned attribute clusters are Communication (39/39 universities); Equity/Social justice (37); Critical analysis (33); Knowledge of discipline(s) (36); Teamwork (32); Information and IT (32); Global context (29); and Lifelong learning (25). If these eight attribute clusters were used as the basis of describing learning outcomes in the proposed framework they would cover 64% of all learning outcomes specified for the sector and would be an important subset to focus on for assessment and subsequent measurement purposes.

Figure 4.2 – Distribution of Australian graduate attributes by Tier 2 Attribute clusters, 2013



These 25 Attribute Clusters form Tier 2 of the conceptual framework for defining the learning outcomes. The next step on the path to outcome specification and assessment is to describe a set of action statements for learning for each of these undergraduate study attribute clusters in the style recommended in the various taxonomies described earlier.

These statements become the Tier 3 learning outcomes in the framework and are in a form which might allow a judgment to be made on potential for assessability.

4.4 Tier 3 – Learning outcomes

4.4.1 The Australian Learning and Teaching Council standards project

The formulation of these learning outcomes statements is supported by work done in the ALTC Standards project in 2010-11. The goal of the ALTC project was to define *threshold* learning outcomes for bachelors or masters level courses as defined in the AQF across a range of disciplines (Australian Learning and Teaching Council, 2011a, 2011c). A total of 10 disciplines participated in the project:

- arts, social sciences and humanities;
- business, management and economics;
- creative and performing arts;
- engineering and ICT;
- health, medicine, and veterinary science;
- law;
- education;
- building and construction;
- architecture; and
- science.

The working definition of 'academic standards' used by the ALTC project was 'learning outcomes described in terms of discipline-specific knowledge, skills and capabilities expressed as threshold learning outcomes that a graduate of any given discipline (or program) must have achieved'. This definition is consistent with the one stated in Section 3.2.1. Therefore the ALTC project appeared to be more about defining a set of threshold learning outcomes than a complete set of 'standards' as defined at the start of this chapter. The conceptual framework shown in Figure 4.1 makes a clear distinction between the standards (Tier 4), which are hierarchical children of the learning outcomes, and the learning outcomes themselves (Tier 3). Hence in the current study's terminology, the learning outcomes from the ALTC projects are closer to the Tier 2 attributes clusters and the ALTC Standards, by virtue of their definition, actually resemble the Tier 3 learning outcomes of the framework. They are also similar to the

type of detailed learning outcomes described in the Tuning process, the HEQCO Competences and the DQP.

Based on the 10 discipline-based reports (Australian Learning and Teaching Council, 2010b, 2011b), it appears the AQF has not been used as a guide to the development of the standards in the ALTC project, nor was there any examination of the Graduate Attributes Project when describing the learning outcomes. Instead ALTC discipline panels refer more commonly to accreditation criteria in professional disciplines and the Tuning process work. Many of the generic skills included in the Graduate Attributes project listings are not mentioned in the ALTC work but are included in the standards statements as the skills necessary to demonstrate the outcomes. The learning outcomes were derived by each of the disciplinary standards panels independently, and the generic skills are used to demonstrate achievement of knowledge or application of skills in professional practice.

The standards statements give more insight into what each discipline requires for a student to be able to demonstrate achievement of the outcome and they generally show considerable commonality across disciplines as shown in Appendix 4.2. There were only two cases where there were more specific statements for the *Application of skills* learning outcome (Engineering and ICT, and Science). For these disciplines there was a tendency to describe very specific standards statements relating to mathematical modeling, types of engineering applications and particular scientific methods, but they could be covered by a general statement about synthesis of theoretical knowledge and skills within the discipline, in the same way as for other disciplines. In fact, it appears that the common features for the disciplines can be covered in generalisable statements of threshold standards by using the term ‘discipline’ rather than the specific mention of the particular area of study. The results of this analysis of the statements and their generalisable form are shown below.

Table 4.3 – Generalisable ALTC learning outcomes and standards for Australian bachelor degrees

ALTC Learning outcome	ALTC standards
Knowledge	• Demonstrate an understanding of one or more major disciplines as relevant to the field of study • Demonstrate knowledge of practices, languages and relevant technologies for the discipline
Ethics and professional responsibility	• Demonstrate an understanding of approaches to ethical decision making and their relevance to the discipline • Recognise and reflect on social, cultural and ethical issues generally and in relation to professional practice
Thinking skills	• Engage in critical analysis to address problems and exercise reasoned choice among alternative solutions • Think creatively to address issues related to the discipline • Make decisions using evidence-based reasoned argument • Use innovative approaches to solve problems • Recognise, evaluate and synthesise various views, arguments and sources to solve problems in the discipline
Research skills	• Interpret, apply or conduct research related to the discipline • Evaluate emergent knowledge in the discipline • Design and plan an investigation in a subject relevant to the discipline
Communication, collaboration and leadership	• Communicate perspectives and knowledge effectively and appropriately to a range of different audiences using written, oral and visual means • Work effectively , responsibly and constructively as a member of a team
Self-management	• Learn and work independently • Reflect on own performance and demonstrate a high level of critical self-knowledge • Show commitment to ongoing learning for personal and professional development
Application of skills	• Synthesise theoretical understanding and practical skills appropriate to the discipline to solve problems and enhance knowledge

While this set of learning outcomes and threshold standards is representative of the ten discipline-based sets of statements from the ALTC standards project, it is not entirely aligned with the types of graduate attributes currently specified by Australian universities (Barrie, 2009). Some are clustered together (for example *Ethics and*

professional responsibility, and *Communication, collaboration and leadership*) whereas these are typically described separately in statements of graduate attributes. If there is a focus on assessability of the outcomes, the elements of these groups may require multiple types of testing to ascertain levels of achievement. The *Ethics and professional responsibility* outcome covers equity, cultural and global perspective aspects. From the policy perspective of individual universities and the Government in each of these components, there is an interest in being able to define and assess specific learning outcomes in each of these areas rather than as a combined element. The absence of outcomes on civic and cultural knowledge is also a gap compared with the previous graduate attribute analysis.

The above synthesis suggests that it may be possible to generalise threshold learning outcomes across all disciplines for bachelor and masters degree levels in the AQF, but more work would need to be done on a range of usable standards for levels of achievement of the learning outcomes, to enable differences in individual and institutional performance to be measured.

4.4.2 A proposed set of new generalisable learning outcomes and associated standards for assessment in Australia

In spite of the 'learning outcomes' derived from the ALTC project not being based on graduate attribute data, they are consistent with a subset of the listed Tier 2 clusters. The Tier 3 learning outcomes of this study's framework have been populated by referencing the generalised ALTC threshold 'standards' statements, simplifying them where possible, aligning them with the Tier 2 attribute clusters, and filling any gaps. A set of standards using the earlier definition in Chapter 3 is then framed with multiple achievement levels in a way that might be conducive to reliable assessment at the individual student level. The format for development of the statements of learning outcomes and standards shown below is modeled on that used in the VALUE rubrics approach, but the learning outcomes are derived as described above, and are considerably fewer in number (38 compared with the 88 of VALUE).

These learning outcomes and their parent Attribute Clusters are shown in Table 4.4. Care was taken in framing these statements to use active verbs which lend themselves to translation as assessment tasks. The meaning of the verbs used are consistent with the application of the Bloom Taxonomy for the cognitive domain. Some helpful examples of how this has been done are available (Otter, p23), (Eber & Parker, 2008) (pp51-53).

Table 4.4 – Tier 3 learning outcomes for Australian undergraduate degrees

Tier 2 Attribute clusters	Tier 3 Learning outcomes
Discipline knowledge	Understand, describe and apply theories and information relating to one or more disciplines
Discipline –related skills	Understand and apply skills related to the discipline to professional practice
Scholarship	Understand the theoretical basis of the discipline and apply its principles to professional practice
Research in discipline	Understand, create and evaluate emergent knowledge in the discipline
Innovation	Create innovative solutions to problems Understand the need for change and apply knowledge and skills to bring it about
Teamwork	Learn collaboratively Participate as a member of a team to achieve joint goals and contribute to the team's outcomes
Leadership	Lead, plan and play the main role in a team project
Communication (oral, written, graphical)	Communicate orally in English Write clearly, coherently and creatively appropriate to audience needs Generate, analyse and communicate numerical information in ways appropriate to a discipline
Equity/social justice	Demonstrate respect for dignity of others and for human diversity Recognise and respect the role of cultural difference and diversity in work and social contexts
Indigenous cultural understanding	Understand Indigenous Australian issues and cultures
Professional ethics related to discipline	Understand and demonstrate professional ethical responsibilities
Personal ethics	Understand ethical issues and apply ethical principles in complex situations
Independent intellectual enquiry	Learn independently in a self-directed manner Apply initiative in setting goals and completing learning tasks
Lifelong learning	Analyse own intellectual capabilities and create opportunities for own professional development
Self-management	Demonstrate initiative and responsibility Manage own time and meet deadlines for learning tasks Demonstrate openness, intellectual humility, spirit of enquiry
Self-awareness	Evaluate learning and performance In tasks Apply feedback on own performance and learning for improvement
Critical analysis	Demonstrate ability to think critically, analyse and evaluate claims, evidence and arguments and evaluate evidence clearly and logically
Problem solving	Apply problem solving processes in novel situations

Tier 2 Attribute clusters	Tier 3 Learning outcomes
Information and IT skills	Identify, organise and evaluate information with emphasis on primary sources Apply information and communication and other relevant technologies to solve problems and in learning tasks
Planning and organisation	Apply sound planning and organisational skills in learning tasks
Employment skills	Demonstrate knowledge of regulatory frameworks and political influences for professional area of study Demonstrate capability to solve real world problems by applying learning and skills for discipline(s)
Flexibility	Apply different learning approaches to suit different tasks Adapt to change and develop personal responses to manage change
Citizenship	Understand what it means to be a well-informed citizen and apply knowledge and skills to contribute to own communities
Sustainability	Understand financial, social and environmental sustainability
Global context	Develop and apply international perspectives in their discipline Demonstrate understanding and acceptance of culturally diverse and international environments

4.5 Tier 4 - Standards

The next step towards measurement of achievement of these proposed Tier 3 learning outcomes for Australian undergraduate degree education is to specify a set of standards or rubrics to enable assessment of individual attainment. These standards represent level 4 of the conceptual framework shown in Figure 4.1. For each learning outcome a standards statement consisting of a criterion and a set of levels of achievement has been developed.

Four levels of achievement for each learning outcome are included in the standards statement - Not achieved; Developing; Accomplished; and Exemplary. There is no magic about choosing the number 4 and a larger number of standards levels creates more options for differentiating student achievement, although the differences between the levels of achievement is sometimes subtle. The choice of the number of standards levels is pragmatic and depends on how the grades given may be used to summarise assessment attainment. For example, in Finland 5 levels are used for similar statements of standards (Ursin, 2014). The number of categories chosen may be important later, where distribution of student achievement should be dispersed enough to allow a range of values to be used in institutional performance calculations relating to the achievement of learning outcomes. For example if only three groups were to be used, with one of these being non-achievement of the outcomes, students' achievement ratings would be

either a score of 1 or 2 representing achieved or exemplary. When aggregated and averaged in some way such a scoring scheme is not likely to provide very great differentiation between institutions. However the downside of a higher the number of categories is a more complex grading system, raising issues about consistency of approach and reliability between markers.

The Tier 4 standards are linked to the above proposed Tier 3 learning outcomes as shown in Figure 4.1. The development of the criterion-based standards statements has been a considerable challenge. Currently the most common form of assessment used internally in universities is standardised tests, which rate student achievement in terms of their performance to their peers. It is difficult to describe standards for different levels of achievement in non-normative terms. But this challenge should be addressed if the assessment is to be truly standards- or criterion-based. If what is actually learned and practised by students is to be measured using the standards, the standard levels of achievement must reflect students' actual understanding and capability and not a deduced rating based on normative practices.

The proposed standards associated with the Tier 3 learning outcomes which attempt to address these challenges are shown in Table 4.5. They are presented in a similar format to the VALUE rubrics.

These learning outcomes and standards for the Australian undergraduate degree level which are based on the conceptual framework shown in Figure 4.1 represent a proposal for consideration as the first step in attempting to find a viable solution for defining and measuring learning outcomes based on what individual students learn in their degree studies.

Table 4.5 – Proposed Learning Outcomes and standards for Australian bachelor degrees by Tier 1 Trait and Tier 2 Attribute Cluster

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Trait 1: Knowledge of discipline					
Discipline knowledge	Understand, describe and apply theories and information relating to one or more disciplines	Unable to describe and apply theories and information relevant to a problem in the discipline	Demonstrates developing understanding of how theories work and information can be used to solve a problem in a discipline	Demonstrates a sound basic understanding of how theories and information can be used to solve a problem in the discipline	Demonstrates comprehensive understanding of theories and their application and provides a well-argued solution to a problem in the discipline
Discipline – related skills	Understand and apply skills related to the discipline to professional practice	Unable to develop or demonstrate skills related to the discipline	Demonstrates developing capability in skills and their application in an area of professional practice	Demonstrates accomplishment in skills and capacity to apply them to an area of professional practice	Demonstrates excellence skills and their application to professional practice in the discipline
Trait 2: Research and Scholarship					
Scholarship	Understand the theoretical basis of the discipline and apply its principles to professional practice	Unable to understand the theoretical basis of the discipline nor able to apply principles to professional practice	Demonstrates developing understanding of the theoretical basis of the discipline and has applied it appropriately to a problem from professional practice	Demonstrates an accomplished understanding of the theoretical basis of the discipline and has applied it successfully to a problem from professional practice	Demonstrates excellent understanding of the theoretical basis of the discipline and has applied it in a transformative way to a problem from professional practice

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Research in discipline	Understand, create and evaluate emergent knowledge in the discipline	Unable to understand emergent knowledge in the discipline	Demonstrates developing knowledge of new ideas and their understanding of their impact on the discipline	Demonstrates accomplishment of knowledge of new ideas and understanding of their impact on the discipline	Demonstrates excellence in knowledge of new ideas and understanding of their impact on the discipline
Innovation	Create innovative solutions to problems	Unable to solve problems except in a previously demonstrated way	Demonstrates a developing ability to look at a problem in a new way and devise a solution based on this approach	Demonstrates accomplishment in looking at problems in a new way and devising solutions based on this approach	Demonstrates excellence in developing innovative solutions to problems.
	Understand the need for change and apply knowledge and skills to bring it about	Resistant to new approaches to addressing problems in the discipline	Demonstrates a developing understanding of new ways of addressing problems in the discipline	Demonstrates an accomplishment in understanding of new ways of addressing problems in the discipline	Demonstrates excellence in addressing complex problems using knowledge and skills to solve them in the discipline
Trait 3: Collaboration					
Teamwork	Learn collaboratively	Does not contribute to work as a member of a team undertaking a task	Demonstrates a developing capacity to work in a team undertaking a task	Demonstrates an accomplished capacity to work in a team undertaking a joint task	Demonstrates an excellent capacity to contribute to the work in a team undertaking a joint task

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Teamwork	Participate as a member of a team to achieve joint goals and contribute to the team's outcomes	Does not work constructively or participate effectively in work or discussions in a team	Demonstrates a capability to contribute actively to work and discussions of a team to achieve goals	Demonstrates accomplished and active contributions to work and discussions of a team to achieve goals	Demonstrates excellent and active contributions to work and discussions of a team and shows leadership of the team in achieving goals
Leadership	Lead, plan and play the main role in a team project	Does not show any leadership in a team project	Demonstrates a developing capacity to lead and plan a project and to play a main role	Demonstrates accomplished planning and leadership of a team project	Demonstrates excellent planning and leadership of a team engaging members and facilitating their contributions to maximise outcomes for the team project
Trait 4 : Communication					
Communicat-ion (oral, written, graphical)	Communicate orally in English	Cannot convey ideas orally to different groups in a presentation	Demonstrates a developing capacity to convey ideas orally to different groups in a presentation	Demonstrates accomplishment in conveying ideas orally to different groups in a presentation	Demonstrates excellence in convey ideas orally to different groups in a presentation clearly explaining concepts and ideas

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Communication - written	Write clearly, coherently and creatively appropriate to audience needs	Cannot convey ideas clearly in writing to various audiences	Demonstrates a developing capacity to write coherently and creatively for various audiences	Demonstrates accomplishment in writing clearly and creatively for various audiences	Demonstrates excellence in writing clearly and creatively for various audiences
Communication - graphical	Generate, analyse and communicate numerical information in ways appropriate to a discipline	Cannot interpret discipline relevant numerical information adequately	Demonstrates a developing capacity to interpret numerical information adequately for the discipline	Demonstrates a developing capacity to interpret numerical information adequately for the discipline	Demonstrates excellence in calculation and interpretation of numerical information to enhance arguments by its use in the discipline
Trait 5 :Equity/social justice					
Equity/social justice	Demonstrate respect for dignity of others and for human diversity	Does not understand the concept of human diversity or the need for respect of others	Demonstrates developing respect for dignity of others and for human diversity in addressing issues in the discipline	Demonstrates basic respect for dignity of others and for human diversity in addressing issues in the discipline	Demonstrates complex understanding of human diversity and respect for the dignity of others in addressing issues in the discipline
	Recognise and respect the role of cultural difference and diversity in work and social contexts	Does not understand the role of cultural difference and its importance to work in the discipline	Demonstrates developing understanding of cultural difference and the importance of it to work in the discipline	Demonstrates accomplished understanding of cultural difference and the importance of it to work in the discipline	Demonstrates excellent understanding of cultural difference and is able to adapt work response to take account of this diversity

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Indigenous cultural understanding	Understand Indigenous Australian issues and cultures	Does not understand Indigenous issues and cultures and relevance to professional practice	Demonstrates developing understanding of Indigenous issues and cultures and their importance to professional practice	Demonstrates accomplished understanding of Indigenous issues and cultures and their importance to professional practice	Demonstrates excellent understanding of Indigenous issues and cultures and their importance to professional practice
Trait 6 : Ethics					
Professional ethics related to discipline	Understand and demonstrate professional ethical responsibilities	Does not understand and/or is unable to demonstrate professional ethical practice	Demonstrates developing understanding of ethical values and responsibilities in professional practice	Demonstrates accomplished understanding of ethical responsibilities in professional practice	Demonstrates excellent understanding of ethical responsibilities in professional practice
Personal ethics	Understand ethical issues and apply ethical principles in complex situations	Does not understand ethical issues and/or apply ethical principles in complex situations	Demonstrates developing capacity to understand ethical issues and apply ethical principles appropriately to practice in complex situations	Demonstrates accomplishment in understanding ethical issues and applying ethical principles in complex situations	Demonstrates excellent understanding of ethical issues and capacity to identify and apply ethical principles in complex situations

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Trait 7 : Self-awareness and self-discipline					
Independent intellectual enquiry	Learn independently in a self-directed manner	Not able to work independently and/or plan own learning approach in the discipline	Demonstrates developing capacity to work independently and to plan own learning approach to the discipline	Demonstrates accomplishment in working independently and planning own learning approach to the discipline	Demonstrates excellence in working independently and demonstrates sophisticated planning of own learning approach
	Apply initiative in setting goals and completing learning tasks	Not able to demonstrate initiative in goal setting or satisfactorily completing set learning tasks	Demonstrates developing capacity to apply initiative in goal setting and satisfactorily completion of set learning tasks	Demonstrates accomplishment in applying initiative in goal setting and satisfactorily completing set learning tasks	Demonstrates excellent initiative in goal setting and satisfactorily completion of set learning tasks
Lifelong learning	Analyse own intellectual capabilities and create opportunities for own professional development	Unable to analyse own intellectual capabilities and does not understand the need for ongoing learning	Demonstrates developing capacity to analyse own intellectual capabilities and understand the need for ongoing learning	Demonstrates accomplishment in analysing own intellectual capabilities and understanding the need for ongoing learning	Demonstrates excellence in analysing own intellectual capabilities and understanding the need for ongoing learning
Self-management	Demonstrate initiative and responsibility	Does not demonstrate initiative and/or responsibility in learning tasks	Demonstrates developing initiative in learning tasks and takes responsibility for own learning	Demonstrates accomplishment in taking initiative in learning tasks and takes responsibility for own learning	Demonstrates excellent levels of initiative in learning tasks and responsibility for own learning

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Self-management	Manage own time and meet deadlines for learning tasks	Does not manage own time adequately and does not meet deadlines for learning tasks	Meets all deadlines for learning tasks and developing management of time and priorities between several tasks	Meets all deadlines for learning tasks and capable management of time and priorities between several tasks	Meets all deadlines for tasks and demonstrates excellent use of own time to achieve goals
	Demonstrate openness, intellectual humility, spirit of enquiry	Does not have intellectual humility and is unaware of own limitations	Demonstrates developing intellectual humility, spirit of enquiry	Demonstrates accomplishment in intellectual humility and spirit of enquiry	Demonstrates excellent self-awareness and openness and explores solutions to problems in depth, showing detailed interest in the subject
Self-awareness	Evaluate learning and performance In tasks	Is unable to evaluate own performance in learning tasks	Demonstrates developing capacity to reflect on the significance of a learning task and accurately assess own performance	Demonstrates accomplishment in the capacity to reflect on the significance of a learning task and accurately assess own performance	Demonstrates excellence in reflecting on significance of a learning task, accurately assess own performance and understand own strengths and weaknesses
	Apply feedback on own performance and learning for own improvement	Does not respond to feedback and does not use it in future learning tasks	Demonstrates a developing capacity to respond well to feedback and to use it in improving responses in future tasks	Demonstrates accomplished capacity to respond well to feedback and use it in improving response to future tasks	Demonstrates excellence in responding to feedback, using it in improving response and applying it more widely in future tasks

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Trait 8 : Thinking and analysis					
Critical analysis	Demonstrate ability to think critically, to analyse and evaluate claims, evidence and arguments clearly and logically	Unable to think critically demonstrated by taking information at face value and lack of arguments in analytical learning tasks	Demonstrates developing capability to think critically, evaluate and analyse data and present logical and clear arguments	Demonstrates accomplishment in thinking critically, evaluating and analysing data and presenting logical and clear arguments	Demonstrates excellence in thinking critically, evaluating and analysing data and evidence and presenting logical and clear arguments
Problem solving	Apply problem solving processes in novel situations	Unable to apply problem solving techniques to address a new problem	Demonstrates developing capacity to apply problem solving techniques to address new problems	Demonstrates accomplishment in applying problem solving techniques to address new problems	Demonstrates excellence in applying problem solving techniques to address new problems
Trait 9 : Skills and their application in employment					
Information and IT skills	Locate, organise and evaluate information with emphasis on primary sources	Unable to locate, organise and use information to support arguments	Demonstrates developing capacity to locate and organise new information from primary sources and use it to support an argument.	Demonstrates accomplished capacity to locate and organise new information from primary sources and use it to support an argument.	Demonstrates excellent capacity to locate and organise new information from primary sources and use it to support an argument.

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Information and IT skills	Apply information and communication and other relevant technologies to solve problems and in learning tasks	Unable to utilise ICT to solve problems or apply it to own learning	Demonstrates developing capacity to utilise ICT to solve problems and apply it to own learning	Demonstrates accomplishment in utilising ICT to solve problems and applying it to own learning	Demonstrates excellence in utilising ICT to solve problems and applying it to own learning
Planning and organisation	Apply sound planning and organisational skills in learning tasks	Unable to plan and organise own resources and time to complete a learning task	Demonstrates developing capacity to plan and organise own resources and time to ensure successful completion of a learning task	Demonstrates accomplishment in planning and organising own resources and time to ensure successful completion of a learning task	Demonstrates excellence in planning and organising own resources and time to ensure successful completion of a learning task
Employment skills	Demonstrate knowledge of regulatory frameworks and political influences for professional area of study	Unaware of or unable to interpret regulatory frameworks and political influences for professional practice	Demonstrates developing knowledge of regulatory frameworks and political influences on professional practice	Demonstrates accomplished knowledge of regulatory frameworks and political influences on professional practice	Demonstrates excellent knowledge of regulatory frameworks and political influences on professional practice
	Demonstrate capability to solve real world problems by applying learning and skills from discipline(s)	Unable to apply learning from the discipline to real world problems	Demonstrates developing capacity to apply learning from the discipline to solve a real world problem	Demonstrates accomplished capacity to apply learning from the discipline to solve a real world problem	Demonstrates excellence in applying learning from the discipline to solve a real world problem.

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Flexibility	Apply different learning approaches to suit different tasks	Unable to adapt preferred learning approach for different tasks	Demonstrates developing capacity to adapt different learning approaches to suit different learning tasks	Demonstrates accomplished capacity to adapt different learning approaches to suit different learning tasks	Demonstrates excellence in adapting different learning approaches to suit different learning tasks
	Adapt to change and develop personal response responses to manage change	Unable to adapt to changed circumstances or to cope with changes	Demonstrates developing ability to adapt to change and work out own response on how to manage that change	Demonstrates accomplishment in adapting to change and working out own response on how to manage that change	Demonstrates excellent ability to adapt to change and work out own response on how to manage that change
Trait 10 : Civic and social understanding					
Citizenship	Understand what it means to be a well-informed citizen and apply knowledge and skills to contribute to own communities	Does not take interest in civic responsibilities or the community in which they work and/or live	Demonstrates a developing understanding of civic responsibilities and involvement with the community in which they work	Demonstrates an accomplished understanding of civic responsibilities and involvement with the community in which they work	Demonstrates an excellent understanding of civic responsibilities and involvement with the community in which they work

Attribute cluster	Learning outcome	Level 0 standard (Not achieved)	Level 1 standard (Developing)	Level 2 standard (Accomplished)	Level 3 standard (Exemplary)
Trait 11: Global understanding					
Global context	Develop and apply international perspectives in their discipline	Unable to apply international perspectives in the discipline	Demonstrates developing understanding of other cultures and recognises the need to adapt approaches to those different cultural environments	Demonstrates accomplished understanding of other cultures and how to adapt approaches to those different cultural environments	Demonstrates excellent understanding of other cultures and how to adapt approaches to those different cultural environments
	Demonstrate understanding and acceptance of culturally diverse and international environments	Unable to understand the need for acceptance of culturally diverse and international environments	Demonstrates developing understanding and acceptance of culturally diverse and international environments in study	Demonstrates accomplishment in understanding and acceptance of culturally diverse and international environments in study	Demonstrates excellence in understanding and acceptance of culturally diverse and international environments in study
Trait 12: Social and environmental sustainability					
Sustainability	Understand financial, social and environmental sustainability	Does not understand the concepts of financial or social or environmental sustainability	Demonstrates developing understanding of financial, social and environmental sustainability principles and practice in the discipline	Demonstrates accomplished understanding of financial, social and environmental sustainability principles and practice in the discipline	Demonstrates excellent understanding of financial, social and environmental sustainability principles and practice in the discipline

4.6 Speculations on the likelihood of measurement

While thought has been given to the potential for assessment in the design of the learning outcomes and the associated standards for the proposed Australian framework, some of the outcomes listed would be easier to measure than others. There may only be a subset of these learning outcomes which are measurable in a common or consistent way to enable institutional performance to be determined and compared, and this decision on the final minimal group of outcomes and standards to be assessed would need to be a pragmatic one.

This set of institutional or sector-wide learning outcomes derived using the framework should be sufficiently comprehensive so that institutions would be able to determine which of those outcomes could apply to their own situations. Since each Australian university has a set of graduate attributes (at the institution level) from which these learning outcomes have been derived, it is reasonable to expect that they have undertaken some mappings of course content and explored different modes of assessment which may be capable of measurement of achievement of some of these learning outcomes by the end of students' courses of study. However as has been seen from international attempts at this exercise, more effort has been expended on specifying broad statements for learning outcomes than on assessment of them. This presumption about institutions' capabilities and experience in appropriate assessment of course or institution-based outcomes is tested in a data collection about assessment approaches in use in all Australian universities.

In summary, it is a challenging task to determine robust, reliable forms of assessment of individual students within a university even at a subject level, let alone in a way that allows valid comparison between institutions at the composite or aggregated level. However, having derived the learning outcomes for the Australian environment by examining in detail institutional intentions in relation to their graduates' anticipated traits and in which there is a considerable amount of commonality, it is hoped that there may also be considerable alignment between forms of assessment which might be appropriate. This assumption is tested in the next strand of the research on assessment of individual achievement of these proposed course learning outcomes.

PART 3: ASSESSMENT OF HIGHER EDUCATION LEARNING

Chapter 5 - Using student assessment to measure individual learning

5.1 A measurement model for learning outcomes

This part of the study explores the nature of assessment in higher education and past and current approaches which may possibly be used to assess learning achieved in a course of study. It also investigates the feasibility of using assessment results for individuals as the basis of indicators of higher education institutions' performance in relation to their students achieving specified learning outcomes. That is, it explores the question: Is it possible to define and calculate new assessment-based indicators to measure institutional performance in student learning in higher education? As has been discussed in Chapter 2, current institutional performance indicators such as retention and graduation rates and employment outcomes, which are proxies for student learning and the quality of institutional teaching, have been discounted as true measures of the extent of student learning in higher education.

While there are many purposes of student assessment (Boud, 2010b; James, McInnis, & Devlin, 2002), the focus on assessment of individual student learning in this study is to gain a better understanding of what graduates know and can do (and how well they do it) on completion of their courses of study in Australian undergraduate degree programs. Because the ultimate desired output from this study is a small group of more meaningful institutional performance measures demonstrating student achievement of the set of generalisable learning outcomes derived in Chapter 4, a key aspect of assessment of interest here is the comparability of tests used and grades awarded in the assessment of students for those learning outcomes between different institutions and disciplines. This focus suggests that particular characteristics of assessment tasks and their application might need to be emphasised to ensure comparability of outcomes and standards between different institutions.

In order to develop such performance indicators, base data on measurement of actual student learning is required. This first involves identification of an appropriate approach to assessment of individual students for course based learning outcomes which lends itself to comparable and consistent results between different institutions. The study approaches this challenge for assessment by addressing the following matters which relate to the goal of reliable institutional performance measurement:

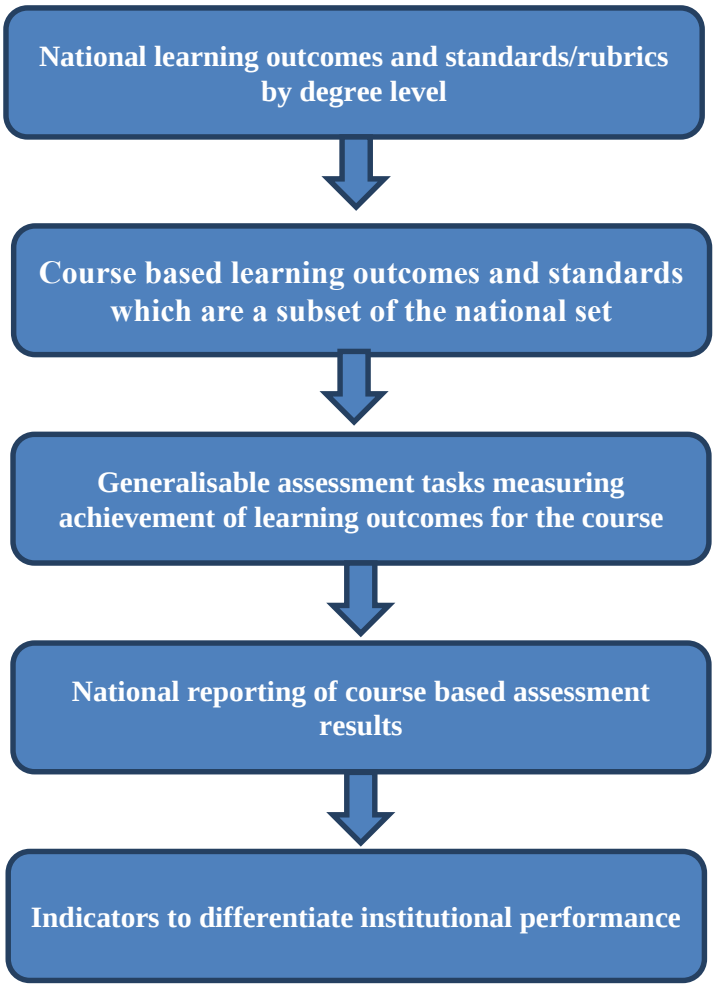
- The specification of graduate attributes at institutional and course levels and their suitability to inform learning outcomes;
- Specification of learning outcomes for each discipline and course of study which are aligned with the graduate attributes;
- Specification of standards or rubrics for each learning outcome against which achievement of that outcome can be measured;
- The need for alignment between what is taught in a course, the specified learning outcomes for the course, and the type of assessment which might demonstrate achievement of the learning outcomes;
- Resolution of the question of whether results of individual student assessment in terms of learning outcomes can be meaningfully aggregated or combined in some way to provide an accurate picture of institutional performance measurement;
- Consideration of how course learning outcomes (or institutional/ discipline based/sector outcomes) demonstrable by the end of a course of study might best be assessed when local assessment of knowledge and skills is mainly done at the subject level within courses; and
- Design of tasks for assessment of course learning outcomes which demonstrate measurement validity and reliability to allow comparison between different course/ discipline and institutional student populations to be made.

The first three of these issues have been addressed in the previous chapter for the Australian higher education environment. There are many assessment types which could address these issues but the aim is to attempt to find an assessment regime which, if possible, makes maximum use of local assessment undertaken as part of normal classroom teaching, is cost effective to run, is suited to approaches already used in Australian universities, and which is able to produce comparable aggregated results between different universities.

The model for the measurement of institutional or course learning outcomes planned at the commencement of this study (described in Chapter 1) is shown in Figure 5.1. This conceptual model is based on being able to specify generalisable learning outcomes by level of course or for a higher education institution as a whole, and assumes that it is possible to assess achievement of the learning outcomes at those levels, as has been

attempted in some overseas jurisdictions. Until now these attempts have generally involved assessment approaches which have not used existing internal subject based assessment regimes, as being explored in this study.

Figure 5.1 – Top down conceptual model for the measurement of institutional performance in achievement of course based learning outcomes



As the study has progressed, the level of uncertainty about the applicability of this model to Australian higher education has increased rather than decreased. Many overseas higher education systems have identified a single set of learning outcomes for a level of study whereas Australian higher education has been characterised by learning outcomes which are specified by each institution and which show some variability. Other differences include the political and educational environments of overseas and Australian universities; the broad spectrum of sophisticated understanding of sector-based learning outcomes and the alignment of assessment tasks with them which exists

across the universities; and the maturity of the systems used to specify learning outcomes and measure student achievement internationally.

Before moving to investigating the feasibility of identifying possible appropriate assessment regimes for the generalisable learning outcomes and standards developed in Chapter 4, a general discussion is presented about the characteristics and use of various approaches in the past and present to higher education assessment.

5.2 The role of assessment in higher education

Boud (2010a) states 'Assessment is a central feature of teaching and the curriculum. It powerfully frames how students learn and what students achieve. It is one of the most significant influences on students' experience of higher education and all that they gain from it. The role of assessment is two-fold - to assist in certification of students' achievements and more importantly to assist in their learning. Assessment is also the key means of providing feedback to students about the extent and nature of the development of their learning over time. The type of assessment required of students directs their behaviour in many ways' (p1).

While many different approaches have been used to assess individual student learning in higher education they have sometimes been in the form of additional tests or surveys administered to students by third parties, thereby focussing on the certification role more than using such tests in the pursuit of student learning. Taking up Boud's point on the influence of assessment on student learning, any forms of assessment chosen in this study to measure achievement of student learning outcomes have potential to send messages to students about what aspects of the curriculum (and which learning outcomes) their teachers think are important. Therefore assessment regimes used in the teaching of a course of study may have a significant impact on what many students choose to learn. The specification of learning outcomes and changes to existing assessment regimes can therefore have the effect of shifting the goalposts of learning for students because of the behavioural changes the greater explicitness of requirements induce in both teachers and students (Baik & James, 2014). The combination of a sector-based set of specified learning outcomes and an associated set of assessment tasks can have a powerful impact on the way teaching and learning is undertaken.

The majority of assessment in Australian higher education occurs as part of the teaching of courses and subjects in universities which have learning outcomes similar to those

specified previously. If it were possible, by building on what universities are doing internally to assess their students, to examine differences in teaching and learning quality between these institutions, more concrete and appropriate information of students' knowledge and skills would be available as feedback to the students themselves during their courses of study, and to potential employers of graduates, whose priorities may differ from the academic aims of university teachers. In addition there may also be a reduced need for reliance on special external testing and surveys, which add to students' assessment workload and increase the costs of learning assessment in institutions.

5.3 Considerations in measuring generalisable course learning outcomes and developing assessment models

While Figure 5.1 outlines the simple model for assessing sector/ institutional/course learning outcomes in a consistent way, there are difficulties in applying it and gaining comparable data between institutions which might form the basis of institutional measures of performance.

The first issue is the generalisable nature of the learning outcomes which have been developed in Chapter 4 and how well these align with the outcomes specified by individual Australian universities for their courses. There is considerable overseas activity in attempting to assess student learning in a comparable way between institutions and there have been some recent trials of such an approach in the United States and Canada where the adoption of common sector-based sets of learning outcomes is now quite widespread in higher education. However in spite of the greater acceptability of the generalisable learning outcomes to institutions than in Australia, recent work in Canada at HEQCO suggests that there is as yet no common assessment framework in place in that country and provision of data on achievement of the learning outcomes is 'voluntary, small scale and focussed on internal expectations' (Lennon & Frank, 2014) (p89) with progress being made in assessment within and not between institutions. In Canada, as in a number of other countries (Finland and Germany for example), the greatest success in measurement of course level learning outcomes to date has been achieved in professional disciplines such as Engineering, Accountancy or Medicine because of their focus on competencies and accreditation.

In Finland, institutional autonomy is also reported as a barrier to universal adoption of consistent forms of learning outcomes and their assessment. Some of the more employment-related specialist institutions have identified learning outcomes and linked those statements to assessment methods and criteria in a similar way to the rubrics and standards and their associated levels of achievement discussed in Chapter 4 (Ursin, 2014)(p166). German researchers have also developed a model for specifying and assessing course learning outcomes in a comparable way between universities, but within disciplines, and are seeking to extend this research to look at inter-disciplinary assessment for common learning outcomes in the KoKoHs initiative (Zlatkin-Troitschanskaia, 2014). The model proposed for course based measurement there involves an approach similar to that proposed for this study of specifying the competencies (or learning outcomes), developing measurement instruments and using a range of psychometric instruments for different fields of study based on current assessment practices in use in individual higher education institutions but able to be generalised across them. The conclusions reached by this initiative suggest that it is possible the test instruments developed in KoKoHs could be applied to other countries to assess learning outcomes.

While Australian higher education at present does not have an agreed single set of generalisable graduate competencies or bachelor level learning outcomes, a study on assuring graduate outcomes in Australian higher education (Barrie, Hughes, Crisp, & Bennison, 2014) focused on subjects rather than courses and emphasised the importance of developing measures of achievement of course-based learning outcomes as essential future work. There has been an intra-disciplinary example of collaboration over assessment in medicine courses in Australia through the Australian Medical Assessment Collaboration project (Schuwirth, Pearce, Australian Medical Assessment Collaboration, & ACER, 2014) but such an approach has not been successfully developed at a more general level. And so, to attempt to apply the model in Figure 5.1 directly to the Australian higher education environment is likely to challenge the autonomy of universities where curriculum, outcomes and assessment regimes are controlled for quality by Academic Boards but remain primarily the domain of individual academic staff and where a diversity of approach is highly valued. The possibility of gaining acceptance of the approach documented in Figure 5.1 within this policy environment would require strong commitment and involvement of the academic

teachers and a willingness to collaborate with colleagues, many of whom would come from outside of their own disciplines and institutions. The potential for future success of such an approach would depend on the extent of the specification and development of course and institutional-based learning outcomes within each university, the similarity of these outcomes between institutions and to the set developed earlier in this study, and the acceptance of a corresponding set of standards or rubrics that would provide a consistent basis for assessment. There are examples overseas of adaptation of statements of learning outcomes and rubrics to suit individual institutions, while still preserving sufficient comparability of assessment results across the sector (Association of American Colleges and Universities, 2010; Banta, Griffin, Flateby, & Kahn, 2009) and this is an approach worthy of consideration for this study.

As already mentioned, most assessment tasks in Australian higher education apply to subjects while the learning outcomes specified by the universities which the students are expected to be able to demonstrate on graduation are at the course level or for the whole institution. While learning outcomes are also commonly specified for individual subjects, information collected in this study suggests that frequently there is no obvious link between the subject outcomes and those specified for the courses or the institution. The cascading approach modelled in Figure 5.1 therefore does not fit the Australian situation well. Hence it seems important to find a way of linking subject assessment to the assessment of learning outcomes for a course or an institution. This challenge has been addressed in other domains with various mapping approaches documented in the literature (Ewell, Mandell, Martin, & Hutchings, 2013; Keshavarz, 2011; Schneider & Ewell, 2013). It is an important part of this alignment process that the assessment tasks used in the subjects of the course allow an holistic view of what has been learned in that course to be obtained.

Designing assessment tasks which shift from assessing subject knowledge, the most common form of classroom assignments in undergraduate courses, to generic skills and personal characteristics (such as reflection, intelligence, ambition, capacity for hard work etc.), is sometimes a challenge for teaching staff in some disciplines (Shavelson, 2009). It suggests that thought is needed about how such traits and skills can be evidenced in the context of the discipline, and the types of assessment tasks which might best allow students to demonstrate both discipline subject knowledge as well as competence in the generic skills. In this context there seems to be an increasing use of *authentic*

assessment – or so-called 'real world problems' – in disciplines which engages students in the type of tasks they are likely to encounter on graduation and entry to the workforce.

This research study first aims to find forms of assessment which might be suitable to measure *absolute* levels of achievement by individual students of learning outcomes by graduation, including what learning they brought to their studies at commencement of their tertiary course. The use of individual assessment to form a view about institutional performance presents several challenges: whether such an approach is meaningful in measurement terms; whether an aggregated outcome of individual assessment results can accurately measure institutional performance and if so, what might be the most appropriate forms of such assessment for these two purposes; and how to differentiate the impact the institution's teaching has had on what the student knows or has learned by the end point of their course. That is, in trying to use individual assessment to develop institutional performance indicators, the relationship between the student inputs and their outcomes may need to be fully explored and moderated for. Astin and Antonio (2012) (p17-19) make this point when they refer to the need for an input-outcome approach to using the results of individual assessment *for performance measurement*. This issue is likely to have an impact on the type of assessment instruments which can be used at the individual student level.

The learning outcomes and the standards described in section 4.5 (Table 4.3) represent a set of criteria against which student achievement at the end of a unit or course of study might be assessed. Assessment of any type involves judgement against standards set by the assessor, the institution, or in some cases externally by governments or a profession. In order to understand what the student actually knows and can do the conduct of assessment and the reporting of results ought to reflect the nature of the learning outcomes and their associated standards rather than the rank the student holds among their peers in the class. In designing criteria for assessment in the standards in Table 4.3 it was difficult to describe these levels of attainment in a way that did not embody comparisons across achievement levels or the use of normative concepts to describe the standards. Attainment of standards or levels of achievement requires the holistic judgment of the academics teaching the subjects in the course, and the effect of individual teacher autonomy is likely to vary between different subjects.

One of the challenges of using local internal assessment approaches rather than external standardised tests is to establish whether the results of individual tests might be calibrated in some way so that differences in achievement of groups of students could be meaningfully compared with those in other courses, disciplines or institutions. In order to compare assessment results between institutions or groups as planned for this study, it is important that any new forms of assessment proposed must have validity and reliability demonstrated. For this study *validity* will be used to mean that the assessment approach is appropriate to all students undertaking the tasks not advantaging students from a particular background or discipline: that is, it must produce accurate results and be relevant to the learning it is testing. To be *reliable* the assessment must produce consistent results when applied on separate occasions either to different groups or over time to the same group.

Finally, not all of the thirty eight generalisable learning outcomes specified in Chapter 4 of this study may be measurable using traditional forms of assessment and not all may be selected by an institution as relevant to their courses of study. But given the desire to obtain a holistic view of student achievement for an institution or for the sector, the number of assessment tasks used would need to be sufficiently broad but not overwhelming, to cover commonly used learning outcomes across a range of universities.

5.4 Possible assessment regimes for generalisable learning outcomes

The task of identifying ways to assess the learning outcomes specified in Chapter 4 is approached by first exploring the nature of assessment and reviewing literature on results achieved internationally on assessment of learning outcomes of the type developed for the Australian system. This involves exploring pedagogical issues relating to what constitutes good practice in higher education assessment as well as more pragmatic views on what is possible in reality. Progress made recently in overseas countries in finding appropriate ways to assess sector wide learning outcomes from some of the typologies analysed earlier is reviewed and strengths and weaknesses of various approaches canvassed.

Interest in the development of standards and their use in assessment is high as evidenced by literature on assessment from Australian experts in the field (James et al., 2002; Sadler, 1987, 2005, 2011, 2014). In spite of this interest there is currently little known

externally to the individual universities about their assessment practices and how they use learning outcomes and standards to inform judgments about their students' learning. However, there have been a number of projects on assessment in Australian universities funded through the previous ALTC and the current Office of Learning and Teaching ((Australian Learning and Teaching Council, 2010a); Australian Learning and Teaching Council (2011a), (Barrie et al., 2014; Boud, 2010b; Freeman, 2009, 2011; Oliver, 2011a, 2011b) which have involved identifying examples of good practice in subject-based assessment in place in particular institutions.

Determining which assessment tasks are used in subjects or courses and the levels of achievement reached by students has always been the domain of the academic teaching staff, but recognition of the importance of student engagement in their learning is leading to changed practices which seek to involve students more in the assessment process. Such initiatives include input into the nature of the assessment tasks and assisting in the evaluation of their own and their peers' performance (Baik & James, 2014; Boud, 2010a). These initiatives have potential to reform the way assessment is used in higher education. In Australia assessment and grading practices have been seen as 'the most important safeguard of quality assurance of academic standards' because of the autonomous nature of the universities (James et al. (2002)) (p17). And yet in this environment the setting of standards and describing learning outcomes has been idiosyncratic to each institution. In addition the term *assessment* is used in many different ways in the literature and often interchangeably with *measurement, evaluation, testing and reporting.* It is important to distinguish between these terms and how they are currently used in higher education.

5.4.1 Definitions

Several authors provide a broad definition of *assessment* which enables its application both at the level of an individual student's learning and at an institution level, depending on the focus or purpose of the assessment approach. Griffin and Nix (1991) (p2), while writing about school education, describe assessment generally as 'all methods used to appraise the performance of an individual pupil or group'. Specifically in their view *assessment* involves the collection of evidence of student achievement and its interpretation while *measurement* means the assignment of grades and numbers, and *evaluation* is the judgment about worth for reporting purposes. *Reporting* is the process

of transmitting information to stakeholders to create an awareness and interest in the policies, goals and operations of the institution.

Astin and Antonio (2012) (p5) make a similar comment on the importance of clarifying use of the terms and describe assessment as 'the operationalisation of notions of excellence'. Their primary interest is in the use of assessment to drive excellence in education, whether for the individual or the organisation providing the education. Their assumption is that 'the basic purpose of assessing students is to enhance their educational development' (p5) and thereby 'advance the institutional mission of teaching and learning'. They therefore describe assessment more broadly as 'the gathering of information concerning the functioning of students, staff and institutions of higher education' (p2). Their interest extends to using individual students' assessment results to determine the performance improvement of an individual or an organisation, in a similar way to this study. The use of the term assessment for both measuring the extent of student learning and for evaluation of an institution's performance is confusing, and so for the purposes of the following discussion about using assessment for ascertaining student achievement against learning outcomes, the following definitions are used.

- *Assessment* is any group of methods used to appraise the knowledge, skills and attitudes of an individual or a group of students at the completion of a course of study.
- *Measurement* is the assignment of numerical scores or quantifying the classification of observations about students' performance in assessment tasks.

Hence in this study measurement will relate to the assignment of the scores specified for each of the standards associated with the learning outcomes statements either for a student, group or institution.

- *Reporting* is the provision of information for stakeholders at the group or institution level on knowledge, skills and attitudes of students resulting from a course of study.

The calculation of a set of performance indicators based on group assessment results is an example of reporting.

- *Evaluation* is the drawing of conclusions and making judgments about performance of these entities based on the results of measurement or reporting.

5.4.2 Selected characteristics of assessment

Assessment, as defined above, is a way of understanding the extent and depth of students' learning. It is usually undertaken at multiple points during the teaching and/or at the end of the student's period of study. It is also diagnostic in identifying difficulties a student may be having in understanding the material taught and provides insight for the teacher on how to change the way they present course material so that students' learning and the quality of teaching are maximised. The most appropriate type of assessment for a particular situation relates to the purpose of the assessment and what knowledge and skills it is testing.

The requirement in most universities for teachers to provide details of the assessment regime they intend to use in each subject contributing to a degree course at the start of the period of study means that many students sometimes use these specifications to focus on particular aspects of the curriculum being taught rather than trying to attain all of the specified learning objectives related to the study. In situations where performance outcomes are being measured or published externally for quality assurance or funding purposes there can also be a tendency for the teacher to 'teach to the test', and not cover all of the learning outcomes in the curriculum, or for the student's learning to be constrained or misguided because of the assumptions made as a result of assessment specifications. It has been reported that a pattern like this has been observed in the NAPLAN testing arrangements in primary schools in Australia and is a salutary warning of how the purpose and advantages of a more transparent generalised assessment regime can be impacted negatively by the behavioural changes of students and teachers once details of the assessment are made explicit.

Assessment has many different aspects, some of which are discussed below. These by no means represent a complete list of characteristics but the selection has been chosen because of the particular relevance of each aspect to the learning outcomes, the need for comparability of results, and the approach proposed in the current study.

Formative and/or summative assessment

Formative assessments are tasks which primarily help in the development of the student. They are used to provide feedback to students about how they are

progressing with their learning and to help them improve later performance. Formative assessment is usually repeated at various stages in a subject or course and to test increasingly complex aspects of learning or more sophisticated interpretation and use of knowledge as the student progresses. Summative assessment tasks form the basis of judgments about the extent of the student's learning for a particular period. Such assessments usually test achievement at the end of study of a particular subject or for the course overall. Summative assessment is more commonly used for accountability purposes when making comparisons between achievements of groups of students against standards (Shavelson, 2009) or more generally in terms of certification. James and Baik (2014) report that there are recent developments of 'nested assessment' design across a course which build on previous assessment tasks and demonstrate true formative assessment while ultimately providing a summative evaluation of achievement by the end of a course.

Norm-referenced or criterion-referenced

Norm-referenced grading systems identify individual students' performance relative to that of their fellow students in a group ((Astin & Antonio, 2012; Boud, 2010b; Shavelson, 2009). Norm-referencing therefore says little about how much students know and can do (ie which learning outcomes they have achieved) but rather whether they do it better than their peers. Common measures of achievement used in universities such as Grade Point Averages, or weighted average marks which are often compared between students in different universities are generally based on norm-referenced subject results. Criterion-referencing measures students' academic performance against a set of standards or rubrics which include the criterion (the learning outcome being tested) and a set of benchmarks, similar to those developed by the AACU in its VALUE matrix or those formulated in section 4.5 of the previous chapter of this study. In criterion referenced systems there is no pre-determined or normalised distribution of grades anticipated and each individual's performance rating is not influenced by the performance of others in the same cohort. There is an increasing trend towards use by individual teachers of criterion referenced grading practices in Australian universities as well as growing use of rubrics and standards as is shown later and by authors in the field (Baik & James, 2014; Barrie et al., 2014).

Assessment of absolute achievement or value added

Absolute achievement is used in this study to mean the attainment of the student at the particular point of time the assessment tasks are performed, while value-added results identify how much the student has developed or improved between two time points, most commonly between commencement and completion of a course. In the literature learning is sometimes defined in these relative terms (Astin & Antonio, 2012; Shavelson, 2009) necessitating the use of assessment tasks which can be appropriately taken at the start and the end of study to allow the amount of learning to be ascertained. While input standards are commonly used to determine whether a student may enter a particular course, and there is usually good correlation between level of achievement on admission and that during or at the end of a course, these 'grades' are usually norm-referenced and so say little about what has actually been learned. In this study assessment relates to learning outcomes achieved by the end of a course of study. These 'results' of assessment will be influenced by what knowledge and intellect (prior learning) the student brings to their tertiary study from prior education and experience as well as the learning achieved during their current course of study.

Discipline based conceptual knowledge testing and testing of higher order skills

Assessment of higher education course learning outcomes involves measuring students' achievements in both understanding complex discipline-based conceptual knowledge as well as skills on how such knowledge can be used effectively in later professional life. These higher order generic skills are commonly developed across a number of different subjects in a course and tested using assessment tasks which are set within these different subjects. There is a risk that the assessment of these generic skills within individual discipline based subjects can become fragmented in the curriculum and not lead to a comprehensive view of student achievement in these areas.

Holistic assessment or a set of individual tasks

The challenge then becomes one of gaining a holistic view of the achievement of course-based learning outcomes through a range of assessment activity within the discipline context. In the past, teachers have tended to focus on testing knowledge of discipline content as part of the curriculum while generic skills testing has

sometimes been taken outside of subject based routine assessment and done through special external testing instruments such as those used in the United States for many years and trialled in Australia in several universities (the Graduate Skills Assessment) (Department of Education Training and Youth Affairs, 2001). Determining achievement of course learning outcomes is likely to involve understanding the relationship between various course and subject based assessment tasks and building up a picture of holistic achievement of the student across a range of assessment regimes.

Embedded assessment or testing regimes external to the curriculum

Standardised testing of students is common in the United States and other countries (Nusche & Organisation for Economic Cooperation and Development, 2008). Generalised external testing is often preferred by government because of its apparent independence and because it usually provides normalised or standardised performance results which make inter-institutional comparisons possible. However, these external assessment regimes are frequently very time-consuming and arduous from the student perspective and costly for institutions. In fact, in most countries reviewed, the majority of assessment of students in higher education courses is internal and embedded in the normal teaching process. Local course assessments are also much more directly focussed on the objectives of what is to be learned in the course rather than an external set of questions (often multiple-choice) which are not directly related to the documented learning outcomes of the course or institution.

Voluntary or mandatory assessment tasks

Assessment tasks can be mandatory or voluntary but there is usually a core set of mandatory tasks to enable decisions and comparisons to be made about performance of individual students. Voluntary assessment tasks are used largely for formative purposes for students in a particular cohort. In some systems where value-added is central to the definition of student learning, and where external testing is used, or when institutional comparisons of attainment are needed, sampling may be used rather than all students being required to take all tests. In a criterion-based assessment system where there is a desire to compare results

across institutions, there is a need at least for a core set of mandatory assessment tasks to be undertaken by all students.

Individual and team assessment

Individual assessment of students by their teachers remains the most common form of knowledge testing in higher education but increasingly with emphasis on some of the employment-based skills such as teamwork and communication the teaching process involves working in groups of peers, who are sometimes called upon to co-assess their fellow students' performance. The use of assessment for developing higher education students' understanding of standards and capacity to make informed judgments about academic performance is also seen as one effective way of engendering some of the common learning outcomes such as self-criticism, and lifelong learning in their discipline with reference to appropriate standards. Team assessment therefore has a place in modern undergraduate higher education, and building an appropriate partnership between teachers and learners in assessment also strengthens the learning partnership between these two groups (Boud (2010b)(p1)).

Authentic assessment and 'real world' problems

The term 'authentic assessment' is increasingly used in the literature (Banta et al., 2009; James et al., 2002) (p10) to describe assessment tasks that are perceived by students as being 'real' – ie they present challenges in applying learning to 'real world' problems which are representative of the types of professional activity the student will probably be engaged in after they complete their degree. Because of this, authentic assessment is likely to engage students more with their studies, resulting in better learning and building capacity to address workplace skills. Authentic assessment tasks also frequently require the demonstration of some of the generic skills such as problem solving and critical analysis as well as discipline knowledge, and so are efficient forms of assessment to test attainment of several of a set of learning outcomes at once.

Assessment tasks and approaches for the learning outcomes specified for Australian undergraduate degrees of this study under the model described in Figure 5.1, could therefore take many forms across the range of courses and higher education institutions. The assessment tasks which might allow valid comparative analysis of student

achievements of learning outcomes by the end of a course of study would have to be good measurement tools for certification purposes but should also have characteristics that encourage and guide student learning and the development of stated generic skills.

Most internal assessment of higher education learning outcomes is still focussed on cognitive learning outcomes (Astin & Antonio, 2012; Ewell, Mandell, et al., 2013; Nusche & Organisation for Economic Cooperation and Development, 2008; Shavelson, 2009) but the outcomes in Chapter 4 also include non-cognitive learning outcomes such as students' attitudes, values, aspirations and social and interpersonal relationships which should be tested if a full picture of individual achievement is to be obtained. Given the measurement model outlined, and a desire to use individual assessment of end of course learning outcomes as the basis of institutional performance measurement, the type of individual student assessment that might be usefully explored for this study is *summative,* and *criterion-based,* which aims to assess *absolute* achievement at the end of a student's course of study. The assessment tasks specified in this approach should be *mandatory*, aligned with the learning outcomes, and *embedded* and aligned within the curriculum and teaching for the relevant undergraduate degree.

5.5 Fitness for purpose of assessment

For some time there has been debate about whether individual student assessment results should or could be used in the evaluation of institutional quality or performance, or whether entirely different types of assessment should be designed for the latter purpose. Much of this debate relates to the definition of assessment and the level of the organisational unit to which it is being applied. When institutional performance or quality is being 'assessed' and the institution is treated as a group of individuals some argue that assessment in this sense is used to evaluate institutional or course effectiveness not individuals' performance. In these cases the assessment becomes an *evaluation* in terms of the definitions outlined above, and measurement is focussed on achievement of objectives or targets set for the relevant organisational level as proposed by Astin and Antonio. Performance indicators themselves could be regarded as assessment tools or independent reviews of progress against targets rather than just reporting tools. In this study the focus is on both individual learning and institutional performance.

There are also measurement theory issues as to whether accurate results of institutional performance can be obtained by combining individual student learning outcomes data in various ways (Sanders & Horn, 1994). This is a very important issue for the current study, which ultimately hopes to use individual direct assessment results combined in some way in the calculation of performance indicators. The current state of play suggests that attempts are being made to use such data for both purposes in higher education (Ursin, 2014; Zlatkin-Troitschanskaia, Kuhn, & Toepper, 2014), but there has been little success so far in using detailed information about the levels of knowledge and skills attainment of individual students, to form a judgment on institutional teaching and learning quality (i.e. how well students taught by each institution are able to demonstrate different levels of achievement of a suite of end of course learning outcomes).

In Australian higher education, other surrogate examples of learning outcomes such as progression rates and attrition rates at different levels of a course are used to differentiate institutional performance where individual data on what subjects a student has 'passed' and their tendency to re-enrol are aggregated to provide insight into their institutions' performance and quality of teaching. While this information does not demonstrate what students have learned, the performance indicators are formed by aggregation of individual student data and have been rigorously tested and used effectively and reliably for many years for differentiating teaching and learning quality (Linke et al, 1991) and to allocate teaching quality funds between universities, hence being used at both individual and institutional levels.

This issue does not appear to be a major concern in most of the recent projects undertaken on assessment of higher education learning outcomes. Ewell (2010) contends that the types of assessment for the two purposes of measuring student achievement of learning outcomes should differ and has devised a taxonomy of approaches to assessment which is based on two different paradigms. His framework differentiates between the focus or purpose of assessment (learning and teaching or accountability) and the level of 'assessment' (individual or group). Performance indicators are identified as an appropriate form of summative 'assessment' for group accountability, as is the aim of this study, while he argues that individual summative assessment should only be a form of gatekeeping for academic standards using tasks such as comprehensive examinations, certification examinations, or capstone

performances. Ewell does state that individual student assessment results can be aggregated and used for course or program evaluation purposes, provided this is for enhancement of teaching and used in a formative or developmental capacity. But it is difficult to see why, with appropriate mapping of curriculum to specified learning outcomes and the alignment of assessment with a set of standards, that data from this form of assessment cannot serve both purposes of evaluation of individual and, in aggregated form, institutional performance. The standards developed in section 4.5 allow for different levels of achievement of learning outcomes which could be seen both as progression of development of knowledge and skills during and as an absolute level of learning at the end of a course of study (ie measure the degree of development of the student over time).

5.6 Internal or external testing of individual student learning?

This section explores approaches to assessment which have been used in the past to test individual student learning and describes the policy context driving each type. The outcomes being tested in these cases are generally equivalent to the high level learning outcomes in some of the taxonomies examined earlier rather than the more detailed statements of learning outcomes and rubrics or standards now used. A brief history of the type of tests used to assess course outcomes in the US in the past provides some insight into practical issues which may need to be faced when trying to assess generalisable learning outcomes, and how the nature of assessment policy has changed over time. This has relevance to the issue of whether it is feasible to use local forms of assessment or whether external standardised tests are the best way of assessing course based outcomes and generic skills.

There have been four distinct eras of assessment in the US from 1900 to the present (Shavelson (2007, 2009) (p33), (p22)) which have led to the widespread use of externally administered tests. Each of these eras had a particular emphasis and the transitions between the eras were related to external changes in higher education structure or policy direction which necessitated a different approach to assessment.

Higher education assessment moved from an initial ‘emphasis on the mastery of academic content’ at the undergraduate level (Shavelson, 2007)(p28) between 1900-1933 using standardised internal essays and examinations which required considerable effort in marking and grading, to large objective external tests of accumulated

declarative knowledge (1933-1947), to the current emphasis on skills and attributes such as the 'personal, social and moral outcomes of general education' through the emergence of general education colleges and then subsequently professional graduate education. Later (1948-1978) new tests such as the Graduate Record Examination (GRE) were introduced by external testing agencies such as the Educational Testing Service (ETS). These tests were also of long duration and arduous for students to take. They focused on the measurement of communication, problem solving and 'values clarification' using a range of different types of assessment including multiple choice as well as short answer questions, essays and direct observation (Shavelson, 2009) (pp29-31). This seems to be one of the first attempts to develop 'authentic' assessment in that real world tasks were formulated and there was a focus on how to solve new problems rather than simply recalling content that had been taught.

Ultimately the high cost of this type of assessment and the political pressure for external accountability of institutions to improve their student learning gave rise to the current era of external accountability (1979 – present) in the US and the emphasis on testing which can allow institutional performance and quality to be differentiated between institutions. In spite of this required accountability in terms of external test results it is reported that academic staff in US universities still prefer their own form of more open-ended problem-based assessment to these tests (Kuh, Jankowski, & Ikenberry, 2013; Shavelson, 2009) because they consider that it more accurately measures what students really know and how they go about solving problems.

Some external tests such as the Collegiate Learning Assessment developed by the Council for Aid to Education in the US do use examples of open ended assessment tasks and a range of grading approaches with input from individual assessors but with electronic marking. This reduces the cost of testing but it is still an additional task for students and one which does not make use of the large quantity of information available from classroom based assessment. The external testing regimes are not usually tied directly to particular institutions' curricula or their stated course learning outcomes and may represent a lowest common denominator approach to assessment of student learning outcomes.

Nusche (Nusche and Organisation for Economic Cooperation and Development (2008)(p14)) summarised international state of play several years ago with large-scale

external assessment regimes using standardised external testing. This pre-dated the AHELO learning outcomes assessment feasibility project funded by the OECD and conducted between 2009 and 2013 ((Edwards & Pearce, 2014) and other assessment approaches being trialled or implemented in a range of European and Asian countries (Coates, 2014).

Large scale standardised assessments are not widely used *within* higher education institutions to assess learning, and such tests are generally not mandatory. Most of these tests are used to measure value-added based on an assumption that the way to separate the impact of the teaching at the higher education institution is to see how knowledge and generic skills have changed during the time the student has spent at the institution. Use of value-added assessment instruments are generally not focussed on precise statements of learning outcomes and standards or rubrics such as those developed in Chapter 4. To measure *what* a student knows in terms of learning outcomes as a result of a program of study, open-ended tests which address the detail of the standards or rubrics and allow judgments to be made about absolute levels of student performance are likely to be more useful.

The external instruments for assessing learning outcomes used in the US today are commonly run by commercial organisations (for example the Educational Testing Service (ETS), the American College Testing program (ACT), and the Council for Aid to Education (CAE)) and are focussed on generic skills not discipline knowledge. Use of these instruments is widespread largely because of the history of use of this type of test and support from accrediting bodies and those concerned with demonstrating quality assurance. The Voluntary System of Accountability (VSA) was established in 2007 by the American Association of State Colleges and Universities and the National Association of State Universities and Land-Grant Colleges following the recommendations of the Spellings Commission. The VSA has the primary purpose of 'evaluating core educational outcomes in public universities and colleges by focusing on skills that are common, multidisciplinary and university-wide' (Lydia Liu Ou, 2009). These common skills were defined as *written communication skills, critical thinking, and analytical skills*. The VSA selected three standardised tests as possible measurement tools: the Measure of Academic Proficiency and Progress (MAPP), the Collegiate Assessment of Academic Proficiency (CAAP) and the Collegiate Learning Assessment (CLA), all of which are mentioned by Nusche in her list of assessment

instruments currently in use. Over 300 institutions in the US have selected one of these testing regimes, and an issue for these testing organisations and institutions is the extent to which the results of the different tests, which all purport to measure these three core learning outcomes are comparable (Lydia Liu Ou, 2009, 2011a; Lydia Lui Ou, 2011b).

This brief history emphasises the changing fashions in higher education assessment during the last century in the US and highlights that the changes which occurred seem to be driven less by educational or pedagogical factors than the policy environments in place and the cost and impact of assessment load on students and institutions. So in spite of the ease of application of generalised tests and their usefulness in comparative studies between institutions, they are not the complete solution to the testing of achievement of all student learning outcomes.

While the focus of this brief history of learning outcomes assessment has been on approaches used in the US there have been similar developments in Australia and in other countries for specific purposes in some disciplines. In particular the Australian Council for Educational Research has done much work on developing tests of to measure learning outcome attainment, often in conjunction with particular interest groups (for example, the Graduate Skills Assessment, and the Australian Medical Assessment Collaboration). The impetus behind these initiatives has been changes in the national policy environment governing higher education from the perspective of improving quality of teaching and learning and the need to demonstrate greater accountability for funds.

5.7 Conceptual frameworks for assessment of individual student learning outcomes

A number of conceptual frameworks have been developed for assessment of graduate learning outcomes. Two of these discussed below are based on the concept of educational improvement. A holistic picture of student learning is favoured in both cases using a range of different assessment instruments. For example, Shavelson (2009) bases his framework on his definition of learning as a relative concept related to improvement in knowledge and skills by a student over a course of study (ie a value-added model). In his view an holistic picture of student learning is best obtained by using a combination of standardised external tests administered to individuals for accountability purposes (ie quality and performance) and ‘internally developed

assessments closely reflecting a campus' mission for improvement' (p19, p42). This suggests that local classroom-based assessment could be validly used in part for the wider purpose of institutional accountability as long as there is some benchmark information also forming part of the evaluation of performance. Nusche (2008)(p8) is of a similar view and believes as long as standards are clearly set and criterion based assessment is used, there should be no difficulty in validly comparing institutional performance derived from individual results from such a set of assessment tasks. But in reality it is not as simple as that.

Shavelson's conceptual framework (2007)(p33) is a two dimensional matrix of aligned assessment of achievement of learning and the collection of snapshots of knowledge and skills both at one point of time (achievement) and over time (learning). The Shavelson assessment framework consists of a 'focus on broad cognitive abilities *embedded* in meaningful, holistic, complex tasks and use of information technologies – to assess academic programs'. In Shavelson's terms, the current study aims to measure individual student *achievement* at the time point of the end of a student's course of study (this has been referred to as absolute achievement on graduation in this study).

The assessment framework developed by Astin and Antonio (2012) is multi-dimensional with variables of type of learning outcome, type of measurement data, and time. The time dimension is described as long-term (ie the end of a course of study) or short term (during studies), while the learning outcomes are classified into two categories: *cognitive* and *affective* (non-cognitive) outcomes. The types of assessment data are described as either Behavioural or Psychological. The cognitive outcomes represent the type of learning outcomes routinely locally tested by higher education institutions and are most likely to be regarded as relevant to the educational objectives of stakeholder groups such as employers because of their association with 'knowledge and higher order mental processes such as reasoning and logic' (p46). Affective outcomes are associated with students' attitudes, values, aspirations and social and interpersonal relationships. The learning outcomes identify *what* is being measured while the types of data describe *how* the outcomes are to be measured (Astin & Antonio, 2012) (p47). Psychological data reflects the internal reasoning of the student while behavioural data 'reflects the interaction of the student and the environment' (p49).

According to this framework the type of psychometric instruments which are likely to be valid in the measurement of individual student learning for cognitive outcomes are either external tests or internal classroom formal assessment while the affective outcomes describe personal traits such as student motivation, attitudes and self-concept, which have been commonly assessed in different ways through the use of instruments such as psychological tests, questionnaires about personal attributes and experiences and student surveys (p48).

Table 5.1 below shows the higher level Tier 2 learning outcomes identified for Australian undergraduate degrees in Chapter 4 distributed across the four quadrants of the Astin model. In this application of the framework they are all long term outcomes as they represent expected attainment by graduation. They are evenly distributed between cognitive and affective outcomes but are biased towards psychological measurement. This distribution is not atypical according to Astin and Antonio.

Table 5.1 –The 25 Tier 2 Attribute clusters for Australian undergraduate education by type of assessment data

	Type of learning outcome	
Type of assessment data	**Cognitive**	**Affective**
Psychological	Discipline knowledge Discipline-related skills Scholarship Research in discipline Innovation Critical analysis Problem solving Independent intellectual enquiry Information and IT skills Planning and organisation	Equity and social justice Indigenous cultural understanding Professional ethics related to discipline Personal ethics Self-management Self-awareness Lifelong learning Global context Sustainability
Behavioural	Employment skills Flexibility	Teamwork Leadership Communication Citizenship

Source of framework: Astin and Antonio (2012), p48

Examples of third party external assessment which might reliably measure the non-discipline related cognitive/ psychological learning outcomes above such as critical thinking and problem solving include the type of VSA tests discussed above. However use of such tests has not been supported in Australian higher education in spite of several attempts to introduce them and their popularity in some overseas countries.

The behavioural instruments identified by Astin and Antonio to assess attainment in affective learning outcomes include questions about students' involvement in co-curricular activities during their period of study, examining graduate outcomes in the workforce on completion, and results of surveys of student satisfaction with their educational experience during their course of study. Similar surveys of student satisfaction, student engagement and graduate employment are conducted annually for Australian undergraduates but were earlier discounted as sole measures for student learning and teaching quality. But Astin's and Shavelson's conceptual frameworks (Shavelson & Huang, 2003) stress the importance of gaining an holistic view of students' achievement of learning outcomes and suggest that such measures can be useful alongside more detailed internal assessment results in forming a view of a student's achievement of the whole suite of cognitive and affective learning outcomes. Therefore using the results of such surveys as performance measures, while they are not directly indicative of the achievement of particular detailed learning outcomes, may be appropriate in a conceptual framework for a complete suite of performance measures for student learning being applied to affective learning outcome achievement. Such issues are canvassed in Chapter 7 on performance indicators.

The logical distinction between cognitive and non-cognitive learning outcomes is seen in several typologies in the literature and the Astin framework is helpful in providing insight into aligning the type of assessment tasks with the type of learning outcome. The above application of the Tier 2 learning outcomes identified in Table 4.3 for Australian undergraduate education suggests that the Astin framework looks like a good fit for thinking further about assessment of the full range of outcomes for Australian undergraduate degrees and in distinguishing between levels of institutional performance. It is therefore proposed that Astin's framework be applied later in the development of a range of assessment tasks which align with the learning outcomes identified in Chapter 4.

Neither Astin nor Shavelson voice philosophical objections about the assessment of individual student learning being used to evaluate institutional performance as long as the assessment instruments are reliable and are founded on a principle of student improvement. Their greater concerns are that if the individual student attainment of learning is to be used for measurement of institutional improvement or performance, then the impact of the initial attributes and skills of the individual students and their personal characteristics need to be taken into account in any final evaluation of institutional performance or quality based on assessment data. This means that the broad concept of the current study which links assessment to performance indicators appears feasible provided care is taken in choosing the forms of assessment of individual students and in drawing conclusions about apparent differences in the indicator values between institutions.

In summary, assessment theory suggests using a variety of assessment tasks which test both cognitive and affective learning outcomes and include external testing of both content knowledge and generic skills 'supplemented by locally devised assessments that are sensitive to campus goals and curricula' (Shavelson, 2010, p41). Both frameworks indicate that the most effective assessment occurs when it incorporates specialised local assessment into a holistic assessment of the full set of learning outcomes.

5.8 Types of assessment regimes currently in use

5.8.1 A taxonomy of assessment approaches

Many different assessment regimes are consistent with these two conceptual frameworks. Coates (2011)(p3) summarises the range of assessment approaches currently in use to measure achievement of course or institutional learning outcomes in a taxonomy which includes five types of assessment regimes. Each of these categories of assessment has different strengths in the application to the problem of comparative measurement of achievement of generalised end of course learning outcomes. In the context of this study, the question becomes one of which combination of these forms of assessment might provide the most accurate view of what students have learned during their courses while being practically feasible to administer and use for comparative institutional purposes. Coates' taxonomy (Coates, 2011) has the following groups of assessment tasks:

- *Local achievement assessments* – these are routine direct assessment tasks used by individual higher education providers and teachers during or at the end of the student's course of study such as subject examinations, essays, laboratory classes, assignments, tutorial performance, capstone subjects and class presentations. They generally lack comparability between courses and institutions, but applications of agreed standards and sharing of information between universities may improve the consistency and usefulness of the assessment data collected through them. They represent the majority of assessment undertaken in higher education institutions;
- *Generalisable achievement assessments* – these are externally developed assessments which involve some form of external moderation to ensure comparability of standards. The special examinations in domain specific knowledge and skills in South America and Mexico mentioned by Nusche are examples of these;
- *Generic capability tests* – These are the specific tests designed to assess learners' generic skills and learning outcomes. For example, the CAAP, MAPP, CLA and GRE mentioned in the previous section;
- *Assessment of discipline capability* - These tests are usually taken at the end of a course of study and focus on topics considered core to the discipline at the final-year level. They may be in the form of tests of suitability for admission to graduate programs or higher level qualifications in a particular discipline; and
- *Licensing examinations* – These provide comprehensive measurement of specific enumerated competencies, may be related to accreditation of professional courses and are usually seen in disciplines such as medicine, law and business.

All but the first of these groups of assessments are additional tasks set above the local assessment undertaken in higher education institutions. Some only apply within particular disciplines and are therefore not sufficient to compare achievement of learning outcomes between disciplines or more generally between institutions. Others focus only on a subset of learning outcomes and alone are unlikely to enable a holistic view of student achievement on graduation to be obtained.

5.8.2 Learning lessons from recent developments in assessment

An international approach

The recent OECD feasibility study on Assessment of Higher Education Learning Outcomes (AHELO) (Coates & Richardson, 2011) was an ambitious example of a globalised assessment of discipline and generic skills ability for the disciplines of engineering and economics at the bachelor degree level. The project followed the approach of defining learning outcomes and devising appropriate assessment instruments aligned with those outcomes but then tested the instruments extensively in the field. Its goal was very broad – to be able to measure student learning outcomes in a comparable way not only across institutions as in this study but across national boundaries – and to provide metrics which complemented routine internal assessments of student achievement.

In order to overcome anticipated opposition to the concept of generalisable assessment tasks, Coates and Richardson addressed this goal by framing an engagement strategy which involved the academic communities (staff and students) from all participating institutions and building a community of scholars which could be consulted so that 'philosophical, political and historical scepticism to assessment innovation' could be understood and addressed (2011, p59). This group was used to develop a set of assessment instruments to assess the specified learning outcomes between disciplines and countries. Issues addressed included how to account for institutional diversity in curriculum, course structure and duration in comparative analysis of test results; the desire of participating institutions to use approaches such as normalised scoring of assessment results and un-standardised rubrics; and adapting non-uniform assessment instruments through an assessment framework so that results from their application might be valid and reliable across institutions and other domains (Richardson & Coates, 2014)(p829). While the AHELO trial was ultimately not extended to implementation, there are some learnings from its approach which may be useful for this study, such as the need for a conceptual assessment framework to assist in the design of appropriate assessment tasks for each of the generalised learning outcomes identified. The example of the Engineering Assessment Framework in the AHELO paper is discipline specific but resembles the statement of *graduateness* for the Australian higher education sector developed in section 4.2 from which the detailed learning outcomes (Tier 2 of the conceptual model for this study) were derived. Edwards and Pearce (2014)(p85)

conclude that the trial offered 'useful insight into the development, engagement, implementation and reporting of large scale common assessments intended to provide measures of learning outcomes'.

The standardised tests discussed earlier for the United States in section 5.6 represent a less complex practical solution in terms of ensuring comparability of individual student achievement between institutions and for measurement of institutional performance and quality, but they are not the complete answer to the challenge of measuring levels of student achievement of detailed learning outcomes as suggested by Shavelson and Astin and Antonio and others. The desire to measure common learning outcomes accurately across state or international borders as yet remains unfulfilled. And so, an approach which explores a new model of assessment of generalisable learning outcomes based on use of local achievement assessment (Category 1 of Coates' taxonomy) and taking account of the issues experienced in the AHELO study appears worthy of further examination.

Because detailed information about classroom and other forms of assessment is generally not available outside of individual universities, it is a challenge to find out exactly how comparable existing assessment approaches and their results are between institutions. But with the pressure on universities from government to demonstrate greater effectiveness and quality in their teaching and students' learning, and from employers and students to be clear about intended outcomes, these internal approaches alone are no longer sufficient for these accountability purposes unless a way can be found of making assessment and results more comparable between institutions and better aligned with stated learning outcomes.

Changes in US institutional assessment practice

In the US where there has been a more intensive focus on specifying generalisable learning outcomes and experimenting with different assessment practices, attempts have been made to see what impact these different approaches are having on the nature of localised assessment. The National Institute of Learning Outcomes Assessment (NILOA) undertook two surveys in 2009 and 2013 of Provosts of all regionally accredited undergraduate degree granting institutions about their use of generalised learning outcomes and various types of assessment in the context of the adoption of the Degree Qualifications Profile, discussed in Chapter 3. The results of these surveys show

that there has been a significant increase in the uptake in the use of the LEAP learning outcomes, and in the number and variety of assessment tasks used to test student knowledge and skills. At present in the US it is reported that 84% of institutions use common learning outcomes for all of their students, and 40% reported that the learning outcomes for their courses aligned with the stated national learning outcomes of the DQP. Course and regional accreditation processes are cited as the main drivers of assessment of learning outcomes followed by an interest in demonstrating improvement in teaching and learning (Kuh et al., 2013; Kuh, Jankowski, Ikenberry, & Kinzie, 2014). The 2013 survey revealed use of increasingly diverse forms of assessment within and between institutions ranging from 'classroom based assessment' (i.e. local achievement assessment) to student surveys to gain an understanding of what students have learned during their courses. Such an approach is consistent with the desire for holistic evaluation of student achievement advocated by Shavelson and Astin for example. Focusing only on the non-survey forms of assessment of graduate outcomes, the popularity of different types of assessment in 2013 in public universities in the US was as follows: Incoming Placement Examinations (76% in public universities); Rubrics (66%); Classroom based performance assessments (65%); Standardised external tests of generic skills (eg VSA tests) (47%); Capstone assessments (47%); External performance assessments (39%) (e.g. professional entry tests for graduate qualifications); and Portfolios (37%). Use of rubrics, external performance assessments and portfolios showed the greatest increase between 2009 and 2013 in that order. These results are consistent with the reported take-up of the DQP and the VALUE Rubrics (p33) in the US as common learning outcomes for liberal education during these four years. In spite of these reported gains in aligning assessment with outcomes, it is concluded that maintaining the recent progress of institutional assessment work must be an ongoing priority with increased academic staff involvement in the collection and use of student learning outcomes data as the basis of assessment design, showing that there is still much work to be done.

Australian practice

Currently there is no similar information available about Australian institutional assessment practice. This is in spite of the several discipline-based large projects on standards and learning outcomes funded by the Australian Government through its Office of Learning and Teaching, which were discussed in the previous chapter. Until

recently less attention has been paid to assessment of these outcomes than in several countries (particularly the US, Canada, and Germany)(Coates, 2014). Also there is not a history of using external standardised tests in Australian higher education and successive attempts made by the Australian Government in 2000 and 2010 to introduce a routine external generic skills assessment through trialling *The Graduate Skills Assessment* and subsequently through an adapted *Collegiate Learning Assessment* failed. This was because of the voluntary nature of the testing, and concerns of the higher education academic community about its validity (Australian Council for Educational Research, 2001; Department of Education Employment and Workplace Relations, 2011).

The AAGLO project

The issue of identifying appropriate assessment tasks to measure achievement of learning outcomes for an institution or courses nevertheless remains a topic of great research interest to the Australian Government and the institutions themselves. Hence a 2014 study also funded by the Office of Learning and Teaching on assessing and Assuring Australian Graduate Learning Outcomes (AAGLO) (Barrie et al., 2014) concentrated on local assessments and addressed two assessment-related questions central to the current study:

- What types of assessment are most likely to provide evidence of student achievement or progress towards graduate learning outcomes?
- What processes best assure the quality of assessment of graduate learning outcomes?

Using a large number of academic practitioners across several Australian and international institutions and seven disciplines, the project focussed on identifying good practice in classroom assessment tasks in subjects volunteered by the participants and how such tasks might best be used to assure the quality of graduate learning outcomes. The project concluded that there had been a 'shift in emphasis from input standards to academic process and outcome standards' under the impetus of changed quality regulatory processes since the advent of the Tertiary Education Quality and Standards Agency in 2012, and that a viable approach to assuring graduate outcomes to the satisfaction of TEQSA would require multiple approaches to assessment and 'a central role being played by academic-led discipline based assessment' which was aligned with

learning outcomes (p7). Being able to demonstrate attainment by graduates of learning outcomes by the end of their courses and alignment of curriculum and assessment tasks during their studies are two of over 40 standards required to be demonstrated by all Australian higher education providers registered by the Government and for accreditation of individual courses run by non-university higher education providers (non-self-accrediting institutions) (Tertiary Education Quality and Standards Agency, 2011).

Mostly the good practice assessment examples identified in the AAGLO project are built on realistic practical problems in professional practice and were frequently similar across different disciplines but were couched within the favoured traditional forms of assignments used in each discipline. The most common forms of assessment were typical examples which have been used for many years: reports, critical reviews and essays, oral presentations, tutorial participation, and examinations in that order. A large number of the tasks were multi-component in nature, testing more than one learning outcome (for example preparation and presentation of a report to an audience based on particular discipline knowledge might test critical analysis, discipline knowledge and oral communication skills).

The leaders of the AAGLO project point out that robust and meaningful assessment for measurement of quality or performance purposes is not just about the tasks themselves but also 'a key feature underpinning the credibility of discipline-based assessment is the appropriate choice of assessment task coupled with the credible assurance of the quality of the assessment task and the assurance of the quality of judgments made on the basis of students' performance of that task' (Barrie et al., 2014). This is of central importance for the current study as even if it is possible to identify some appropriate assessment tasks which align with the generalisable learning outcomes specified in Chapter 4, the consistency of application of those tasks for each learning outcome between institutions, disciplines and courses, the marking of student submissions on the tasks and the assurance that the tasks across different disciplines and institutions are of comparable difficulty are issues requiring attention.

Through the involvement of a large group of practitioners from the academic community the AAGLO project identified a number of key policy issues which might be contributing to the current incapacity to assess learning outcomes at a course or

institutional level adequately in Australian higher education. The policy issues include fragmented design of assessment at the course level, assessment policy gaps and inconsistencies in some universities and the use of norm-referenced moderation of assessment results. The solutions proposed to improve assessment of learning outcomes include whole of course planning of assessment and agreement on standards and criterion-referenced assessment. These solutions are focussed on improvement within an institution and while they do not directly address the issues of transferability of assessment tasks between institutions or disciplines to enable reliable measurement of quality or institutional performance, they provide useful exemplars from which to build more generalised assessment task construction.

The best practice examples of assessment tasks included are for individual subjects across the seven disciplines represented in the project and they led the authors to conclude that 'effective practice in the assessment of Graduate Learning Outcomes exists throughout the Australian higher education sector' (p41). This seems a rather confident generalisation from the available data. While the examples do show use of innovative assessment in a selection of subjects, one of the project conclusions is that there is a need to promote whole-of course approaches to assessment planning to be able to truly demonstrate achievement of course based learning outcomes. While the single subject good practice examples identify the characteristics of assessment tasks which have potential to be adopted more generally, the project does not make suggestions on how this might be done, or on the extent to which any Australian universities have made progress in measuring achievement of their stated course or institutional learning outcomes by students on graduation. It is this gap that the chapter dealing with the Australian undergraduate environment which follows tries to fill.

5.9 A possible assessment framework and principles of assessment

This review of the characteristics of assessment generally and recent developments in assessing learning outcomes locally and internationally highlight issues to consider when attempting to find an appropriate assessment regime for the generalisable Australian undergraduate learning outcomes and standards specified in Chapter 4.

5.9.1 Conceptual framework for assessment of the prescribed learning outcomes

Referring to the findings of the AHELO project that a conceptual framework of assessment is a necessary 'theoretical base on which educational achievement can be

investigated' (Richardson & Coates, 2014) (p829), consideration has been given to what such a framework for advancing the assessment of learning outcomes at the undergraduate level in Australia might include. The proposed framework, which has been developed through the strands of investigation of this study and is used in the following chapter to explore possibilities for assessment of generalisable undergraduate learning outcomes, consists of the following:

- The statement of *graduateness* in section 4.2
- The 38 detailed learning outcomes (Tier 3 of the taxonomy for undergraduate learning outcomes) derived from this statement and used to varying degrees by individual universities (section 4.4.2);
- The standards derived in section 4.5 against which quantitative measures of achievement can be determined at several levels;
- A set of principles of assessment which are couched in a measurement model and which support good practice in comparability of assessment of the learning outcomes; and
- A set of generalisable local achievement assessment tasks which could be used by individual higher education providers and teachers during or at the end of the student's course of study.

The principles of assessment enunciated below are framed in the context of the institutional performance measurement aim of this study and so they focus on characteristics which most lend themselves to the comparative measurement of the set of generalisable learning outcomes which form part of the above measurement framework. In the literature there are many other examples of sets of principles that have a different emphasis (Boud, 2010a; Griffin & Nix, 1991; James et al., 2002). For example Boud's set is focussed on conditions which encourage student learning while James' set is about the quality of student assessment generally. The following principles are more pragmatic and utilitarian but have been based on the ideas and frameworks discussed earlier in this chapter.

5.9.2 Principles of assessment

The proposed set of principles of assessment derived from these considerations which could be used to form a judgment on the effectiveness of assessment regimes for measuring learning outcomes and for institutional performance purposes are:

1. Assessment is holistic in relation to the specified learning outcomes at the course, discipline or institutional level;
2. Assessment tasks form part of the routine local achievement assessments of students in their courses (ie they are Coates category 1 types of assessment);
3. There are multiple types of assessment used in the set which contribute to an overall testing of achievement against the specified learning outcomes and standards but the set should pragmatically not be too large, so that differences in patterns of achievement can be distinguished for each course and institution;
4. Assessment tasks are able to be used summatively to determine attainment of learning outcomes on graduation but ideally also formatively during the period of study to encourage and assess development of students through their courses;
5. Each assessment task preferably addresses more than one learning outcome and draws on the context of the discipline for which the learning outcomes are being assessed, particularly across the categories of discipline knowledge and generic skills or other attributes;
6. The tasks are developed collaboratively and implemented with significant input from the teachers of the course content within and between institutions;
7. They are able to be moderated or calibrated for validity and for reliability (ie describe student achievement in a fair, relevant and accurate manner); and
8. The results of the assessment are transparent and able to be easily communicated to all higher education stakeholders (students, academic staff, and employers).

These principles will be used to evaluate options for assessment regimes for Australian undergraduate learning outcomes discussed in the following chapter.

The exploration of these general aspects of higher education assessment and the identification of a possible conceptual framework for thinking about assessment of the undergraduate learning outcomes derived earlier hopefully assists in the formulation of some potential approaches to using local achievement assessments to measure individual attainment of learning outcomes.

Chapter 6 - Towards a new approach to learning outcomes assessment

6.1 Introduction

In this chapter various options are explored for measuring individual student achievement of a subset of the 38 generalisable graduate learning outcomes included at Tier 3 of the conceptual framework for specification and assessment of learning outcomes shown in Figure 4.1. In this investigation, it is assumed that the levels of achievement of students can be measured in terms of the standards presented at Tier 4 of that framework. But first it would be helpful to establish a view on the current state of assessment policy and practice in the Australian higher education sector and to gain a clearer understanding of the approaches used within universities to evaluate whether their institutional or course based learning outcomes are actually being achieved. It seems that more information is required about these base line practices and the commonality of their assessment regimes before new comparative approaches to assessing generalisable learning outcomes can be proposed.

This task is approached by building on the general characteristics and taxonomies for assessment presented in the previous chapter as the theoretical base, but exploring pedagogical issues as well as more pragmatic approaches on what may be possible to achieve in reality.

As discussed in Chapter 3, the Australian higher education system is less advanced in specifying and using generalised learning outcomes at the sector, institution or course level than many international systems, but interest in assessment and the development of standards and their use in grading within institutions is high as evidenced by writing and seminars on assessment from Australian experts in the field (for example (Australian Learning and Teaching Council, 2011a; Coates 2010; Sadler, 1987, 2005, 2011)). In addition to the AAGLO project there have been a number of other studies on assessment in Australian universities funded through the previous ALTC and the current Office of Learning and Teaching (Australian Learning and Teaching Council, 2010a; Barrie et al., 2014; Boud, 2010b; Freeman, 2009, 2011; Oliver, 2011a, 2011b) but these are more general in nature (eg broad guides to assessment practice, peer review of assessment, moderation of results etc) and address issues of principle rather than describing particular assessment practices.

In order to identify common practices and to compare Australian higher education assessment practices and models with those in other countries, a collection of data on universities' use of learning outcomes and their current assessment practices was undertaken. This data collection also helped identify the frequency of specification of the 38 learning outcomes derived in Chapter 4 within universities to examine the fit of the generalisable set with actual practice in the institutions. It was hoped that this information might form the basis of a set of proposals for assessing in a comparable way the individual results of achievement of the generalisable learning outcomes included in the conceptual framework identified in section 5.9, but the detail provided in the data collection was not sufficient for this purpose.

Based on the information collected through these pieces of qualitative research, three hypothetical options for advancing approaches to assessment, which may lend themselves to comparative measurement of student achievement between institutions, are developed and their strengths and weaknesses discussed. Where possible a minimal set of generalisable assessment tasks for those learning outcomes that are most commonly cited by the universities and which lend themselves to psychometric measurement, can be considered for the three assessment options. As discussed by Astin and Antonio not all learning outcomes are directly assessable psychometrically and it may be that a combination of direct assessment and other information proves to be the best pragmatic solution to the question of assessing the whole set of learning outcomes.

6.2 Understanding the Australian context

6.2.1 Learning outcomes and assessment policy settings from Australian university websites

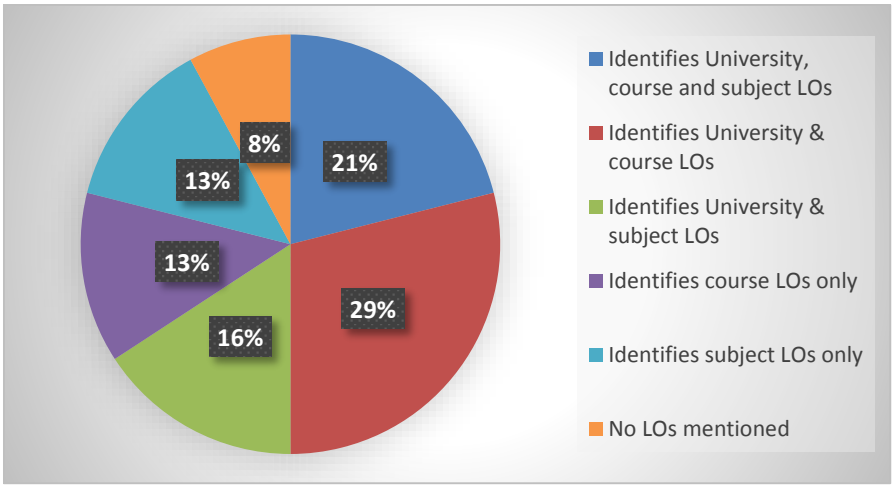
As noted in Chapter 5 the outcome assessment models used in the US, Canada and some European countries are commonly based on sector-wide or discipline-based learning outcomes; it is planned that assessment tasks should be designed around these broad learning outcomes which can be applied across all institutions, disciplines and courses; and national reporting often occurs in terms of standardised externally designed and managed test results. This model, as depicted in Figure 5.1, for course learning outcomes assessment and reporting was the one planned to be used for the Australian higher education sector, but the absence of a Government-prescribed or consortia-accepted set of learning outcomes and the rejection of government attempts to introduce a sector-wide standardised external testing system suggests that a more nuanced model

is needed for the Australian higher education sector. In addition, as argued earlier, there may be a number of benefits to exploring the use of routine internal classroom assessment tasks (internal achievement assessment) such as recommended by Shavelson and Astin to more effectively demonstrate student learning in terms of the Tier 3 outcomes identified in Chapter 4, rather than using externally conducted tests. The wide range of different approaches currently used in Australian undergraduate education are highlighted through a collection of data published on Australian university websites and a special data collection survey of all institutions undertaken for the current study.

All Australian universities' websites were searched for references to learning outcomes, course mapping, and assessment. It is by no means certain that the information on these sites is kept up to date, but since their content is of critical importance to students and their learning, it is hoped that it gives at least a broad indication of institutional practice relating to learning outcomes specification and approaches to assessment of them.

The results of this search are shown in Figures 6.1 and 6.2 below. They reveal that while all universities, as expected, refer to learning outcomes in their online course and subject materials and have policies on assessment it is not really clear how they go about assessing whether students have achieved their stated outcomes.

Figure 6.1 - Learning outcome policy settings, Australian universities, 2014



Source : University websites related to learning outcomes, extracted August 2014

Figure 6.1 shows that:

- 29% of universities indicate that they specify institutional and course level learning outcomes but do not mention subject learning outcomes and how these relate to the outcomes for a course to which the subjects contribute;
- 21% provide details of a fully cascading model specifying learning outcomes at institution, course and subject levels;
- 16% specify institutional and subject learning outcomes but none for courses;
- 13% specify learning outcomes only for courses and not subjects or institution;
- 13% specify learning outcomes only for subjects and not courses or institution; and
- 8% of the universities' websites make no reference to learning outcomes at all.

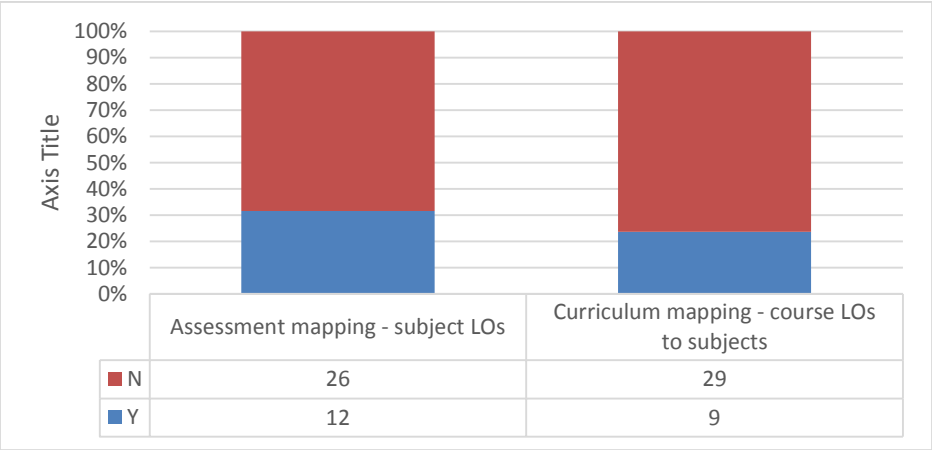
The majority (66%) of the 38 Australian universities surveyed specify general learning outcomes at the institutional level for all their graduates (one of the universities had no valid email for the staff member to whom the data proforma was to be sent). These outcomes are not usually differentiated by level of study and are sometimes referred to as graduate attributes or graduate capabilities as well as learning outcomes on the websites. The focus of 50% of the universities is on identifying learning outcomes for subjects but what is often not clear is whether the outcomes identified for subjects have any relationship to any of the course or institutional graduate outcomes that may have also been specified.

As expected *local achievement assessments* as described by Coates are usually undertaken for subjects while in 79% of cases, universities are specifying learning outcomes at the institutional level or course level.

A common way of ensuring these higher level outcomes are being covered in the curriculum and assessed in a way that demonstrates their achievement is to identify in which subjects of each course the particular learning outcomes are being taught and assessed (Bloxham & Boyd, 2007; Ewell, Mandell, et al., 2013; Keshavarz, 2011). This builds a picture of assessment of course or institutional learning outcomes bottom-up from the outcomes identified for subjects included in the curriculum rather than the more top-down approach at course or institution level used in some jurisdictions overseas.

Figure 6.2 shows the extent to which Australian universities appear, from their websites, to engage in these mapping activities. This information is extracted from the webpages on assessment and learning outcomes for each university.

Figure 6.2 - Status of mapping of learning outcomes to curriculum and assessment, Australian universities, 2014



Source : University websites related to learning outcomes, curriculum mapping and assessment, August 2014

If the information on university websites is an accurate guide to Australian universities' practice in this area, less than a quarter of the 38 universities engage in any form of curriculum mapping of learning outcomes from subject to course, and a slightly higher percentage commonly have adopted a practice of aligning assessment tasks within subjects to their specified learning outcomes at this level. The above simple analysis of website content seems to reveal that universities are stronger on the rhetoric of the need to specify graduate learning outcomes at several levels than the practice of assessing against the learning outcomes and demonstrating that students achieve them. One of the other objectives of undertaking a separate data collection about the types of assessment activity in use at universities was to try to gain a better understanding of the extent to which the generalisable learning outcomes identified in Chapter 4 were common to universities and their courses and whether the forms of assessment in use were aligned to them.

6.2.2 Assessment data collection from universities

The data collection proforma sent to all Australian universities' senior executives with responsibility for teaching and learning sought information about the acceptability and

frequency of use within universities of the type of learning outcomes identified in Chapter 4, and the nature of assessment practices in use in their institutions which might measure their student achievement of the outcomes. The proforma aimed to enable the:

- Exploration of the commonality between the 38 generalisable learning outcomes and those specified by individual universities;
- Identification of the state of the Australian higher education sector in assessing their learning outcomes and the type of assessment tasks used;
- Potential identification of good practice examples from the sector for the application of the principles of assessment for university learning outcomes;
- Exploration of how comparable the assessments used across disciplines were within the institutions;
- Provision of information on how and where data is stored about achievement of the learning outcomes in universities' computer systems; and
- Discovery of whether they would be willing to allow access to de-identified data from those systems for particular types of assessment tasks to allow trialling of options proposed in this study.

Of the 38 data collection proformas sent out, 19 universities responded to the request for data. The distribution of the responses received is shown in Table 6.1. These were achieved after a follow-up email was sent to those institutions which had not responded within a month of receipt of the original request. Two of the proformas were completed following an interview with the relevant executive and a further one was completed using data provided from the university's teaching and learning website, for which permission was given by the university. A copy of the data collection proforma is attached as Appendix 6.1.

Table 6.1 – Summary of responses to the request for information on learning outcomes and assessment, Australian universities, 2014

Type of response	No. of responses	% of all forms sent out
Completed proforma	14	37%
No learning outcomes	2	5%
Refusal to participate	3	8%
Total responding	19	50%
No response	19	50%
No valid email	1	-
Total	39	100%

Some of those institutions surveyed considered that the proforma was too complex and believed it would take too long to complete fully. Others simply did not reply. In the end the information provided by those completing the proforma was less conclusive than originally anticipated because of its non-specific nature. However what was collected was sufficient to do some qualitative analysis and what was derived from this is summarised below.

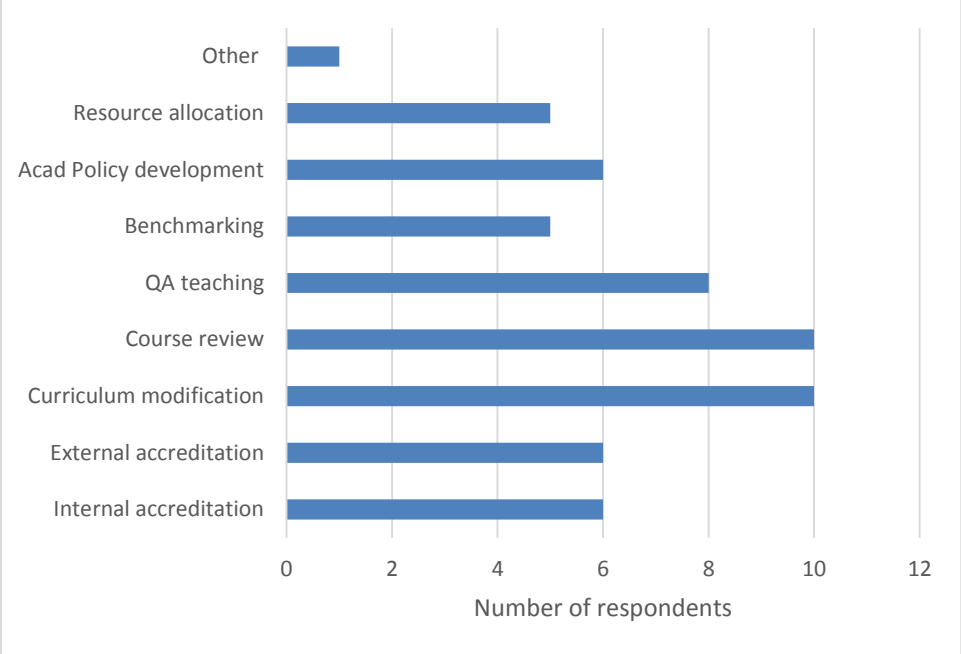
Twelve of the fourteen respondents who completed the form indicated that their universities had whole of institution learning outcomes while the other two used learning outcomes but at a lower level – for subjects or courses or both. The institution or discipline learning outcomes specified by each of these 14 were a subset of the 38 generic learning outcomes identified on the proforma suggesting that the group derived earlier might be a reasonable representation of sector practices. This is hardly surprising since they were originally derived from statements of graduate attributes published by these universities but it is interesting that these statements are being applied not just at the institutional level but in three cases also at each level down to subjects in their course hierarchies.

The number of institution-wide learning outcomes specified by the respondents ranged from 3 to 9 with one university identifying 13 detailed learning outcomes in four higher level categories. Thus all universities providing data had embraced the concept of having graduate learning outcomes at some level, and only two institutions (of 39) reported that they did not use learning outcomes at all. Three of the respondent

institutions explicitly stated that while they specified learning outcomes they did not assess against them. The remaining 11 institutions all claimed to assess against the outcomes specified, but from the information provided on assessment on the proforma it was not clear how or at which level. It is also of note that there is considerable overlap between the institutions which have no details of curriculum or assessment mapping on their websites and those who did not reply to the request for data about their assessment practices.

The most common reasons given for explicitly specifying learning outcomes were related to quality of courses and teaching and were as shown in the following chart. Multiple options were allowed.

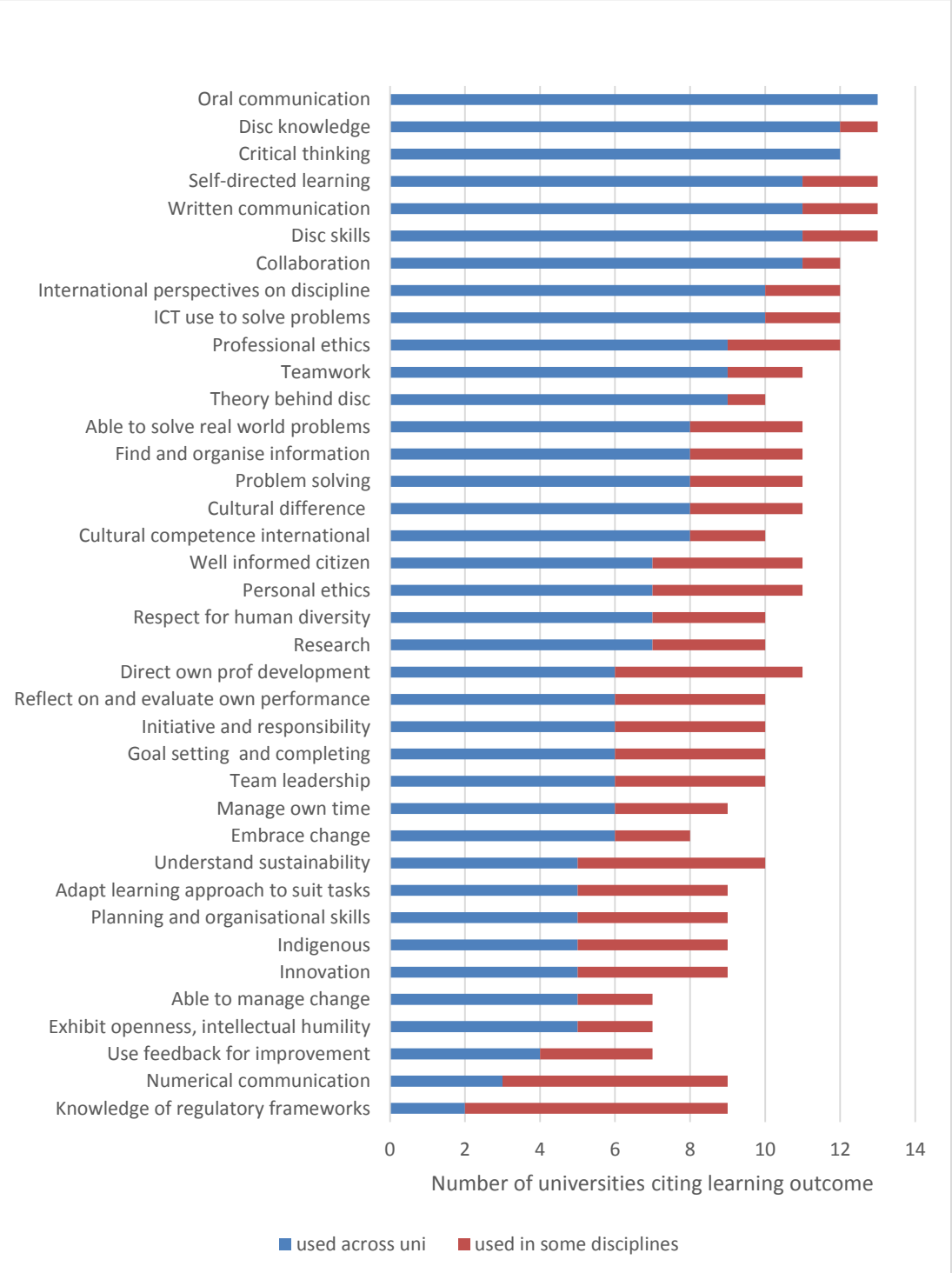
Figure 6.3 – Purpose of using learning outcomes



The majority of the 14 universities fully completing the proforma indicated that they used similar outcomes to the generic set across the whole university or at least in some disciplines. In most cases all of the thirty eight outcomes were selected. This did not align well with other data provided by those respondents for the number of institutional learning outcomes in use and suggests that while institutions subscribe to the rhetoric of institutional and course learning outcomes, they are reporting an amalgamation of all learning outcomes across individual subjects, many of which do not match their institutional-wide ones.

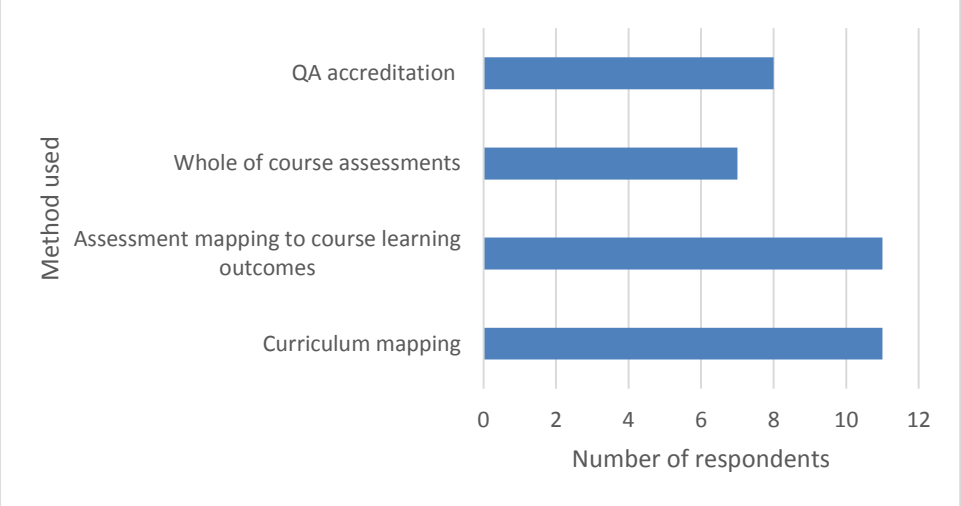
The following chart provides information about the frequency of respondents' use of the 38 generalised learning outcomes derived in Chapter 4 and which were included in the survey proforma. The data is sorted by number of institutions indicating the learning outcome is used across the whole university.

Figure 6.4 – Use of 38 Tier 3 generalisable learning outcomes by Australian universities for undergraduate study



Universities were also asked how they ensured that their institution-wide learning outcomes were aligned with their curriculum and assessment in individual courses of study. Their responses are shown in Figure 6.5 and indicate that the most common stated approach is to map the curriculum and assessment tasks (probably at the subject level). In some fields universities relied upon their own internal quality assurance processes or the requirements of accreditation in their professional degrees to ensure alignment between content or assessment and the learning outcomes.

Figure 6.5 –Means of ensuring alignment between learning outcomes and assessment



The percentage of respondents reporting use of whole of course assessments is higher than expected from the review of websites but it is not clear from the responses what form this assessment takes (eg portfolios to be submitted by the end of a course or some form of capstone assessment or surrogate surveys of student experience as was the case for two institutions). In some cases the responses were also internally inconsistent with the responses given against the separate learning outcome statements in the data collection instrument. For example, the number of course learning outcomes stated as being used in the early part of the data collection ranged from 3 to 13 course based learning outcomes while most respondents ticked all of the 38 generalisable learning outcomes as being used in the institution either across the board or in at least some disciplines. There was little alignment between the institutional level learning outcomes identified by the universities and the forms of course based assessment reported. However the responses indicated use of a wide range of similar outcomes possibly for

individual subjects taught as part of courses or in particular disciplines and examples of assessment which might be applied to such subjects in order to establish achievement of those outcomes.

Use of curriculum mapping and assessment/learning outcome alignment

Eleven of the nineteen responding universities (58%) reported they were using some form of curriculum mapping to identify in which subjects the course learning outcomes are taught and a similar number also claim that they have mapped assessment approaches at the course level. The 50% response rate for the collection of data and the insights provided by reviewing the websites suggest that this percentage would not translate to a similar overall percentage of the sector using curriculum mapping and assessment alignment (in fact the figure was 24% from the website review as shown in Figure 6.2). Also actual information about the specific types of assessment used for each of the learning outcomes was often not provided or looked similar for quite different learning outcomes suggesting it was not specifically designed for or aligned with the particular learning outcome.

Criterion or norm-referenced

Ten universities (53% of those responding) reported that they did assess against the stated learning outcomes, that the type of assessment used was criterion-referenced and that it was both formative and summative, but these examples are likely to be related to using the course based learning outcomes at the subject level, given the context stated above. One university responding to the data collection stated that its assessment was norm-referenced only. Several reported that because assessment occurred at a subject level there were too many diverse approaches in the control of individual subject coordinators to be able to state in a general way what type of assessment was or could be best used across a course of study. This was an unexpected response because, based on the scan of statements made about course based learning outcomes on the institutional websites, it appeared that most universities specified course and/or institutional learning outcomes and so might have been expected to have clear assessment strategies for measuring achievement of the outcomes at these levels.

Types of assessment tasks

One respondent reported that the only ‘assessment’ regime used across the whole university for some of the course or institutional learning outcomes was a student

survey which asked students about the extent to which they believed they had achieved the stated learning outcomes. This may be the situation for holistic assessment at the course or institutional level for generic skills learning outcomes but also suggests that there has been no comprehensive attempt to make use of local achievement assessment tasks for the measurement of learning outcome attainment of knowledge learned and skills developed for subjects which have been linked to course based learning outcomes. Two universities provided some detailed statements of the type of assessment employed for each of their learning outcomes but made the point that this was in effect an 'averaged' result across subjects and the reality was likely to be more diverse.

Because of these factors the data collected in this study on type of assessment was not comprehensive or specific enough for particular learning outcomes to enable direct comparisons to be made with the US survey results discussed previously. The AAGLO project gave some examples of frequency of assessment types for each of the seven disciplines involved in that project which provides some supplementary information for the survey results. The most frequently mentioned assessment tasks for graduate learning outcomes identified were reports; multi-component tasks; critical reviews/essays; oral presentations; tutorials/rehearsals; reflective pieces; examinations; performance; work placements; and working demonstrations in that order. There was no mention of use of rubrics or standards in spite of the reported use of criterion-based assessment, but this is not surprising given that each institution was acting independently and their aim was not necessarily to obtain results which could be compared between institutions or disciplines.

Local storage of assessment results data

Finally, another unexpected result from the survey was that a number of universities apparently hold no easily retrievable information in central computer record systems on students' assessment results for components of subjects or individual assessment tasks. In fact only 5 institutions reported that they held such information in their systems. This was surprising given the considerable length of time learning management systems have been implemented in most universities, are often used to test student knowledge, and the rising interest in student analytics. These systems typically contain information about curriculum taught and assessment tasks set for individual subjects. While student management systems are used to record a final mark and/or grade for each student/subject combination largely for certification purposes, it had been expected that

more detailed information about levels of achievement would be available in the related learning management systems. The survey results showed that more commonly universities hold these subject component results from assessment tasks in local departmental systems such as spreadsheets and the results are not available outside of the department or even shared with other teaching staff in the same department. This places severe practical limitations on comparing results in similar assessment tests across disciplines or courses within an institution, let alone across several institutions as was originally hoped for the current study. Newer smaller higher education providers which are subject to greater regulation in the sector may have more modern systems and be able to demonstrate clearer links between assessment and stated learning outcomes for subjects and accredited courses but this was beyond the scope of the sample. No below-subject level assessment results for higher education institutions are reported to the government in Australia and hence not stored or reported by the sector.

6.2.3 Conclusions from the data collection

First, there is considerable agreement on the appropriateness of the generalisable learning outcomes developed in this study. At least 80% of the 14 full respondents to the data collection cited the following 15 common outcomes of the 38 specified as being used in their universities (as shown in Figure 6.4):

- Discipline knowledge
- Discipline-related skills
- Learn collaboratively
- Participate as a team member
- Oral communication
- Written communication
- Recognise cultural difference and diversity in work
- Professional ethics
- Learn independently in a self-directed manner
- Direction of own professional development
- Problem solving
- Critical thinking
- Using ICT to solve problems
- Able to solve real world problems

- International perspectives on the discipline.

This suggests that the above subset could potentially be regarded as a common set of Australian undergraduate learning outcomes for which a set of assessment tasks might be developed for use in measurement of achievement by students in a range of disciplines at different universities.

Secondly, two thirds of respondents cited learning outcomes at a university level, but with a few exceptions, do not appear to test student achievement against these in any holistic way. Those which do assess achievement against learning outcomes seem to either consider only individual subject learning outcomes in isolation or in some cases attempt to align course outcomes and the curriculum. They do this by undertaking mapping of individual course content against course based learning outcomes which may or may not be nested within the university wide set of learning outcomes. Hence the main focus at present seems to be more on ensuring that content related to learning outcomes is taught rather than on assessment being appropriate to measure achievement of those course learning outcomes. While criterion-based assessment is reported in the data collection as common, it appears that normalised grading schemes are also in use and universities often require conformity or standardisation in their subject grading systems with information stored for certification purposes according to their assessment policy documentation. While such an approach gives an impression of comparability of results, the validity of measurement is within the subject, and not necessarily within the course or discipline. This data collection suggests that while learning outcomes are stated at multiple levels (institution, course and subject), there is rarely an institutional view in their implementation of them through assessment and curriculum as advocated by Shavelson and Astin, for example.

Thirdly, in Australia it seems that there might be a narrow baseline of experience in universities for meaningful holistic assessment of course-based learning outcomes even though there may be widespread effort in linking assessment and subject based learning outcomes. There is little evidence external to the institutions on whether the forms of assessment used to measure student attainment are well aligned with the institutional learning outcomes they specify. Using local achievement classroom-based assessment at a course level, linking this to the course learning outcomes, and then to institutional performance measurement as planned in this study is a long way from the paradigm

which exists in most higher education institutions in Australia at present. Instead their teaching staff are focussed on improving teaching practices at the subject level to increase individual student learning within their discipline. This seems to fit with the model of institutional autonomy and the role of the subject teacher in Australian higher education rather than taking a perspective of learning between courses, disciplines and the institution as sought in this study. The question remains as to whether it is possible to use these diverse types of local achievement assessments at the subject level within an institution and adapt or generalise them so that comparable valid results can be obtained on achievement of the generalisable learning outcomes by students in different courses, disciplines and universities.

And yet while the approaches to assessment used in Australian universities appear to remain idiosyncratic the institutions seek to be active organisations in the global higher education environment. In order to demonstrate international competitiveness, the Australian higher education system should be able to provide evidence that it is at least of the quality of overseas peer institutions and that its graduates do develop prescribed knowledge and skills by completion of their courses of study as sought by each of their stakeholder groups. This measurement of student achievement should be done in a way which is consistent with the Australian higher education sector's learning environment. To achieve this, universities may need to be willing to explore approaches used elsewhere and test them in their own environment so that a consolidated view of the sector achievements in student learning can be presented as part of a robust quality assurance/ performance culture. Most of these assessment approaches are likely to involve psychometric measurement but as proposed by Shavelson, and Astin and Antonio, it may also be appropriate to include other forms of assessment for affective learning outcomes or to use criterion-based local assessment in parallel with standardised externally administered tests to gain the most complete picture of student learning against specified graduate outcomes.

Some questions which are raised in examining possible new assessment regimes for the suite of learning outcomes include:

- How many of the 38 learning outcomes specified in Chapter 4 can reasonably be measured;

- How many are needed to be assessed to give a sufficiently comprehensive picture of how well students have acquired the stated knowledge and skills described by these learning outcomes by the end of their studies;
- Which ones can realistically be measured taking into account measurement issues, pragmatism about the availability of data, and the willingness of teaching staff in universities to adapt their approaches to assessment; and
- What will be the impacts on student learning and the pedagogy of assessment of each approach tested.

These questions and a discussion of the strengths and weaknesses of various proposals, are explored in identifying some options for approaches to assessment of student learning outcomes using local assessment tasks, which have varying potential for comparability of results across institutions and courses.

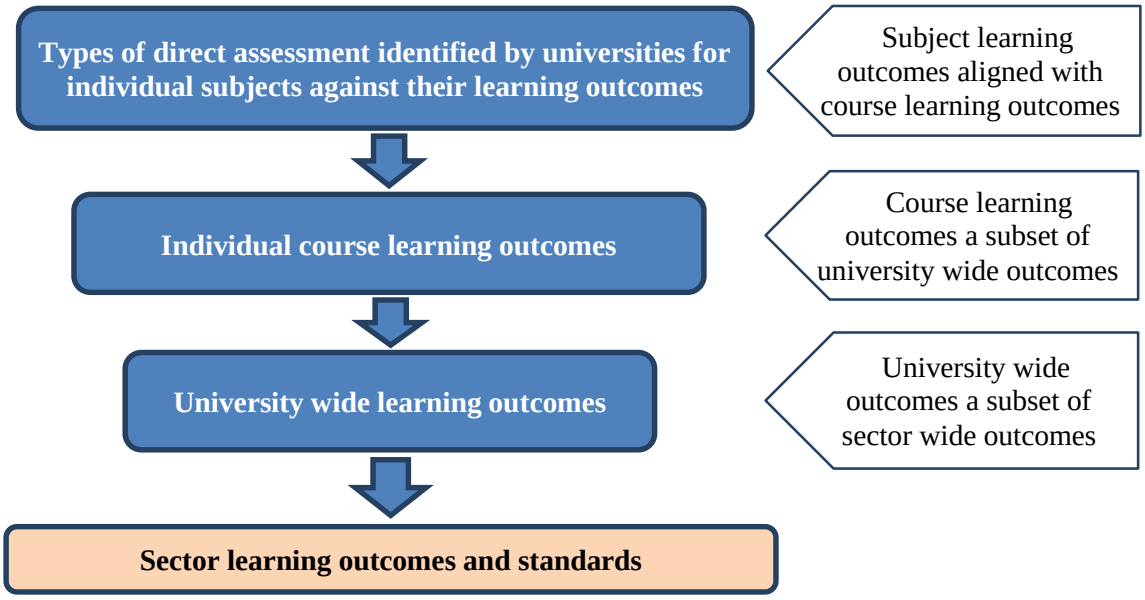
6.3 A general model for the Australian learning and assessment environment

The most common approach to assessment of learning outcomes identified through the survey of Australian higher education institutions involves curriculum and assessment mapping between courses and constituent subjects (11 out of 14 respondents providing details of assessment in the survey) as depicted below. It is a model which builds up a picture of evaluation of the achievement of university or course level learning outcomes through bottom up assessment of attainment in subjects linked to the learning outcomes of the courses to which the subjects are credited. Whole of course assessments were reported separately from this approach by 7 of the 14 complete responses to the data collection. This may be in addition to or instead of the subject alignment model. Four of these indicated what forms of assessment they actually used and all said it applied only in some disciplines. Capstone subjects were mentioned in three of the four cases and the other institution reported the use of viva voce examinations in some professional disciplines.

This model can be used to form a consistent and generalised approach to course or institutional based learning outcome assessment, as has been done with measurement of learning outcomes in some professional course accreditations. Using this approach, the first step in determining a comparable approach between institutions and disciplines is to adopt a consistent way of mapping each course's learning outcomes to the subjects

included in it and to ensure that the learning outcomes being assessed are selected from the common set of sector level learning outcomes specified for that course or level of study. In this way the coverage of content associated with each learning outcome is assured and the level within the course where the relevant material is taught is identified so that progression in the student's knowledge and skills during their course of study can be ascertained. It may be possible then that the type of local assessment tasks applied in these subjects might be developed to ensure consistency in complexity between different institutions for similar learning outcomes and used with the specified standards related to the learning outcomes to demonstrate the level of student achievement.

Figure 6.6 – Bottom up model for assessment of course or institutional learning outcomes in undergraduate Australian higher education



The approach of mapping subject curriculum and assessment to learning outcomes is similar to many of those described in the literature (Bloxham & Boyd, 2007; Keshavarz, 2011). The above model does not prescribe the nature of the various assessment tasks, only the way meaningful tasks aligned with learning outcomes can be used to ascertain whether the learning outcomes have been achieved or not.

The UK Quality Assurance Agency (QAA)'s 2006 Subject Benchmark Statements were an early series of discipline-specific statements of learning outcomes at various levels of

study which were at a higher level than course. The expectation that student achievement would be able to be demonstrated against these statements led to a strong focus on how to assess students meaningfully so that their attainment of the embedded learning outcomes and standards could be demonstrated at discipline and institutional levels. Implementation of this concept of quality assurance included a requirement for each university to prepare a 'programme specification' for each of their courses which specified how its learning outcomes would be assessed (Bloxham & Boyd, 2007) (p159).

Bloxham and Boyd (2007) discuss the validity of assessment tasks in this environment. They advocated use of a six point process of 'constructive alignment' and taxonomies for describing learning outcomes and related assessment tasks to ensure that the two were well aligned, as well as the use of *authentic* assessment tasks (see earlier definitions) as a means of engaging students in both their learning and assessment (p27). Such an approach is similar to that described above as a possibility for linking course and subject based outcomes in the Australian environment and using the set of learning outcomes developed in Chapter 4. Bloxham and Boyd also stress the need to match assessment and learning activities to the learning outcomes not just curriculum, and to identify practical constraints on assessment such as load and time constraints on students and staff and course rules about graduation to ensure that the certification part of assessment is also considered.

While the assessment examples are developed at the subject level as individual tasks the emphasis remains on course based assessment to demonstrate *course* learning outcomes. Arguments are made as to why it is only meaningful to assess graduate learning outcomes by looking at assessment across whole courses (ie a holistic approach), and not at the subject level only. This emphasis is consistent with the aim set out by Barrie *et al* in the AAGLO project. The reasons given for this focus include the greater range of graduate outcomes expected to be achieved in a course of study; the possibility of being able to demonstrate progression over years in the attainment of knowledge and skills; the ability to assess for the promotion of learning; and assessment of learning for comparative analysis.

To gain consistency and comparability of learning outcome assessment results between courses, disciplines and institutions some adaptation of current approaches and tasks is

likely to be required so that institutional performance in terms of students achieving stated learning outcomes can be accurately evaluated. There may be resistance from the academic community to any suggestions for greater uniformity or conformity in the type of assessment activity proposed to ensure comparability of grades because the nature of assessment is central to the teaching process, and these concerns should be carefully considered and taken into account. But a different approach to assessment which is more closely aimed at the achievement of stated learning outcomes may also provide an opportunity for improving student learning and teacher and institutional accountability.

6.4 Possible opportunities and barriers to changes in assessment

There are many recent examples of embedding learning outcomes in the curriculum and developing local assessment regimes which measure achievement of students in institutional or course level learning outcomes but a critical issue for this study is whether such assessment results can be translatable or comparable between different domains even though they may be working to an agreed common sector- or discipline-wide set of learning outcomes. There are also examples of good practice assessment which could be drawn on (Astin & Antonio, 2012; Carless et al., 2006; Ewell, Kuh, & Ikenberry, 2013; Keshavarz, 2011) but trying to find consistent ways to extend these specific examples between institutions is not a simple matter.

There appears to be potential for advancing measurement of common learning outcomes for the higher education sector as a whole if assessment tasks could be generalised in a similar way to the specification of learning outcomes to allow local adaptation but in a way which would not endanger comparability of measurement. This is a controversial idea as there seems to be a prevailing view that assessment tasks cannot be generalised to cover different disciplines or courses and these views are strongly held by practitioners. Perhaps such a generalised approach might be acceptable if particular approaches to assignments which test attainment can be agreed and administered consistently in the context of each institution's courses but graded against a sector-agreed set of standards or rubrics. But as indicated by Sadler the concept of achievement standards is intricately tied up with academic freedoms and identity (Sadler, 2011).

In addition, as discussed earlier, assessment serves many purposes in higher education and to a considerable extent impacts on students' attitudes and motivation to study.

Changing assessment regimes in institutions may therefore have significant impact on the quantum and quality of student learning characteristic of that institution and it may not be possible, because of academic beliefs and different discipline cultures, to translate or generalise assessment tasks in ways which would allow the comparability sought. In any case even if these challenges could be overcome there may be unintended consequences and greater unpredictability to results of such assessment tasks used in performance or quality measurement because of the impact of the changes and the messages sent about the importance of particular assessment tasks on student behaviours in different environments (Baik & James, 2014) (p282).

If these institutional factors were able to be managed, any change of approach to assessment would need to be driven by academic communities of scholars across a comprehensive range of disciplines in benchmarking assessment tasks and couching them in appropriate disciplinary knowledge.

6.5 Feasibility of reliable and valid assessment - what to measure using local achievement assessments

Some information was provided on the nature of the assessment tasks used by institutions responding to the data collection described earlier in this chapter. On the surface there appears to be a lot of similarity between types of assessment used for different learning outcomes across disciplines. The responses received on the nature of assessment tasks used and also examples provided in the AAGLO study at the subject level are summarised in the table in Appendix 6.2 for the full set of 38 learning outcomes, together with a view on their assessability and possible types of assessment tasks appropriate to the standards.

Using the classification outlined by Astin (see Table 5.1), most of the cognitive learning outcomes appear as if they might be directly assessable using standard psychometric means. Some others which are related to non-cognitive characteristics may be measurable using other forms of assessment such as student opinion surveys or observation but these approaches generally would not lend themselves to grading at multiple-levels as suggested in the statements of standards proposed for the outcomes in Chapter 4.

While the data collected on the nature of the assessment tasks in use is broad, it is hoped that the information contained in Appendix 6.2 may assist institutions to move towards

aligning assessment with those outcomes. The full set of 38 learning outcomes appears comprehensive but it may not be a practical set for assessment for the purposes of inter-institutional performance measurement because of the range of approaches used. It is therefore proposed that a pragmatic approach should be taken to answering the questions posed earlier in Section 6.2 about how many of the 38 learning outcomes could and should be measured by selecting a common subset which applies to the majority of Australian universities and which has potential for reliable and consistent measurement based on the model developed by Astin and Antonio.

6.5.1 A measurable common subset of Tier 3 learning outcomes

Based on the information presented in Appendix 6.2 and expanding the Tier 2 learning outcomes to Tier 3 statements, it seems that 23 of the 38 learning outcomes are potentially psychometrically assessable and 15 are assessable by different types of behavioural instruments or a combination of psychometric and behavioural tests. Of the 23 learning outcomes which could possibly be measured using psychometric data, 15 of them were quoted by over 80% of respondents to the data collection request as indicated earlier (see Figure 6.4) and so represent a core set of learning outcomes which might be measurable in some way. Table 6.2 classifies the 38 learning outcomes developed in Chapter 4 by type of assessment data using the Astin and Antonio framework and whether the outcome was in the 'core' set of fifteen outcomes identified through the data collection survey (see section 6.2.3). Based on this analysis it is proposed that the list of ten core learning outcomes which might be assessable with psychological data and form part of the common set for Australian higher education institutions (ie those in the top left quadrant in the table) be listed as the outcomes for which options for assessment regimes could be developed. The remaining five core affective learning outcomes would need to be assessed using other means to give insight into students' behaviour and stated opinions. If this is possible, the form of assessment for these outcomes could provide data which might be used later for performance measurement but in a different way to the psychometrically assessed outcomes.

Table 6.2 – Learning outcomes by assessability and popularity with institutions

	Cognitive learning outcomes **Assessed using psychological data (internal traits of the student)***	**Affective learning outcomes** **Mainly assessed using behavioural data (relating to student's observable activities) ***
In common set (15 outcomes)	• Discipline knowledge • Discipline-related skills • Oral Communication • Written communication • Recognise cultural difference and diversity in work & social contexts • Understand professional ethics • Critical analysis and thinking • Apply ICT to solve problems and in learning tasks • Solve real world problems • Develop and apply international perspectives in discipline (10)	• Learn collaboratively • Participate as team member • Learn independently in self-directed manner (not assess) • Analyse capabilities and create own professional development opportunities • Evaluate own learning and performance in tasks (5)
Not in common set	• Understand theoretical basis of discipline & apply it • Understand, create and evaluate emergent knowledge • Create innovative solutions to problems • Understand need for change and apply knowledge to bring it about • Quantitative communication • Respect human diversity • Understand Indigenous issues • Apply problem solving processes in novel situations • Understand ethical issues and apply principles • Locate, organise & evaluate information from primary sources • Be a well-informed citizen & contribute to community • Understand and accept culturally diverse international environments • Understand financial, social and environmental sustainability (13)	• Lead, plan and play main role in a team project • Apply initiative in setting goals and completing tasks • Demonstrate initiative & responsibility • Manage own time & meet deadlines • Demonstrate openness, intellectual humility, spirit of enquiry • Apply feedback on own performance for improvement • Apply planning and organisation skills • Know regulatory frameworks and political influences • Apply different learning approaches to suit different tasks • Adapt to change and develop personal responses to manage it (10)

* (taxonomy of Astin and Antonio p47-49)

6.5.2 Preparatory steps to measurement

The next task is to explore possible approaches to assessment for this subset of 10 psychometrically assessable learning outcomes. Three options are described and each is reviewed against the assessment principles identified in the previous chapter (section 5.9.2) and its strengths and weaknesses discussed. Each of these options assumes that constructive alignment of learning outcomes and subjects has first been first undertaken for each course on which the assessment regime could be applied. This would establish a matrix for each course taught by the institution at the undergraduate degree level with the 10 core learning outcomes as the columns and the subjects for the relevant course as the rows. The main purpose would be to identify in which subjects the learning outcomes were taught and hence the subjects that would need to have assessment tasks identified to demonstrate whether or not students have achieved the outcomes. The X markers in the cells of the matrix indicate that material is taught in the subject in that row related to the learning outcome for that column. Such a matrix would also identify the year level at which the subject is taught in the course, and the student load (in EFTSL) (ie the proportion of a year's full time study load) associated with an enrolment in that subject. This information would enable appropriate weighted averages to be calculated to provide equivalence in subject offerings between courses and institutions.

The achievement of outcomes that the student would have on graduation will be an accumulation of the achievements in the subjects contributing to each learning outcome from the matrix but it is expected that the highest levels of knowledge and skills gained would be in those subjects which are taught in the higher year levels of the course. Table 6.3 shows what such a matrix would look like.

For a particular learning outcome, assessment tasks could then be designed for the subjects marked with an X in the matrix which allow students to demonstrate achievement in that subject's learning outcomes and hence the course, which can be measured against the standards set for the particular learning outcome as documented in Chapter 4.

Table 6.3 – Example of a matrix aligning course learning outcomes to be assessed with subjects taught using a constructive alignment approach

Course Bachelor of XXX

Subject	Subject Level (year taught)	EFTSL	Generalisable Cognitive Learning outcome 1	Generalisable Cognitive Learning outcome 2	..	Generalisable Cognitive Learning outcome 9	Generalisable Cognitive Learning outcome 10
Subject 1	1	0.125					X
Subject 2	1	0.250	X	X			X
Subject 3	2	0.125				X	
Subject 4	3	0.250		X		X	
…							
Subject n-1	3	0.5	X				
Subject n	4	0.75	X			X	X

This example illustrates how it might be possible to identify achievement of learning outcomes through assessment results of subjects contributing to them, and how those subjects are put together to form the curriculum for the particular course of study. That is, to build a bottom up view of what is learned for each learning outcome specified in the above core list. So provided the assessment tasks are well aligned and appropriate for each learning outcome associated with the subjects taken by students, a view on the achievement of the learning outcomes by the students ought to be able to be obtained.

This approach of constructive alignment shown in Table 6.3 between subjects and course learning outcomes is similar to those advocated by several other groups whose learning outcome typologies were discussed in Chapter 4 of this study. For example, the final report of the HEQCO Tuning project (Lennon et al., 2014), papers by Ewell and his colleagues (Ewell, Kuh, et al., 2013) on how to assess and demonstrate achievement in the DQP dimensions; several other assessment exemplars published by NILOA (Baker, Jankowski, Provesis, & Kinzie, 2012; Blaich & Wise, 2011; Volkwein, 2011) and assessment of the VALUE rubrics by the AACU (Rhodes & Finley, 2013b) all

suggest similar starting point steps to developing an appropriate regime for assessment and a set of assessment tasks for their generalised course or sector learning outcomes which involve knowing where in courses the learning outcomes are taught and standards/rubric criteria by which assessment results will be measured.

Each of these groups of researchers have identified diversity of assessment as a key issue to be addressed in order to make progress on ensuring comparability of assessment and results between institutions for a common set of learning outcomes. The AACU and NILOA work explore the issue of comparability of results based on their own approaches to assessment and in each case trial an approach for their own jurisdiction. The findings of these trials provide some insights into what might work as a feasible assessment regime for the Australian environment, although each approach has limitations and has not yet proved to be a workable universal solution to the problem of measuring course level learning outcomes so far (see for example (Banta et al., 2009)).

6.6 Possible general assessment models

As the earlier discussion of types of assessment used in Australian universities has shown, a wide variety of assignments and assessment tasks are employed in the testing of student achievement in courses of study. However, if the results are to be validly compared across subjects, courses and disciplines, it needs to be demonstrated that the assessment regimes used produce similar grades and results for comparable tasks and levels of difficulty between different cohorts of students. One of the first steps to achieving this comparability across different courses or institutions is to find a way to ensure that the marking and grading systems are comparable and are similarly adhered to. As argued in Chapter 5, for the model of assessment used in this study it is proposed that criterion based assessment be used for all subjects identified through constructive alignment as contributing to the learning outcomes and, for later use in performance measurement, that the grading of achievement be done in terms of the four common levels of standards identified for each learning outcome identified. Over time a student would accumulate a number of grades for the subjects taken which align with the learning outcomes, and it is these grades determined using the criteria in the standards developed in Chapter 4 which would provide evidence of level of achievement of the student in the various learning outcomes.

The data collection undertaken in Australian universities shows that most have advised that criterion-based assessment is the form they already use, although information was not sought about the particular criteria used for the grading of cognitive learning outcomes identified earlier.

The grading at four levels in the standards allows considerable academic judgment for each subject assessed against the specified criteria for each learning outcome. Each of the standards statements identifies the following levels of achievement:

- Not achieved (numerical score 0);
- Developing (numerical score 1);
- Accomplished (numerical score 2); and
- Exemplary (numerical score 3).

This is a very simple scheme and results could be translated into these four grading categories from other grading regimes which are often broadly based around these groups of levels of attainment. A significant amount of change in the way some universities currently process assessment may therefore be required for this approach to enable movement towards comparability of subject (and hence learning outcome) results. This would potentially reduce the current levels of diversity of assessment used and autonomy exercised by academic staff in relation to judgements in the subjects they teach. Such changes would be contentious in the sector and there may be resistance to changing grading schemes and forms of assessment from traditional approaches.

6.6.1 Options for assessment of subjects

Many different approaches are possible for assessment of the learning outcomes as has been seen from the data collection outcomes in Appendix 6.2. Each has its own strengths and weaknesses when considered as a potential basis for between institutional comparisons of student achievement.

Three options for assessing learning outcomes within the conceptual framework of this study have been selected which potentially show varying degrees of comparability of results between courses and universities. The options have been chosen using a framework for assessment which has two dimensions: diversity of the assessment regime used, and comparability of outcomes.

Being able to assess against learning outcomes in a consistent way between courses or institutions and using such assessment outcomes later as a basis of institutional performance measurement is a necessary condition for the pursuit of the main objective of this study. What is required is a practical assessment proposal to address the need for valid cross-institutional comparisons based on common standards related to each desired student learning outcome. So a possible approach to assessment might be one which is easy to implement at a practical level across the sector (ie not too complex or diverse) but which also has the capacity for consistency of application and measurement.

The three options chosen for assessment of subjects relating to the cognitive learning outcomes in this study, which span the spectrums of the conceptual model (Figure 6.7), are based on common approaches used in higher education to varying degrees, based on the analysis of the assessment literature on course learning outcomes:

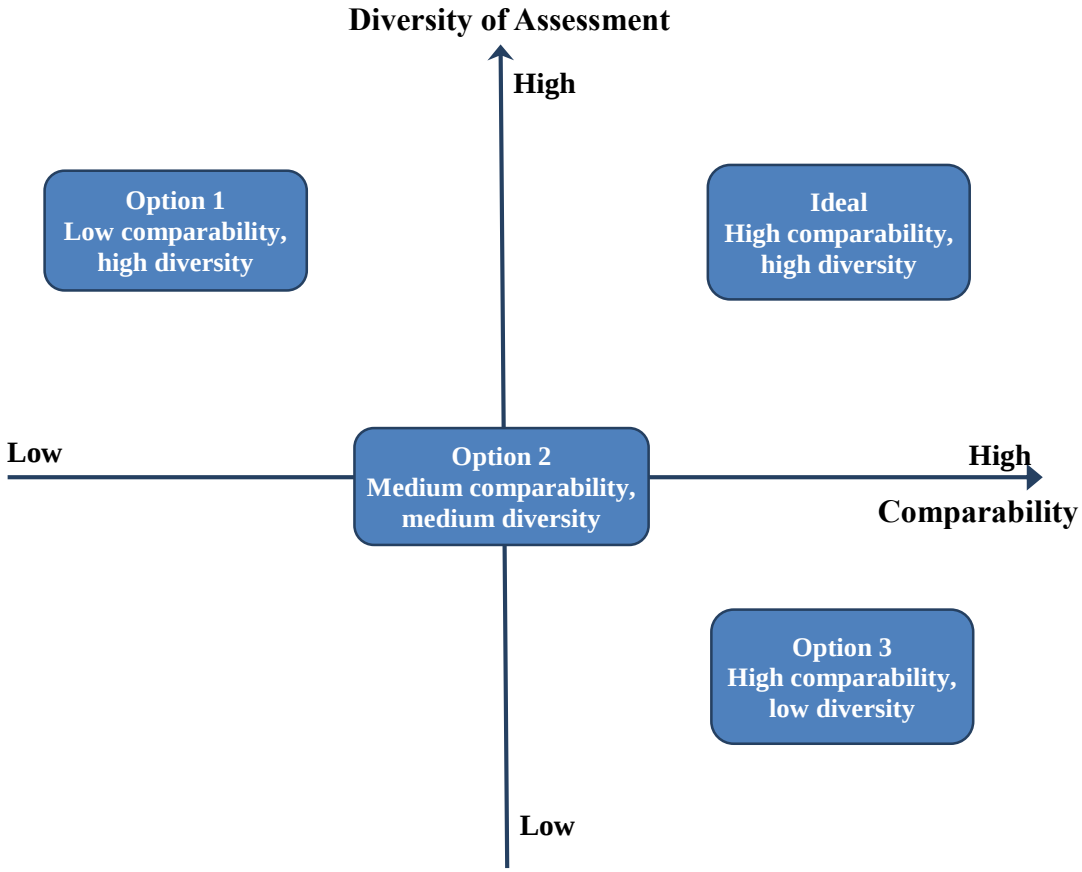
- Institutional assessment tasks with collaborative quality assurance on comparability of tasks and results;
- (e)Portfolios which record results of assessment tasks related to learning outcomes; and
- A common set of assessment tasks (signature assessments) specifically devised to test achievement against learning outcomes.

These are but a sample of the large range of assessment schemes which could be positioned on this model. Their positioning on the grid of the model is shown in Figure 6.7, as well as a possible ideal for the future.

As shown, an ideal assessment regime with the most desired characteristics would be placed in the top right quadrant of the chart having both a range of different assessment tasks and high results comparability for them. The actual current situation is in the upper left quadrant of the model and has a high diversity of assessment and low levels of comparability of assessment results. This is described as Option 1. The portfolio based scheme identified as Option 2 below could be described as ‘feasible’ given that it has potential for both medium diversity of tasks and results. It potentially has some comparability based on its construction but also considerable diversity in its assessment tasks, and Option 3 on the bottom right quadrant would represent an approach which is

based on a relatively small number of assessment tasks linked to learning outcomes but designed to produce consistent and comparable results for students. It is assumed that all of these options would use local assessment achievement tasks as the basis of the assessment schemes.

Figure 6.7 – Conceptual framework for assessment regimes



6.6.2 Option 1 – high assessment diversity and low comparability of measures

The first assessment approach considered is one where the usual institutional local achievement assessment tasks could be used to test student achievement of the particular subjects related to the individual learning outcomes. From what has been seen on examples of assessment tasks linked to particular learning outcomes in Australian higher education (eg the AAGLO project) teaching staff have developed subject assignments which have high content validity and are leading to good learning outcomes. The question is whether the results from these assignments can be used for cross-institutional evaluation of performance given the diversity of content that might

be taught by different institutions and staff. The way to partly ensure comparability is to require that the tasks designed by the academic staff in each institution be closely linked to the learning outcome statements developed in Chapter 4, and achievement of the outcomes be assessed using a common set of standards.

In this option no attempt would be made to change the tasks set by the relevant teachers or to require convergence of different institutions' approaches to assessment through the setting of particular assignments or examination questions. This approach therefore involves considerable diversity and variability both within and between institutions in the range of assessment used, the tasks set and the curriculum content for the subjects contributing to the learning outcomes. From the perspective of the academic staff in the institutions, there would be little change from current assessment approaches. But in order to be confident the results being achieved this way are comparable in standard between disciplines, courses and institutions and that they are well aligned with the learning outcome statements, an overlaid external quality assurance system using various forms of moderation of the assessment process would be needed. If this moderation were to be rigorous enough it might then enable aggregation of these diverse individual achievements to demonstrate accountability and performance of an institution.

Such systems have already been used informally between similar types of universities. In particular the Australian Group of Eight (Go8) universities developed a Quality Verification Scheme (QVS) (Group of Eight, 2011) which they piloted to assure the quality of the Group's courses, assessment and standards. Its aim was to examine assessment in a sample of final year subjects in undergraduate programs and to benchmark grades awarded in similar courses across the Go8 universities. This was a relatively small scale project between very similar universities but the principle might also be applied to a larger group of institutions for which comparability of achievement of learning outcomes is desired.

At present using what is commonly a standardised and similar set of grades across the sector for graduate certification purposes, current accountability systems do not tell the sector what the individual students know or have learned by graduation. Measures such as the GPA or weighted average marks obtained in these tests are assumed to be

indicative of the students' grasp of knowledge and skills specified for the relevant course. Such measures are used comparatively in selection of students for higher level courses without concern but there is no information on whether the GPAs achieved at one institution are comparable to those of another university except in broad form and range of values.

The proposed Option 1 would allow individual teachers to test the knowledge in their own way, while the overlaid quality system in which peer groups of academic staff would review samples of tests for particular learning outcomes from their own and other universities for comparability of difficulty of the assessment and the standard of achievement would provide some assurance that the application of different assessment tasks and the standards produces comparable learning outcomes for the different institutions or courses. This would require sampling of assessment scripts, independent marking and comparison of the outcomes in terms of the achievement of the characteristics of the learning outcome at a specific standard. Such a quality assurance process attempts to bring together incommensurate tasks and raters to deliver more comparable results from different tasks.

This form of moderation of both the assessment tasks and the results obtained by the students (ie the assessment judgements made by academic staff) may in the longer term lead to convergence in design and assessment tasks by the group of academic staff involved in the moderation review but it would be a cooperative activity. Moderation may take many forms depending on the stage of the assessment process being explored for consistency (Bloxham, Hughes, & Adie, 2015) but the general aim is to improve the quality of components of assessment through providing assurance of academic standards. It might relate to the design of assessment tasks or provide calibration between tasks through a shared understanding of task requirements and standards or allow a judgement to be made about the quality of assessment process and actions taken by teaching staff. As envisaged here, the peer group assessment community would build on local assessment developments but be able to verify that assessment tasks were of similar quality and complexity, and assure that the results were comparable for similar student capability. Again the comparability of standards and deep involvement of individual staff members in devising assessment tasks at the centre of this approach would rely on the agreed common standards.

While such a process does not lead to complete conformity between institutions in terms of assessment tasks it makes an attempt, through the use of common standards statements and involvement of academic peers in moderation and quality assurance, to ensure that the moderated results obtained by individual institutions are able to differentiate more reliably between the achievement of individual students than using the raw assessment results.

Strengths and weaknesses of Option 1

The strengths of Option 1 include that there is little interference with current academic approaches used within institutions and the proposed regime; it is likely to gain cooperation from the academic community as the practices are well known within universities; it is likely to promote greater understanding of difference in current practice and may lead to greater consistency of approach between institutions when these difference are understood; and may stimulate collaborative ventures. Change which may be necessary to ensure greater consistency of outcomes between student populations relates to ensuring consistency of rating approaches used and clear identification of how the assessment tasks set by the relevant staff members are commensurate in difficulty across different courses and aligned with the agreed learning outcomes.

The main weaknesses includes the lower probability of reliability of results between institutions in spite of the use of generalised learning outcomes and a set of standards for assessing achievement because different tasks are being considered. Several studies (Bloxham, den-Outer, Hudson, & Price, 2015; Sadler, 2009, 2010) discuss difficulties in achieving consistent marking and grading even when using specific assessment criteria. These suggest that this approach may not deliver the desired consistency of results to be used as the basis of institutional performance measurement. The extent to which these weaknesses impact on comparability will depend on the rigour of the quality assurance on the curriculum design and its linkage to the specified learning outcomes.

6.6.3 Option 2 – Portfolios - medium assessment diversity and medium comparability of measures

Option 2 consists of a mixture of diverse internal achievement assessment tasks and some core tasks relating to specific learning outcomes which would be common or

highly calibrated for achievement. The core set of tasks would be what could be used to measure achievement of the learning outcomes for the course or institution. The device of an (electronic) student portfolio system would be used to store information on assessment employed by the institution. Portfolios have become popular in several universities in Australia and overseas in the quest to enable students to demonstrate achievement in a range of activities during a period of study which correspond to the stated learning outcomes for the course or institution in which they are studying. Their use in digital form, which is becoming more common, allows the collection of information to be used as a digital repository about the student's achievements as well as the assessment tasks themselves. Generally those elements stored in the portfolio relating to assessment are known as *artefacts*. Artefacts may include records of knowledge and skills as they are built up over the duration of the student's course of study as well as the grades achieved and so are good records of formative and summative forms of assessment of a graduate's achievements. They might also be useful in identifying achievements or activities undertaken for the affective learning outcomes described earlier, as well as results of more formal assignments which relate to cognitive learning outcomes. In this sense portfolios have potential to take a more holistic view of assessment in determining which types of evidence of achievement are to be recorded in the portfolio and taken into account as evidence of student learning.

In this assessment option of the use of a portfolio as a vehicle for holistic assessment at the course or institutional level, it is assumed that the artefacts would include pieces of information associated with each of the subjects identified through constructive alignment with the particular learning outcomes identified for assessment. This would still allow considerable diversity in the types of local achievement assessment tasks set by teachers being included in the portfolio, and on the surface has low comparability between institutions. But such an approach could be constrained by each institution including some particular common assessment tasks relating to the core set of learning outcomes to ensure some direct comparability of results as has been the case in some US colleges and universities (Banta et al., 2009) (p 9). Such an approach to assessment would still require considerable involvement of the teaching staff of each university in a similar way to Option 1 but would work towards identifying a core set of learning outcomes and set of assessment tasks for inclusion in the portfolios. To provide some comparability, the standards for assessment of each learning outcome would again be

employed in the marking and grading of the results to be recorded in the portfolios of the students – this would create a form of common currency between institutions and courses for the agreed assessment artefacts but would allow greater flexibility for the remainder of the assessment artefacts included in the portfolio.

Hence option 2 would still allow considerable diversity of forms of assessment in the portfolio but would impose over it a limited core set of assessment artefacts which would be required to demonstrate achievement of the relevant learning outcomes in a more comparable way than the quality assurance and moderation approach of Option 1. Option 2 is consistent with choosing a smaller agreed set of core learning outcomes which should be tested and whose results would be recorded in a similar way between institutions to allow comparability of student performance.

Strengths and weaknesses of Option 2

Some of the advantages of the portfolio approach are that it would result in a more coherent and holistic evaluation of the achievement of the learning outcomes at the institutional or course level and may encourage student learning through reflection and engagement on what they have learned in the process of demonstrating achievement of the core learning outcomes in the portfolio. Because academic staff would also be involved as a group, possibly in the design of the core assessments and in how they could be recorded in a consistent way in the portfolio format, the differences in the nature of the assessments between institutions could be expected to be minimised.

The portfolio approach to assessment as described here is still institution-specific but with an overarching framework for what artefacts are important and may be core in the demonstration of the agreed learning outcomes. The portfolio can contain a diverse range of assessment tasks and outcomes and have potential for greater student engagement in their own learning and assessment and therefore improved learning outcomes by students.

The main weaknesses of the approach include the potential for the lack of coherence in the artefact set; the potential for lack of reliability and reproducibility over time; and the fact that combined information stored in portfolios has not been tested in a quantitative performance measurement environment similar to that desired for this study.

6.6.4 Option 3 – low diversity of assessment and high comparability of measures - a common set of assessment tasks

Option 3 is the most radical proposal of the options considered. It proposes an approach to assessment with lower diversity of assessment tasks but potentially high comparability between disciplines, courses and institutions if it could be made to work in the academic environment. The option is based on work undertaken by the AACU using the VALUE rubrics (Rhodes & Finley, 2013a) which is the parallel to the specified learning outcomes and standards developed in Chapter 4 of this study, for the Australian undergraduate higher education sector. A similar approach was also used in the AHELO project (Richardson & Coates, 2014) as described earlier, though in that case the comparability sought was within a set of disciplines across a number of international domains.

Option 3 would involve defining a limited set of common but customisable assessment tasks for the subjects identified through the constructive alignment process covering the 10 core cognitive learning outcomes identified earlier. The assessment tasks, which would each involve solution of authentic problems with meaning in post-graduation practice, would test a common type of knowledge or skill in the context of the discipline but not be framed specifically for a single discipline. As with Options 1 and 2 the assessments would be marked for the relevant subjects and graded in accordance with the standards and their levels of achievement to provide the basis of comparability and reliability. The tasks so developed would be common to subjects contributing to the relevant learning outcome at the appropriate level and could be supplemented by other local achievement assessment tasks unique to the particular institution, thus maintaining some academic flexibility. Specification of these 'generalisable' assessment tasks would necessarily involve extensive consultation and collaboration between academic staff from a range of disciplines across the higher education sector (a form of moderation of the design process for assessment). The AACU reports that for the trial of a set of generalisable assessment tasks it went through a consultative and testing process with many campuses which had implemented the VALUE learning outcomes and rubrics across a range of disciplines.

This approach and a small set of cross-course/discipline 'signature assessments' developed collaboratively has been tested at a practical level by volunteer institutions in

the US working with staff of the AACU and by rigorous calibration of assessment results being undertaken by academic leaders (Banta et al., 2009; Rhodes & Finley, 2013b). Their conclusions from these trials were that 'campus based case studies have provided considerable supporting evidence for the reliability of the VALUE rubrics' in measuring achievement (p28). Two state-wide efforts in the US to assess student learning in this way are in train. This trial has since been extended in a collaborative venture between AACU and State Higher Education Executive Officers using common assignments and rubrics for assessment and the results reported in Inside Higher Education (Lederman, 2015).

Similar initiatives based on their own learning outcomes classifications and standards are also under investigation in some Asian and European countries (Coates, 2014). For example a similar assessment framework to that described earlier has been applied in Finland (at the JAMK University of Applied Sciences) with criteria developed for different levels of achievement of specified learning outcomes but there has been criticism that in spite of this framework, assessment methods used in the institution are still 'traditional' and competence based assessment methods should be introduced (Ursin, 2014). 'Real-world problems' or authentic tasks set as signature assignments which would allow students to both demonstrate knowledge in the discipline but also to apply generic skills such as problem solving and critical thinking in the context of the discipline might address this issue. Identification of these signature tasks would require considerable collaboration between academic staff in a range of universities and disciplines.

Another project which relates to collaborative development of common assessment materials for medical courses in a similar way to that proposed here, has been undertaken in Australia (the AMAC project) (Edwards & Pearce, 2014). A repository of assessment items (mostly involving online multiple choice tests) created in a bottom-up way through contributions from academic staff in the collaborating medical schools and a consultative approach was used effectively in providing common tasks and enabled comparable assessment and reporting to be undertaken within this discipline across a range of institutions.

These projects appear to provide promising evidence of comparability of assessment results for a generalised approach to measurement of achievement of individual student

learning outcomes which could be applied locally but which might allow reliable and valid inter-institutional and inter-disciplinary comparisons to demonstrate accountability and performance levels.

Possible signature assignments for Option 3

As proposed the description of signature assignments is general but the application to each discipline would be finessed by the relevant teachers of the subjects in which it might be used as was the case in the AACU trial. The type of knowledge and skills being tested and the standards against students would be graded would be consistent across disciplines and only the specific discipline problem would vary. A further challenge in creating generalisable assessment assignments would be to ensure that the authentic assessment tasks are equally difficult between disciplines and appropriate for the level in a courses at which the learning outcome material is taught. The following table provides examples of some possible types of generalised signature assignments which might be used in a similar way in the Australian environment to assess individual student learning outcomes.

Table 6.4– Possible signature assessments for 10 most common assessable cognitive learning outcomes in Australian undergraduate education

Learning outcome	Signature assignment	Other learning outcomes also tested
Discipline knowledge	For a new key problem or issue in the discipline, describe and apply the relevant theories underpinning the problem or issue and describe how to go about addressing it	Written communication
Discipline-related skills	For a skill being taught in the subject which is important for practice in the discipline, undertake a practical example of applying the skill in a real world context relevant to the discipline.	
Oral Communication	Prepare a short talk on an issue related to the discipline and present it to an audience of fellow students and staff.	Organisation and planning
Written communication	Research an aspect of the discipline and write a brief paper on the aspect and how it is relevant to practice of the discipline.	Research

Learning outcome	Signature assignment	Other learning outcomes also tested
Professional ethics	Identify an ethical issue facing the profession/ discipline and develop an argument for a course of action which might be taken to address the issue.	Written communication
Recognise cultural difference and diversity in work	Identify approaches or issues in the discipline being taught which might be differently addressed by people from other cultural backgrounds, and how work practices or instruction could be changed to incorporate these different perspectives.	Equity and social justice Indigenous culture
Critical analysis	For a new issue in the discipline, analyse the relevance of current knowledge to this new issue as taught in the subject, form a judgment on how to address the issue, and defend the position taken in a report or presentation.	Solve real world problems Oral & written communication
Apply ICT to solve problems and in learning tasks	Undertake a project in the discipline which requires use of ICT packages to research answers and produce a report or presentation on findings	Research Critical analysis
Solve real world problems	Report or other written presentation of findings from an investigation of a topic of contemporary significance and relevance to the discipline	Critical analysis Problem solving Research Written communication Planning
Develop and apply international perspectives in discipline	Undertake a project relating to real world practice in the discipline which has been placed in an international context; identify factors which could impact on a successful outcome and which require addressing because of the international context.	Solve real world problems International perspective

Strengths and weaknesses of Option 3

Option 3 is an assessment regime which would require much greater collaboration and consultation up front in the design phase about the nature of the tasks and may require agreement of the academic community to change the nature of some of the assignments set for the subjects they teach. Any changes to assessment potentially have an impact on student learning and the proposal in Option 3 may encounter resistance because it would introduce a degree of conformity not required at present within institutions. Use of such consistent ways of assessing to demonstrate student achievement of the learning

outcomes identified earlier in Table 6.2 as common and assessable may also lead to views of conformity and lack of diversity in the nature of graduates produced through such a system, even though this might be disputed because of the flexibility afforded in the full set of assessment tasks which might be set by the academic teachers. In spite of these limitations, strengths of Option 3 include that the majority of assessment tasks applied in different courses could still be institution or course-specific allowing ongoing diversity in courses and graduate outcomes while there is an overarching common assessment task framework which potentially ensures greater validity and reliability of assessment outcomes between different institutions. The signature assessments would address authentic new problems in most cases and produce directly comparable results as long as there is prior agreement between the collaborating partners on the nature of these tasks. Their development has potential to build cooperation and collaborative solutions for the sector and through the use of best practice examples of assessments that have worked well in some institutions, is likely to improve knowledge and practice of teaching staff in the institutions.

The weaknesses of Option 3 include that it requires acceptance of the philosophy behind the signature assessment tasks; doubt in the academic community that the approach is educationally sound; and it would still require overarching quality assurance mechanisms to ensure that the signature assignments are formulated with the same levels of difficulty of the tasks and rigour and consistency of the marking and grading processes.

6.7 Evaluation of options

6.7.1 Comparison with principles of assessment

The following table summarises the characteristics of each assessment option in the context of the principles of assessment documented in Section 5.9.2.

The principles were framed in the context of the need for comparability of assessment results between different disciplines, courses and institutions for this study. While to some degree each of the options can be seen to be consistent with the principles of assessment, Option 3 appears to offer the greatest potential for comparability of measurement and use in indicators which would measure different performance of institutions in terms of their students achieving the set of learning outcomes. However it

would also require the greatest amount of change of individual teacher approaches to setting assessment tasks and marking and grading them.

Table 6.5 – Comparison of Options 1-3 with principles of assessment

Principles of assessment	Option 1	Option 2	Option 3
Comprehensive and holistic across learning outcomes	Perhaps	Perhaps – depends on selection of assessment artefacts included	Yes for the subset of outcomes
Be part of the routine local achievement assessments	Yes	Yes	Yes
Multiple types of assessment used in the set but set should not be too large	Yes but set is large and diverse	Yes, diverse	Yes, set of assessment tasks is contained
Tasks able to be used summatively and formatively to test progress through studies	Yes	Yes	Yes
Address more than one learning outcome	Perhaps	Perhaps	Yes
Collaborative development of tasks and implemented with significant input from teachers	Yes to some degree through moderation	Yes but to a lesser degree than options 1 and 3	Yes to a large degree
Able to be moderated or calibrated for validity and reliability	Yes but most difficult	Yes but more difficult	Yes
Results transparent and able to be easily communicated to all stakeholders	Not transparent More complex to communicate	Not transparent	Most transparent

6.7.2 Pedagogical issues

There are several important pedagogical issues associated with each of these assessment approaches. To achieve tasks that are comparable in difficulty and marking outcomes Option 1 would require teachers to be subject to peer review from outside their own institution about their preferred forms of assessment and their academic judgements, and to be willing to change the nature of their assessment approach as suggested by their peers. Options 2 and 3 which respectively involve nominating and developing a

particular set of common assessment tasks associated with the specified learning outcomes would require a willingness of teachers to work in teams and agree to use generalised signature assignments in the suite of tests and tasks set in a course of study. Working this way in a networked model of experts 'that draws on the judgments of many' is likely to lead to a more convergent model of assessment of student knowledge and capabilities which reduces assessor variations (Baik & James, 2014), p3). Both Option 2 and Option 3 rely on the establishment of this more formal approach to setting assessment tasks to achieve greater consistency of assessment approach and graded results, but it requires cooperation by the network members and commitment to approaches which may not be their preferred individual positions in relation to assessment.

Options 2 and 3, but particularly 3, are dependent on the feasibility of being able to design assessment tasks, which can allow students to demonstrate required learning outcomes, couched in different discipline knowledge that are of equal difficulty between courses and disciplines. Much of the work done to date on assessment of learning outcomes restricts the assessment scope to disciplines and raises doubts as to whether cross-disciplinary exercises are valid or meaningful (e.g. Turing, AMAC) but the work done by AACU using the VALUE rubrics and much done on the DQP learning outcomes by NILOA addresses this issue and seeks comparability of sector-wide outcomes across disciplines and institutions.

For any one of these options for assessment to be feasible would also require agreement to use a common set of standards as rubrics to assess the students' achievement of the learning outcomes, and the use of particular criterion-based assessment. The standards developed in this study allow demonstration of progression in the development of knowledge and skills and so could be used for both formative and summative assessment of a final result in each subject. By the assignment of a numerical value associated with each level of attainment they lend themselves to being used in the measurement of a group's level of achievement of learning outcomes by aggregating or combining the results of individuals in that group in some way. Academics may see this structured approach linked to the criteria for assessment as limiting their decisions and exercise of professional judgement on the quality of students' work and lead to questioning of any performance measurement based on such an approach.

From the perspective of students' learning, Options 2 and 3 in which some agreed broad assessment tasks adapted for disciplinary differences by a peer group of academic staff would be included in the set of assessment tasks, run a higher risk of teachers teaching to the test represented by these broad tasks and the possibility of using the assessment outcomes to 'game' any performance measures which might be based on these outcomes. This is a common concern with any system in which the contributing data is in the control of individuals within the institution particularly if it is known that 'good' results for any performance measure based on the data is likely to be viewed positively by the management of the institution or may lead to financial or status rewards. As a result if Options 2 or 3 were to be implemented there would be a need to perform an external audit assessment of the quality of the application of the approach within institutions. Changes in behaviour are not restricted to the teaching staff and the implementation of an assessment regime which makes it clear that certain assessment tasks will carry greater weight in the demonstration of learning outcomes will inevitably have an impact on students' behaviour and motivation. This could lead to a concentration of effort by students so that their demonstration of achievement of these outcomes assists them in obtaining suitable employment or further study options rather than gaining a broad education.

6.7.3 Conclusions

As can be seen there are many issues surrounding the use of individual assessment data as the basis for comparing the performance of institutions in relation to their students attaining specified generalisable learning outcomes. But recent advances in other countries discussed here on measuring generalised learning outcomes give hope that these problems could be overcome to some degree and show that clear specification of learning outcomes and standards in the context of an assessment framework may ensure that outcomes from such approaches to assessment are sufficiently reliable and consistent to be used as the basis of institutional accountability and performance evaluation.

None of this approach is easy and there are barriers to such a scheme being adopted in the higher education sector, for many of the reasons outlined. But hopefully some of the ideas proposed in this chapter for the assessment of learning outcomes have potential to advance the sector towards a viable solution for institutional performance measurement. This study now goes on to examine the feasibility of developing some institutional

performance indicators based on an assumption that Option 3 for individual assessment could be developed further and implemented in a range of institutions, creating a data set of achievement outcomes at the institutional and discipline levels.

PART 4: REPORTING OF INSTITUTIONAL PERFORMANCE

Chapter 7 - Measurement of institutional performance in relation to learning outcomes achievement

7.1 The modelling so far

So far this dissertation has built up a model for the measurement of graduate learning outcomes for Australian undergraduate education using the following stepwise process:

1. Identification of a common generalisable set of graduate learning outcomes for degree level study in higher education;
2. Development of a set of standards for grading results of assessment of the learning outcomes to facilitate calibration of levels of student achievement across courses, fields of study or institutions; and
3. Exploration of types of local achievement assessment tasks which have potential to produce comparable results.

The objective of this study – to develop a suite of teaching and learning performance indicators to measure differences in universities' performance in attainment by their students of specified learning outcomes - requires the addition of a further two stages to the above model: the manipulation of quantitative data on individual assessment grades for learning outcomes obtained through stages 1 to 3 above; and the definition of a set of performance indicators based on the manipulated data sets which might validly and reliably differentiate performance between universities within a year and allow exploration of changes in levels of student achievement over time.

None of this is easy to achieve. The model proposed is built up in a step-wise fashion and each component relies on the earlier parts. Its validity as a total concept for measurement of learning outcomes is based on the appropriateness and the strength of each of the underpinning concepts. Whether it could lead the higher education sector to a more viable approach to the measurement of learning outcomes would depend on acceptability to the academic community of the various proposals put forward in each of the parts of the model and a demonstration that each part would be practical to implement in reality.

The underlying motive for seeking to define such performance indicators is to identify the level of attainment of students from each university for each of the specified learning outcomes from stage 1 above and to compare the effectiveness of Australian universities at producing well qualified graduates with the desired knowledge and skills. The performance indicators would be an attempt to measure teaching and learning quality in terms of the learning outcomes specified by the universities as the objectives of undergraduate study. They would provide insights into the nature and quality of teaching and the identification of those institutions which produce the 'best' results in terms of level of student achievement across all courses and disciplines. Because the performance indicators would use the graded outcomes of assessment tasks conducted within each institution using the set of common standards identified in Stage 2 above, they potentially ensure comparability of assessment results between disciplines and different courses. These assessment results measure levels of achievement of knowledge and skills gained throughout and by the end of students' courses of study.

The performance measures described below are direct outcome measures of performance in terms of achievement of knowledge and skills by graduates of institutions by the end of their courses of study. If the indicators developed are robust in terms of reliability of measurement, and they differentiate well between levels of performance, they could also represent a tool for government, institutions and employers to gauge the acquisition of the attributes graduates have from different universities relative to each other, and to reward or incentivise universities to improve their teaching and learning outcomes.

The assessment results obtained through applying the standards from Stage 2 to the tasks from Stage 3 of the model can be used to calculate the proposed performance measures. While a preference was expressed in Chapter 6 for assessment Option 3 as the means of assessing the ten cognitive learning outcomes identified, any of the assessment options canvassed could be used to obtain the base data on individual assessment which would populate the institutional performance indicators. The only issue is the degree of comparability of the results between institutions or courses. For the purposes of the development of appropriate performance indicators what matters is that there is a set of data on student grades available for each learning outcome for a range of institutions or disciplines and that the range of values covered by the data is consistent with the standards developed in stage 2 of the above process. The indicators

developed later primarily relate to the ten cognitive learning outcomes documented in Chapter 6 although some suggestions are also made for some indicators for the five affective learning outcomes using different types of data.

The history of the use of performance indicators in higher education was discussed in some detail in Chapter 2 and the characteristics of a good set of such indicators were enumerated. These are referenced in the development of the proposed indicator set relating to the specified learning outcomes, standards and the assessment approaches outlined in the previous chapter.

7.2 Base individual assessment data

According to the literature examined in Chapter 5 on assessment, a good assessment regime would take a holistic approach to the range of tasks set for students. The aim of the process for measuring individual achievement of learning outcomes was therefore to identify a number of assessment tasks which together represented a holistic approach covering as many as possible of the subset of the fifteen learning outcomes found to be common to the majority of Australian universities. This should enable a comprehensive measurement of achievement of these learning outcomes between different universities.

The five affective learning outcomes listed in Table 6.2 present different challenges and may require different approaches to assessment of institutional performance. Use of surveys of student engagement and other student opinion surveys which provide insight into the collaborative learning, teamwork, and evaluation of students' own and others' learning may be useful in these cases (Astin & Antonio, 2012; Shavelson & Huang, 2003).

In the analysis which follows it is assumed that the bottom-up model described in the previous chapter for the measurement of learning outcomes in Australian higher education (see Figure 6.6) would be implemented so that each university would have a set of assessment task results for each of the specified cognitive learning outcomes for all of its students in each of its undergraduate courses. These results would be graded into four categories based on the associated standards applied in the criterion-based assessment process. For each student the test grade results would be recorded against the particular subjects contributing to each learning outcome as identified through the constructive alignment process described earlier.

Different data sets would be required for the two groups of learning outcomes. The universities which specify the learning outcomes for their curriculum would own the assessment results which contribute to the data files, but the creation of the data files is made more complex because of the need to approach the learning outcomes through the assessment of subjects rather than courses or the institution itself.

7.2.1 Cognitive learning outcomes

The data set for the ten cognitive learning outcomes and all institutions would consist of a file of unit records at the individual student-subject level. This sounds simple but the mechanism proposed for gathering data on student achievement results for the cognitive learning outcomes is complex and multi-layered as shown in Figure 6.6. The aim would be to have a set of records for each student, each of which would contain the achievement score in terms of the standards for the subjects linked to the relevant learning outcomes. Unlike the international measurement of course or institutional based learning outcomes in terms of a customised single test or a small number of course-based assignments, the path between the local achievement assessment for a subject needs to be linked to a learning outcome and a result in the achievement of the learning outcome imputed from the relevant subject results. The results for each student (a score between 0 and 3) would be recorded against each of the particular subjects taken and mapped back through the course enrolment to the chosen learning outcomes associated with the course, and from there to the institutional learning outcomes. Because any data files of results created would need to be manipulated to obtain learning outcome achievement ratings for courses/disciplines and institutions, the file would need to be structured in a way which allowed processing and aggregation at different levels. An example of the data structures and flows associated with this approach is shown in a technical appendix (Appendix 7.1).

There are other issues associated with the aim of measuring the achievement of learning outcomes on graduation through a form of annual data collection. Thinking about building a longitudinal achievement record over time leads to several possibilities for gathering and processing results data. First there is a question about the scope of the data which should be included in such a file. For example, should the file be a retrospective set of records for all of the subjects taken over the whole duration of the student's course enrolment from commencement to graduation which map to the course learning outcomes, or should it be a set of annual records of achievement which builds

up a picture over time where each annual record is in a sense a formative assessment of achievement of the learning outcomes for the student? There are several options for addressing this issue. Each approach has its advantages and disadvantages as discussed below.

One approach would be to identify the graduating students in any year and extract from their previous annual academic records the results for the subjects identified as contributing to the particular learning outcomes in the constructive alignment analysis described earlier. This would produce a record for the student of results in all subjects taken over a number of years which contributed to the knowledge and skills gained from their course of study. However, this would require a long time series of data collected using the approach identified in Chapter 6, which would be complex to compile. Also the final performance indicator result would not be available for many years until all of the subjects were assessed. This approach is not really practical given the lagged nature of the data and the way students' enrolments often vary over time.

A simpler approach which would still model the concept of skills and knowledge learned *by graduation* would be to identify the highest level of subject taken by students for their institution from the constructive alignment contributing to each learning outcome and examine the grade achieved by the student in that subject. This assumes that the subjects contributing to each learning outcome would become increasingly more complex as their level in the course rises, and so the highest level of subject contributing would become the summative grade for the learning outcome. This process would collect data for each currently enrolled student but only use the data on grades relating to the highest level subject for the learning outcome at that institution. The data set would not be comprehensive but would provide some insight into the fial achievement levels of students in each university.

A third approach is favoured which would produce more comprehensive data for all students and learning outcomes. It would not, in a single year's data, specifically address the knowledge and skills attained *on graduation*, but by examining the grades achieved for each learning outcome for the whole population of students in that year, greater insight would be obtained into how the student population in an institution is progressing as a whole towards the achievement of the ten cognitive learning outcomes than either of the approaches outlined above. Time series data in subsequent years

would build up a picture of each student's cognitive learning outcomes achievement by graduation, provided the course structures remained stable. Because this approach is simpler in concept and practice than the above two, and it would still over time give a comprehensive picture of achievement of the cognitive learning outcomes across all subject levels in a course, it seems to be a good compromise for capture of the achievement data for the cognitive learning outcomes which could ultimately populate the performance indicators developed below.

These network relationships between students, courses in which they are enrolled, subjects which contribute to the cognitive learning outcomes content and their assessment tasks are complex and multi-dimensional as shown in Appendix 7.1 and so the data needs to be stored in such a way and be retrievable to reflect these connections. Using the preferred approach documented above, each institution would create an annual data file of grades awarded for specified assessment tasks in various subjects which link through the constructive alignment process to achievement of the ten cognitive learning outcomes.

Such data structures would allow the unit record data of subject results appropriate to each of the ten cognitive learning outcomes to be filtered or aggregated and cross-tabulated to form the base data for the institutional performance indicator calculations.

7.2.2 Affective learning outcomes

As identified in Chapter 6 the five affective learning outcomes common to the majority of universities in Australia are:

- engage in collaborative learning;
- participate as a team member;
- learn independently in self-directed manner;
- analyse own capabilities and create professional development opportunities; and
- evaluate own learning and performance in tasks.

Consistent with the findings of Shavelson, and Astin and Antonio (Astin & Antonio, 2012; Shavelson & Huang, 2003), it is proposed that the data which is likely to be more appropriate to measure this type of learning is obtained from behavioural assessment instruments. These could include various surveys on student satisfaction and student engagement or self-rating by the student of the level of their own learning. This seems at odds with statements made earlier in this dissertation that measures such as good

teaching scales from student opinion surveys are proxies for true performance measures and do not measure what students know and are skilled to do by graduation, as intended by this study. This is still the view but the use of some targeted information on student engagement and experience in terms of the affective learning outcomes might be used to augment the other achievement data on the outcomes of local assessment tasks as is now proposed. Use of two types of learning outcomes data can be stronger than either approach alone, as suggested by Shavelson.

The current QILT (Quality Indicators for Learning and Teaching) indicators published by the Australian Government for its universities and some private higher education providers, mentioned in Chapter 2, may assist in addressing the measurement challenge for these affective learning outcomes. A customised survey of students could possibly provide better information in the longer term but as an interim measure it is worth exploring the type of questions asked in the QILT surveys of students about their learning for relevance to the five affective learning outcomes.

Six QILT indicators are published from the University Experience Questionnaire and some of the issues canvassed and covered by these indicators are relevant to the affective learning outcomes identified in the current study:

- overall quality of the educational experience;
- teaching quality;
- learner engagement;
- learning resources;
- student support; and
- skills development.

Two of these University Experience Survey scales, Learner Engagement and Skills Development, ask students questions which are relevant to three of these affective learning outcomes: collaborative learning; participation as a team member; and learning independently in a self-directed manner. While the responses from students do not represent independent assessment of these learning outcomes in the same way as for the cognitive learning outcomes, the responses do indicate whether or not the student has been exposed to a learning environment which recognises these learning outcomes as

important and how satisfied the student is with the opportunities to experience them through their studies.

In 2015 performance indicators related the University experience were published on the QILT website. These indicators represent the percentage of students who respond that they are satisfied with the particular aspect of their study experience. Answers to questions within each scale are averaged to give an overall score. Not all questions asked in these two scales are relevant to the particular affective outcomes identified here. The relevant questions are:

1. How frequently have you worked with other students as part of your study? (Learner engagement Q2) – collaborative learning
2. To what extent has your course developed your ability to work with others? (Skills development Q3) - teamwork
3. To what extent has your course developed your confidence to learn independently? (Skills development Q4) – independent learning

Using just these questions and scales, performance indicators relating to the three affective learning outcomes could be defined in similar ways to the approach described for the complete scales in the technical note which accompanies QILT (see QILT methodology – UES based indicators – www.qilt.edu.au).

Using this response data for undergraduate students, the data set for the affective learning outcomes and all institutions could be extracted into an institutional file of response data as shown in the technical information in Appendix 7.1.

This file would only contain the student responses to the three questions from the UES relevant to the achievement of the 'collaborative learning', 'teamwork' and 'independent learning' learning outcomes but these responses might be used to populate some simple performance indicators which would augment the set on achievement of the cognitive outcomes. The records of the data file could be aggregated and analysed in the same way as for the cognitive learning outcomes data and should be able to provide average performance data similar to that published for the total scales of the University Experience Survey.

7.3 Manipulating the base data

The data in the institutional files of assessment data described in Appendix 7.1 need to be manipulated and aggregated to allow easy calculation of a set of performance indicators, depending on whether the indicator of interest relates to course, discipline or institution.

If the approach suggested is workable, for the cognitive learning outcomes at any of these levels of analysis, the file of data described in Appendix 7.1 would contain records of each student-subject combination in a course, together with the standards-related numerical value obtained by the student in that subject. These values could be mapped to an institution and a discipline using other reference data in the file. The grade values (between 0 and 3) could then be aggregated into groups for learning outcomes using the mapping of the subjects to the particular outcomes. This would result in a matrix of semi-aggregated results information by course, institution, and discipline, which would provide an indication of the levels of achievement of the learning outcomes obtained through the assessment process.

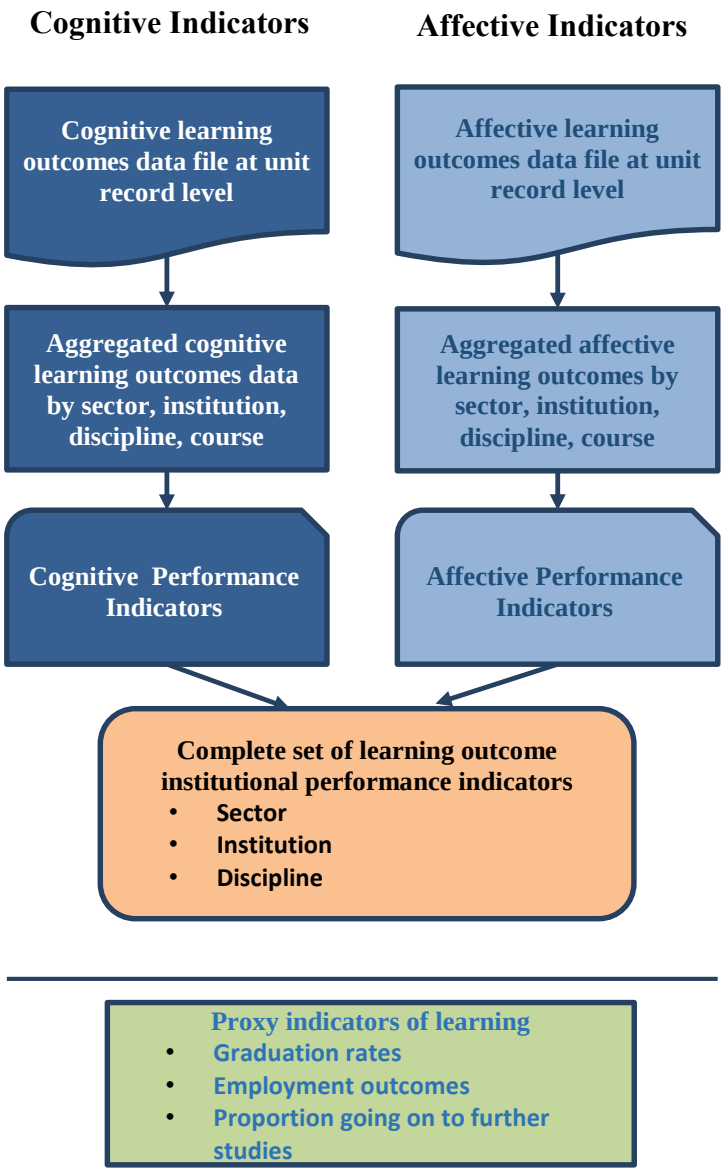
A similar approach could be used for the file on achievement of affective learning outcomes by institution. Instead of using the grades from the standards statements to measure achievement, the values manipulated would be the values of the student survey responses (rated 0 to 5) for the particular questions relating to the affective learning outcomes as assessed above. The student responses would be aggregated into groups representing the particular learning outcome, resulting in another matrix of semi-aggregated survey results by course, institution and discipline, which could then be used to populate performance indicators for each group.

7.4 A speculative set of institutional performance indicators related to the learning outcomes

The framework proposed for the set of performance indicators which could be populated by the aggregated elements of these files is shown below. It is a merging of two distinct sets of indicators for the cognitive learning outcomes and the affective outcomes based on the data which could be obtained from the use of assessment outcomes and QILT survey data respectively, as discussed earlier. These indicators might also be augmented by some of the commonly used measures of learning outcomes related to employment or further study used in the sector. Although they do

not measure what students know on graduation, they do provide a context as to whether the students' learning has equipped them well for employment, and whether their achievements have been successful in the eyes of other higher education stakeholders.

Figure 7.1- Structure of the reporting of performance indicators



All of the data elements in the two files relate to *achievement* of the specified learning outcomes at the time of the collection of the base data, as defined by Shavelson and as opposed to *learning*, defined in a value-added sense. This data reflects the measurement of total knowledge and skills achieved by students during and by the end of their course of study but can also be affected by what knowledge and personal characteristics each

student brings to their study in the first place. Performance measures can be developed based on this data within an institution and can be useful in looking at improvement in achievement of learning outcomes by an institution over time but may be less valid for comparing performance of different institutions in terms of their students achieving the learning outcomes being tested. This may be because the institutional performance indicator values derived from this source may be affected by a range of factors separate to the quality and effectiveness of the teaching that has been delivered to students in the various subjects of their degree studies. This relates to the issue of value-added which is central to US systems' definition of student learning and its assessment as discussed in detail in Chapter 4 on assessment (see 5.4.2). The implications for comparing performance measurement between institutions from such indicators are also discussed in more detail in section 7.6 below.

7.4.1 Cognitive learning outcome indicators

Several performance indicators are proposed for measuring the absolute knowledge learned for the ten cognitive learning outcomes, for different aspects of the data.

Sector and institution analysis

For the sector as a whole or at whole of institution or discipline levels**,** analysis of distribution data by learning outcome, provides basic performance information in relation to absolute achievement by students in the sector. This can be represented graphically as a set of simple distribution plots by grade of achievement or in terms of the standard distribution parameters such as mean scores or percentage of scores achieved in the top standard level for each of these learning outcomes. This may be revealing because it may highlight the normative tendency of grading systems even when criterion based assessment is used, or may show the degree of bias in the grades awarded by institutions under the standards. The proposed indicators are:

KPI 1 cluster: **Distribution of grades achieved by learning outcome**

KPI 1.1: Distribution of grades achieved by learning outcome by sector

KPI 1.2: Distribution of grades achieved by learning outcome by discipline

KPI 1.3: Distribution of grades achieved by learning outcome by institution

These indicators would be presented as plots of percentage of grades awarded at each level for the sector, or discipline, or for each institution. They would provide a quick insight into differences between disciplines or institutions in the application of the standards for each learning outcome specified.

KPI 1.4: Mean grades achieved by learning outcome by institution

For learning outcome i, and institution j, the indicator KPI 1.4 = mean grades achieved

$$= \frac{\Sigma \left(\text{student subject grade results contributing to learning outcome i in institution j}\right)}{\text{Subjects enrolled}}$$

$$\frac{\Sigma \left(\text{student subjects assessed for learning outcome i and institution j}\right)}{\text{Subjects enrolled}}$$

where subjects summed over are those contributing to learning outcome i as identified through the constructive alignment process for all courses at institution j.

KPI 1.5 (a) and (b): Percentage of grades achieved at the low and high end of achievement by learning outcome by institution

This indicator shows the proportion of grades awarded at the highest and lowest levels identified in the standards by each institution, and helps in identifying any biases in assessment outcomes..

KPI 1.5 (a)*:* For learning outcome i and institution j the percentage of grades awarded with score = 3 is calculated as

$$\frac{\Sigma \left(\text{student subject grade results of 3 contributing to learning outcome i in institution j}\right)}{\text{subjects}} \times 100$$

$$\frac{\Sigma \left(\text{student subjects assessed for learning outcome i and institution j}\right)}{\text{subjects}}$$

where subjects summed over are those contributing to learning outcome i as identified through the constructive alignment process for all courses at institution j.

KPI 1.5 (b): In a similar way the percentage of grades awarded with score = 0 (the lowest grade) is

$$\frac{\Sigma\,(\text{student-subject grade results of 0 contributing to learning outcome i in institution j})}{\Sigma\,(\text{student-subjects assessed for learning outcome i and institution j})} \times 100 \text{ subjects}$$

The performance indicators KPI 1.4 and KPI 1.5 could be presented in tabular form as shown in Table 7.1 below for some hypothetical test data.

Table 7.1 – Presentation of hypothetical performance data for cognitive learning outcomes for a single university: KPIs 1.4 and 1.5 by learning outcome and institution

Cognitive Learning Outcomes for Institution X	KPI 1.4 Mean grade achieved	KPI 1.5 (a) Percentage of grades equal to 3	KPI 1.5 (b) Percentage of grades equal to 0
1. Discipline knowledge	1.7	10%	25%
2. Discipline-related skills	2.0	18%	20%
3. Oral Communication	2.1	25%	12%
4. Written communication	1.8	15%	5%
5. Professional ethics	2.5	20%	0%
6. Recognise cultural difference and diversity in work	2.7	25%	10%
7. Critical analysis	1.1	25%	18%
8. Apply ICT to solve problems and in learning tasks	2.4	30%	25%
9. Solve real world problems	2.1	12%	30%
10. Develop and apply international perspectives in discipline	1.9	11%	24%

For KPI 1.4, the higher the mean value the greater average achievement of the learning outcome. In this example, the outcomes of *Professional Ethics* and *Recognising Cultural Differences* are the ones showing highest levels of achievement. Similarly for KPI 1.5(a), a higher percentage reflects a greater level of achievement, and would commonly be accompanied by a lower value of KPI 1.5(b).

The extent to which institutional performance will be differentiable at this macro level will depend on how rigorously the standards have been applied and the size and capability of the student groups in the institutions for the particular learning outcomes. At this high level of aggregation the differences in distributions identified for the individual subject results may be smoothed out and a tendency for normative assessment patterns may emerge.

The variability in results is also influenced by the fact that in this study the number of levels of achievement in the standards is 4 (ie scores of 0-3). As discussed in section 3.11.5 inclusion of more levels in the criteria for the standards makes it more difficult to describe the differences in the criteria used for assessment and there is a resultant tension between this and lower variability in accumulated results.

Institutional and discipline achievement of learning outcomes

There is great diversity in undergraduate degree course structures and the number, type and level (in terms of stage of the course) of subjects which contribute to the various learning outcomes can also vary significantly between courses. While a simple mean of the grades achieved for a particular learning outcome such as defined above for KPI 1.4 for an institution or discipline may highlight differences in broad performance, a more accurate measure would also take some account of the quantity of learning associated with the learning outcome and biases in the grade distribution (Adelfio, Boscaino, & Capursi, 2014). A way of doing this would be to build into the performance indicator the student load values of subjects taken which contribute to the different learning outcomes. This would provide insight into the differences in effort required by students to demonstrate attainment of the standards at particular levels.

Performance indicators related to pass rates of subjects offered by institutions have been in use for nearly 30 years in Australian higher education as a proxy for achievement of learning outcomes. These indicators, known as *subject unit pass rate* and *subject load pass rate*, were first described by Linke *et al* in 1991 and have been in continuous use in the Australian higher education system since that time. They are relatively simple to calculate and the student load based rate has been proven to differentiate well between the performance of different institutions and disciplines. The SPR (student progress rate) is a weighted average of student grade outcomes (1: withdrawn not failed; 2:

passed; 3: failed or withdrawn failed; and 4: incomplete) by the student load in EFTSL associated with the subjects.

A similar approach might be used in framing a weighted grade average for achievement of learning outcomes using the achievement data in the files described here. The following performance indicator shows how this could be defined.

KPI 2 : Student Achievement Score by learning outcome and institution

It is proposed that this indicator would be calculated as follows.

KPI 2 for learning outcome i and institution j =

$$\frac{\sum(\text{student subject grade x subject EFTSL contributing to learning outcome i in institution j})}{\text{Student subjects}}$$

$$\frac{\sum(\text{student subject assessed x subject EFTSL for learning outcome i and institution j})}{\text{Student subjects}}$$

This represents a weighted grade average of student achievement for each learning outcome in terms of the grades used in the standards. The ratio calculated this way takes into account the differences in the contributions the subjects, identified through the constructive alignment process, make to the total achievement of the learning outcome for the institution or discipline. The indicator would have a value between 0 and 3, and the higher the value, the better the performance of the institution. This indicator could be used to compare achievement levels by learning outcome within an institution as well as between institutions for the same learning outcome.

In terms of the production model of performance (van Dooren et al., 2015), KPI 2 relates the inputs of the curriculum and the teaching provided to the outcomes achieved by the students (ie intermediate and final outcomes) from a range of learning activities. It is therefore a measure of the effectiveness of the educational process and environment in delivering results consistent with the learning objectives of the institution and its courses as framed through the statements of expected learning outcomes.

The definition of each of these indicators allows for differences in the level of aggregation of the achievement data and so they are likely to be useful for comparing performance between disciplines and institutions. For example the approach could be

used to compare performance at the discipline level both within and between institutions, and over time.

It is hoped that, in the same way as the SPR calculations in use over a long period of time have proved reliable in distinguishing between different levels of performance these KPIs would also differentiate well between student achievement in learning outcomes at particular levels. But this needs to be tested with real data.

Table 7.2 illustrates what the set of cognitive performance indicators might look like for three universities and how they could be presented for comparative purposes.

Table 7.2 – Presentation of hypothetical performance data for cognitive learning outcomes for a selection of institutions and discipline levels

Learning outcome	KPI 1.4 – mean grades achieved			KPI 1.5(a) - % of grades =3			KPI 2 – average achievement score		
	Inst A	Inst B	Inst C	Inst A	Inst B	Inst C	Inst A	Inst B	Inst C
LO 1	1.7	1.2	2.2	10%	5%	16%	2.3	1.4	2.2
LO2	2.0	2.1	2.2	18%	20%	27%	1.6	1.7	2.6
LO3	2.1	1.8	2.2	25%	19%	27%	1.8	2.0	2.5
LO4	1.8	2.0	2.3	15%	18%	25%	1.5	1.8	2.7
LO5	2.5	1.5	2.6	20%	15%	20%	2.0	1.7	2.6
LO6	2.7	1.3	2.7	25%	16%	25%	2.7	1.4	2.5
LO7	1.1	1.8	2.0	25%	30%	35%	1.5	2.0	2.8
LO8	2.4	2.0	2.5	30%	21%	30%	2.5	1.7	2.6
LO9	2.1	1.6	2.3	12%	10%	15%	2.4	1.8	2.5
LO10	1.9	1.2	2.2	11%	5%	13%	2.0	1.0	2.3

In this exemplar Institution A shows the best student achievement in LO5 and LO6 in KPI 1.4 but when the average takes into account the EFTSL weight of the subjects contributing to the learning outcome in KPI 2, the best results are for LO9 and LO 6. The clustering of the indicator values suggest greater differentiation might occur if more grade achievement categories were used in the standards.

7.4.2 Affective learning indicators

As discussed in Chapter 2, the University Experience Survey report (Graduate Careers Australia & The Social Research Centre, 2015)) contains a set of learning and teaching performance indicators (QILT) derived from student responses to the questions in the six scales of the survey. The responses, which are either levels of satisfaction with the

quality of teaching and learning or information about the level of the student's engagement are on a 5 point or 4 point rating depending on the scale of the question, are pooled within the scale and the results averaged. The QILT indicators are 'the percentage of students who are satisfied with 'the quality of teaching or the learning resources or the support provided by their university', for example or in the case of skills development the extent to which their university developed each of the particular skills mentioned through their course of studies. The percentage of students satisfied is the ratio of the number of graduates satisfied with the aspect of their higher education experience as demonstrated by selecting the two highest ratings in the appropriate scales, and the number of students with a valid response to the UES questions. The UES report pools the response data over the last two survey periods (2013-14) for each of the scales. From the published data it appears that the indicators so formed are robust, consistent over time and distinguish between different levels of satisfaction or experience. However in some of the scales, the levels of satisfaction or engagement show similar values across a number of institutions.

This survey seems to have some similarities to the type of engagement surveys mentioned by Shavelson, and Astin and Antonio as providing alternative types of information about affective learning outcomes. In the absence of a tailored survey on these questions, it is proposed that the three questions from the UES identified earlier should be used to develop three performance indicators. These would be narrower in scope than those included in the UES report but would focus on collaboration, teamwork and independent learning only. The positive responses to these questions provide some indication of the institutions explicitly addressing the particular issues during students' courses of study.

The proposed indicators are :

KPI 3.1 – Collaboration

KPI 3.1 is the proportion of undergraduates satisfied with the level of collaboration with others in their study (ie that they have worked with other students *often* or *very often* as part of their study).

KPI 3.1 = Number of undergraduates who worked with other students *often* or *very often*
Number of undergraduates with a valid response

where undergraduates with a valid response are those who responded to this question (item 4) of the learner engagement scale in each of the 2013 and 2014 surveys.

KPI 3.2 – Teamwork

KPI 3.2 is the proportion of undergraduates satisfied that their course developed their ability to work with others (ie responding that their course has developed their ability to work with others *quite a bit* or *very much*).

KPI 3.2 = No. of undergraduates who worked with other students *quite a bit* or *very much*
Number of undergraduates with a valid response

where undergraduates with a valid response are those who responded to this question (item 3) of the skills development scale in each of the 2013 and 2014 surveys.

KPI 3.3 – Independent learning

KPI 3.3 is the proportion of undergraduates who believe that their course developed their confidence to learn independently (ie responded that their course developed this capacity *quite a bit* or *very much*)

KPI 3.3 =
Number of undergrads whose courses encouraged independent learning *quite a bit* or *very much*
Number of undergraduates with a valid response

where undergraduates with a valid response are those who responded to this question (item 4) of the skills development scale in each of the 2013 and 2014 surveys.

To calculate these indicators institutions would need to extract undergraduate student responses to these three UES questions, compile a file of unit record data and aggregate the responses.

The feasibility of splitting off the data related to these questions and the robustness of the indicators based on the single questions to reflect differences in the focus placed on these learning outcomes by different institutions would have to be trialed for reliability and validity as was done for the QILT indicators, before they could be used with confidence.

The following table illustrates what the affective learning outcomes performance indicators might look like, based on overall results published in the University Experience Survey report.

Table 7.3 – Presentation of hypothetical performance data for affective learning outcomes for several universities: KPIs 3.1, 3.2 and 3.3 by learning outcome and institution

KPI	Institution A	Institution B	Institution C
KPI 3.1 - Collaboration (% of students working regularly with other students)	48%	52%	60%
KPI 3.2 - Teamwork (% of students working in teams with other students)	81%	79%	81%
KPI 3.3 - Independent learning (% of students developing confidence in working independently)	78%	70%	80%

7.5 The complete speculative indicator set

The complete speculative set of performance indicators which fits the framework shown in Figure 7.1 is summarised in Table 7.4.

Table 7.4 – The complete proposed performance indicator set

Cognitive Performance Indicators		Affective Performance Indicators	
KPI 1.1	Distribution of grades achieved by learning outcome by sector	KPI 3.1	Collaboration – percentage of students who worked often or very often with other students
KPI 1.2	Distribution of grades achieved by learning outcome by discipline	KPI 3.2	Teamwork – percentage of students who believed their course allowed them to work with other students quite a bit or very much
KPI 1.3	Distribution of grades achieved by learning outcome by institution	KPI 3.3	Independent learning – Percentage of students who believed their course encouraged independent learning quite a bit or very much
KPI 1.4	Mean grades achieved by learning outcome by institution		
KPI 1.5 (a)	Percentage of grade results obtained at the high end of the grade distribution		
KPI 1.5 (b)	Percentage of grade results obtained at the low end of the grade distribution		
KPI 2	Student achievement score by learning outcome and institution		
Proxy indicators			
KPI 4	Undergraduate graduation rate for institution		
KPI 5	Undergraduate employment outcomes (% in full-time work; % in part-time work; % in casual work; % unemployed)		
KPI 6	Undergraduate further study outcomes (% going on to further study)		

7.6 Appropriateness of the proposed indicator group

The set of six performance measures proposed above for the ten cognitive learning outcomes and the three for affective learning outcomes represent a possible limited suite which could be trialled on real data to establish whether they have the capacity to differentiate reliably between performance of Australia's higher education institutions based on the individual assessment achievement results of students as outlined above.

The indicators discussed here are instruments which have been developed with a degree of pragmatism, and they do not alone fully describe the attainment of students in relation to learning outcomes. But they are likely to be more sensitive, are directly related to what is being taught, and are more informative about specific academic achievement than the broad proxies in use alone at present.

Chapter 2 described the history of the development of teaching and learning performance indicators in higher education and use of a production model of performance development (van Dooren et al., 2015) as the conceptual base rather than the more common output based indicators currently in use to measure the outcomes and quality of learning by students. It also outlined some properties a set of performance indicators should have to be useful and valid measures of learning outcomes (see section 2.5.1). Comparing the indicator set developed above with these properties shows that they mostly exhibit these characteristics. Provided the necessary base data can be collected by the institutions the indicator set could be used as a sound set of measures of institutional performance. The comparison of the indicators with the properties discussed earlier is summarised below and reveals some vulnerabilities in the construction of the indicators.

The indicators should:

- *Be clearly related to the prime functions and objectives of the institution*

The set of indicators has been developed to measure the effectiveness of a university's teaching related to the attainment of a set of learning outcomes (or objectives) for undergraduate Australian students. Such an objective relates to the prime function of teaching and learning which is central to universities' activity.

- *Form a coherent set but be relatively small in number*

The six performance indicators for cognitive learning outcomes and three for affective outcomes form a relatively small set based on local achievement assessment. Together they could provide information about how well students have achieved the learning outcomes identified in Chapter 3 of this study for the Australian higher education sector and reveal how the grading practices vary between universities, disciplines and learning outcomes. Since they are based on actual achievement ratings by academic staff in subjects aligned with the cognitive learning outcomes, the subset of 6 indicators should

be are a coherent set and are quite different to the type of proxy indicators developed in the past and currently in use. However they complement some of these other measures such as graduation, employment and further study rates while providing new information about the amount and level of learning and skills development achieved by students during their studies.

- *Be valid, objective, reliable and verifiable, and be defined and collected in a uniform and agreed fashion*

The proposed indicators are based on a collection of data which are the results of local assessment by academic staff. The previous chapter describes a process which would ensure that these data are collected in a similar way using the set of standards for assessment described in Chapter 3, provided marking and grading of assessments are consistent between institutions. Subject to the extent to which the comparability of outcomes of assessment is valid at the individual subject level, the way these indicators are calculated is reliable and verifiable. The formulae for the performance indicators provide a way for the indicators themselves to be created in a uniform fashion.

- *Have a connection between the theoretical concept to be measured and the empirical variable used in the indicator (ie have content validity);*

The proposed indicators have content validity in that they attempt to directly measure the level of achievement of students for specified learning outcomes, and the empirical variable used in the performance indicators is the grade result obtained by the students during their course of study.

- *Be relative and not absolute;*

The indicators themselves are ratios and not absolute numbers and have the potential to measure the relative achievement of the universities against their goals to improve quality of learning through the achievement of specific learning outcomes. What they are measuring though is absolute knowledge and skills gained by students by the time of graduation.

- *Be calculated over time to demonstrate progress towards the achievement of the agreed target;*

They are designed to be calculated annually over time and increased values of the performance indicators would show positive progress made towards increased quality of teaching and learning. The approach taken using constructive alignment between the learning outcomes to be measured and the curriculum studied by students may vary over time and impact on the values obtained for the indicators. Care would need to be taken that the between-years data obtained through this process is measuring the same characteristics over time to ensure that any changes in indicator values reflect improvements in the quality and effectiveness of teaching and not changes in methodology.

- *Not be 'gameable' – ie their values cannot be manipulated by the institution they are measuring;*

Because the indicators are based on individual grades awarded by teachers to students in subjects which form part of the curriculum of each institution, their values are potentially gameable by the allocation of a large number of category 3 results from the standards. The achievement score indicator would be more difficult to 'game' because of its greater complexity with weightings by subject EFTSL values. The main vulnerability of the approach to gaming is the nature of the assessment tasks chosen, whether they change over time and the distribution of grades obtained across the four categories of the standards against which they have been assessed. This would need to be addressed if the indicators were to be implemented.

- *Focus on what needs to be measured not on what is easy to measure;*

The design of these indicators and the way they are built on the nature of routine assessment used across the curriculum means that they focus on what needs to be measured (ie the attainment of the learning outcomes) not what is easy to measure. There are no special standardised tests given to provide the base data for these indicators because of the embedding of the signature assessment tasks with those used more generally for each institution's courses. The greatest vulnerability would be knowledge gained by students and staff over time on the nature of assessment associated with the signature assessments proposed as part of the Option 3 for

assessment. To avoid this, the particular problems identified as part of the signature assessments should be changed from year to year while retaining the emphasis on the type of assessment used.

- *Not be driven by a desired political outcome*

The use of the indicators for student learning improvement purposes should not encourage behavior which improves the measurement without actually improving the learning and skills attainment which is the real goal of applying the indicators. For example, if the indicators were to be used to allocate funds between institutions or to populate published league tables of student learning then there may be political influences within the sector as a whole and within individual institutions that drive the approach taken to collect the base data used in the indicator calculations. Care would need to be taken by stakeholder groups when determining whether or how the indicators could be used publicly.

7.7 Issues impacting on quality and comprehensiveness of the performance indicators

The proposed performance indicators assume that the data files of individual student results, coded in terms of the four levels of competence in the standards developed for each of the learning outcomes, would be available from each institution, and that the assessments were marked in a consistent way between institutions. This depends on the academic credibility of the proposed assessment approaches in the sector and their acceptance by teachers of the subjects contributing to the learning outcomes. While it is reported that academic staff are keen to be involved in the assessment of course based student learning outcomes and support assessment approaches using routine tests and assignments which lead to comparable data sets of results across institutions, there is some resistance to attempts to devise common assessment approaches such as the signature assessments outlined for assessment Option 3 discussed earlier. There is usually strong association of teachers with their disciplines and a view that it is not possible to generalise assessment across disciplines as is suggested here. The success of defining performance indicators which measure the capacity of students in an institution to demonstrate achievement of learning outcomes associated with their course is dependent on the degree to which common approaches to the definition of standards, assessment tasks and grading systems are implemented across institutional borders.

Given the relatively high levels of teacher autonomy in these academic matters in Australia, the implementation of the approach proposed and its requirement of a degree of conformity presents a significant challenge for the staff and managers of universities and the Government, which may wish to use indicators like these for ranking or funding purposes.

The reliability of the indicators in measuring institutional performance is a function of the reliability of the assessment regime used. There must be consistency of the approach used for assessment between different disciplines and institutions for the indicators to differentiate accurately between the achievement performance of those groups. To be useful as performance measures, the indicators must be able to be reproduced in a consistent way over time. The values of the indicators depend on the structure of the courses students are enrolled in and the subjects taught which contribute to the learning outcomes. The constructive alignment process described would need to be checked from year to year to ensure that the same elements are contributing to the indicator value.

7.8 Do these indicators really measure institutional performance?

7.8.1 Issue of value-added from base learning

The indicators as defined could be used to show whether or not within an institution or a discipline there has been improvement in the absolute levels of achievement of students over time during their studies. In this sense they are measures of teaching performance and effectiveness. But there is a difference between a simple outcome measure of absolute knowledge obtained by the end of an educational program and a measure of program impact on the student's knowledge and skills. (Astin and Antonio, p67). The latter involves dual measurement of the students' knowledge and characteristics at commencement and completion using the same tests. The difference between the two outcomes is the amount of value-added to the student's knowledge and skills through the institution's teaching and curriculum (i.e. related to the institutional performance). The level of absolute achievement in this context may also depend on the personal characteristics of the students at commencement of their higher education courses. These factors include achievement level in secondary schooling; socio-economic status; disciplines studied in secondary schooling; type of enrolment (part-time, full-time, external studies); sex and so on. In order to 'level the playing field' in performance measurement and identify the acquired learning during the higher education course, a

way of estimating the impact of those pre-study factors on the absolute learning indicator would have to be found.

As has been discussed in chapter 4 the definition of learning in some countries is conceptualised as change in knowledge and skills between these two time points in a student's course of study. The way these performance indicators have been formulated using specific statements of learning outcomes and customised criteria for assessment of the learning outcomes during and by the end of study does not lend itself to using the same assessment model for knowledge and skills held on commencement of study. This is unlike the testing that could be done if a standardised external knowledge test were used to measure the learning outcomes as is the case in many countries. However it may be possible to adjust performance indicator scores based on a correlation analysis which takes account of the individual factors influencing the outcomes of learning undertaken. There have been many attempts to estimate value-added attributable to the influencing factors on the outcomes identified above as opposed to teaching and to identify what the 'true' institutional performance is.

7.8.2 Other approaches

Considerable research has been done on the issue of value-added (Lydia Liu Ou, 2011a) using the Voluntary System of Accountability (VSA) in the United States with standardised tests such as the ETS Profiicency Profile and the Collegiate Learning Assessment. Assessment outcomes from these tests have been compared for students at different stages of courses of study without using the same student sample as they progress through their studies – ie a cross–sectional design to the comparative testing has been used. This is not a possible approach using the model for assessment of learning outcomes proposed in this study because the nature of the assessment used varies based on the structure of the curriculum in different institutions and its aligned assessment approaches and the outcomes are meant to be achieved by graduation, not at the commencement or earlier in the courses. More relevant to this study is an approach developed by Sanders and McLean in 1984 on the use of student achievement data as a basis for teacher assessment in secondary schools. This has been applied in Tennessee with a statistical mixed-model methodology to 'eliminate many of the previously cited impediments to incorporating student achievement data in an educational outcome-based assessment system'(Sanders & Horn, 1994) (p299). In the analysis presented the authors discuss the issue of partitioning the impact of teacher and external influences

on student learning outcomes separate from a range of other influencing factors by modelling results obtained in school tests as a regression model with both fixed (student characteristics) and random effects (teacher influence). This approach for the secondary school system showed that the assessment results were uncorrelated with socio-economic and racial characteristics and overcame many of the problems which had been cited as reasons for not doing educational outcome assessment from student achievement data. It therefore appears that there are methods available to strip out the impact of student personal characteristics and levels of knowledge present on entry to study on institutional performance derived from routine assessment results.

A further study is reported in 2014 about measuring value-added in higher education in Texas in the US (Cunha & Miller, 2014). The paper acknowledges the differences between secondary education, where tests of student achievement can be undertaken yearly, and higher education in which similar standardised tests are not generally administered and in which there is loss of students through attrition (p65). In their study the authors use the outcome measures of 'persistence' (or retention), graduation and earnings and correlate these with student characteristics of race, gender, and SAT scores on commencement to investigate the impact of these characteristics on their outcome measures. It is argued that the base conditioning factors (ie the student characteristics) can be used to gain a more accurate estimate of institutional performance without the need to use the same form of assessments in the value-added calculations. Hence it may be possible in the more detailed statements of learning outcomes used in the current study and the performance indicators developed in this chapter to use a similar methodology to achieve comparability of the impact of the institution whose performance is being tested on student achievement of these learning outcomes. The authors warn against using performance measures with allowances for the effect of personal student characteristics to drive funding allocations, to incentivise universities to improve teaching quality, or for ordinal ranking of institutions because of the varying charcteristics and goals of the institutions themselves, and instead urge consideration of establishing benchmarks for performance and thinking in terms of broad categories of performance.

What these examples of attempts to identify the value-added by the teachers and the educational environment show is that care needs to be exercised in use of the 'raw' performance indicators such as those identified in this study, but there may be ways of

making adjustments to the performance measures to allow for the differences in student personal characteristics and provide better comparative measures of institutional performance. More work would need to be done on this were the performance indicators to be trialled with real data. It is interesting to note that in the Australian higher education sector the proxy teaching and learning performance indicators of retention rate, student progress rates, and graduation outcomes have generally not been controlled for the nature of the student intake or other personal characteristics but have been used to allocate funds and to compare institutional performance.

7.9 Possible behavioural impacts on stakeholders

The expectation of implementing performance indicators to measure the achievement of student learning outcomes in different institutions is that such measurement would encourage those institutions to enhance and reform their teaching and curriculum design so that the effectiveness of student achievement of prescribed learning outcomes would be improved and institutional goals and objectives met. However the introduction of new measures sometimes produces different behaviour in the systems to which the measures are applied and these changes in behaviour might lead to spurious outcomes in performance indicator measurement (Henman et al., 2015).

There could also be resistance to implementation of the indicators if there were to be funding adjustments or publication of the results generated by their application to the sector, and it is important to understand any patterns of behaviour which may compromise the usefulness of the indicators.

In the case of the indicators defined in this chapter, there would be some pedagogical issues related to the assessment regime used to collect the individual base data on achievement. One of these factors would be the tendency to describe the standards and their associated criteria for grading in normative terms even though the assessment forms recommended are criterion based. It is important for the legitimacy of this approach to not assign grades based on students' performance relative to their peers but to make a holistic judgment on achievement based on the rubrics and criteria used in the standards. This would be expected to result in a set of grades in terms of the standards which does not resemble a normalised distribution. A risk is that there may be differential and inappropriate use of the standards criteria by teachers which could be inherent in the collection of the data.

If the signature assessment approach (ie Option 3) were to be used as the assessment regime for the learning outcomes as described in the previous chapter, it would make the students more aware of the nature of the testing they would have to undertake to achieve the particular learning outcomes. This may encourage students and staff to focus more clearly on these particular assessment tasks because they know these results are to 'count' in quite explicit ways in the performance indicator calculations, in terms of the performance of the students and their teachers. There may be a tendency in these circumstances for teachers to teach to the assessment test if the outcomes are published or funding is linked to performance, as has been found in some of the achievement testing in primary and secondary education.

For a consistent and reliable implementation of such indicators it would be necessary to address these possible behavioural changes with remedies and conditions on use of the outcomes data such as regular auditing of the base data and the way the assessment tasks are applied between different institutions.

The purposes of undertaking this investigation of the possibility of developing more meaningful institutional performance indicators of what students know and learn during their university studies included providing greater accountability to government for resources spent on teaching and learning and greater clarity about what distinguishes particular graduates to prospective employers. If there is confidence in the reliability of the performance indicators, governments may tend to simplify the indicators or not note qualifiers which the institutions may place on the indicators and the use of their data.

However, no set of performance indicators will fully describe the achievement of learning outcomes by students. There will always be a range of contextual factors related to the universities and their student profile which impact on the performance measured in these simple ways. However, because they have been designed around desired statements of knowledge and skills and their values relate to a proposed set of standards, the indicators and their scores should convey more about student attainment than other output-related indicators inaccurately cited as outcome measures which have been in use now for over 20 years.

PART 5: BRINGING IT ALL TOGETHER

Chapter 8 – Bringing it all together

8.1 Overview

The ultimate objective of this study was to develop a set of institutional performance indicators which would accurately measure what students had learned and could do as a result of their degree studies in higher education. This aim was based on a premise that it might be possible to develop a generalised solution to the problem of specifying and measuring course, discipline- or institution-based learning outcomes for graduates in Australian higher education, as was being pursued in many other countries.

It was an ambitious project given that the issue of measuring course or institutional learning outcomes has been a subject of active interest for about 20 years in the higher education sector. In that time there has been no solution for individual learning outcomes assessment in spite of some large international projects attempting to do so, and no attempt to measure institutional performance in terms of knowledge and skills gained by students. There has been much discussion of the need to measure achievement of learning outcomes *in a comparable way* between institutions and disciplines as is required for benchmarking of performance, but no study undertaken in this area to date has been able to achieve this in any sustainable sense. Also, little work has been done on trying to use various forms of local assessment as the basis of individual student achievement measurement because of the diversity of these local assessment regimes.

There has however been increased interest in the comparability issue internationally in the last two years and there have been several new projects undertaken in uniform measurement for a range of disciplines within universities (Coates, 2014). Two large international project trials in Europe and the United States are in train at present which are attempting to address this problem of comparability more generally and which use routine student assessment as the basis of the comparisons (Measuring and Comparing Achievements of Learning Outcomes in Higher Education in Europe' (CALOHEE) sponsored by the European Commission and the Tuning Academy; and pilot testing of a model 'that is rooted in campus/system collaboration, in authentic student work, and in faculty curriculum development and teaching activity' by the State Higher Education Executive Officers Association (SHEEO) and the Association of American Colleges

and Universities (AACU)) (International Tuning Academy, 2016; State Higher Education Executive Officers, 2015). The SHEEO project has reported some comparisons of student outcomes between institutions which suggest that the model posited in this study may have some potential for reliable measurement of individual student achievement of learning outcomes.

While the intractability of the general research problem over a long period suggested at the outset that a complete broad solution was unlikely to be found to the measurement questions in this study, it was hoped that the proposed model articulated in this study (see Chapters 6 and 7) could be applied to Australian higher education with some useful results. The scope of the study was therefore narrowed to defining learning outcomes and trying to measure them for graduates from Australian undergraduate education.

This chapter reflects on the potential application of the model and the outcomes of the research study. It explores what is possible and what is not for each of the building blocks of the model, and the impact these factors might have on theory and practice in assessment and institutional performance measurement. The feasibility of national application of this model is discussed and an indication given of which components of it might be supported in the sector informed by data collection on learning outcomes in use in Australian universities.

8.2 The model

The study began with formulating a staged model for specifying and measuring learning outcomes as outlined in Chapter 2 which was based on a review of the literature on higher education learning outcomes and their measurement and an examination of approaches taken internationally and in Australia. Each of the building blocks of the proposed model, which is shown in Figure 5.1, has areas of contention amongst the higher education academic community about whether such an approach is theoretically and educationally sound or could feasibly be implemented.

The model assumed a top down structure starting with the specification of desired sector learning outcomes and how those for an institution or course might be selected from this larger set. This was followed by an examination of how the particular learning outcomes might be assessed for individual students, and then how these results might be aggregated or combined to populate a set of institutional performance indicators related to student learning. Each of the model steps involved discrete strands of research with

each component based on previous conclusions from the prior step of the methodology. Within each of the building blocks of the model, the research involved both the examination of past and current theory and practice relating to that step and then an attempt to use this information to derive outputs from the application of the model to the Australian higher education sector.

8.3 What is possible – application of the model to the Australian environment

The research outcomes and strengths of the model which arose when applying it to the Australian higher education environment are discussed below under each of the building blocks which comprise it.

8.3.1 Specifying learning outcomes

The first component of the proposed model involved enunciating a discipline-independent set of learning outcomes for the higher education sector from which universities might be able to select particular outcomes for their own institution and courses. Most educational practitioners believe that the discipline context to specifying learning outcomes is of vital importance and there has been greater success in measuring learning outcomes for particular disciplines (for example various professional accreditation studies or other projects in which assessment exemplars are identified cooperatively between a group of discipline practitioners) than in generalist solutions. Three facts related to learning outcomes and their measurement in the Australian context were identified which encourage a counter-view that it may be possible to find a generalised approach. The first is that Australian universities rarely specify their graduate attributes, competencies or learning outcomes at a discipline level or by level of course and so in practice there appears to be no differentiation by discipline between the groups of aspirational graduate learning outcomes which the model aims to measure. Second, when discipline-based research studies are undertaken independently, comparison of the outputs seems to suggest that there are very few differences between the learning outcomes for different disciplines, their associated standards and the range of assessment tasks which could lead to their measurement. Third, several international attempts to define and measure learning outcomes have led to discipline-independent outcomes, rubrics and assessment tasks (e.g. the Canadian HEQCO system, the VALUE rubrics of the US, and the KoKoHs initiative in Germany). Based on these observations it was considered appropriate to pursue trying

to find a generalisable set of learning outcomes and assessment tasks as identified in the model in order to see whether it was possible to find a single approach to measurement of graduate attainment in Australia. A generalist approach may also be more appropriate to increasingly common multi-disciplinary studies in which learning outcomes are not specific to a particular discipline.

The analysis of graduate attributes specified by all Australian universities discussed in Chapter 4 identified a great deal of similarity between institutions. This led to the development of new four-tiered conceptual framework for the specification of learning outcomes and standards for assessment starting from the idea of *graduateness* (or what qualities society expects to be the characteristics of a graduate). These broad traits derived from the statement of graduateness were then converted into detailed statements of learning outcomes that would demonstrate structural and semantic properties recommended in the literature based on Bloom's taxonomy. This approach yielded a set of 38 sector-based learning outcomes (at Tier 3 of the conceptual framework) of which 8 were among a common group used by the majority of the universities. The opinions of universities on the proposed set of learning outcomes for undergraduate education in Australia were sought through the survey of universities conducted to capture data on assessment practice.

There was wide acceptance of the sector-wide learning outcomes developed in the study. Institutions were already using such learning outcomes across a range of disciplines as well as at the institutional level. This provided enough comfort that the full set of learning outcomes developed and the four tier conceptual model on which their derivation was based were plausible outputs from the first building block of the application of the model to the Australian environment. While this research showed that it was apparently possible to identify meaningful sector-wide learning outcomes for Australian higher education, it also revealed that there were considerable discrepancies between what Australian universities specified in their promotional materials and on their websites about their practices in relation to the specification of learning outcomes and assessment, and what they actually did. This highlights considerable differences between the espoused approaches of universities and those used in practice.

8.3.2 Criterion-based assessment and standards

The second building block of the model is the nature of assessment which might be used to obtain information about students' levels of achievement of the agreed set of learning outcomes. Application of this section of the model would depend first on acceptance by the sector of the principle of criterion-based assessment for the tasks used to assess the learning outcomes and then the use of uniform quantitative grading for each of the criteria categories. There appeared to be no issue with the use of criterion-based assessment as shown by the majority of universities' responses to the data collection survey, but universities expect to define their own criteria for their own assessment tasks.

The standards (or rubrics used in the assessment of learning outcomes) proposed for this step of the model were also generalised in nature and were designed to address the particular wording of each of the learning outcomes developed in Chapter 4. Together the learning outcomes and the standards represent the basic scaffolding underpinning the measurement of achievement in the model. Therefore the descriptions of the learning outcomes represent the criteria against which students are to be assessed and the standards work as descriptors for the grading of achievement.

It was difficult to develop meaningful descriptions of standards across the four achievement categories chosen in the application of the model that did not impose normative characteristics on the grading ratings. Keeping the standards statements reasonably broad and few in number was intended to allow academic raters to continue to form holistic judgements about each student's capabilities in assessment tasks as happens at present in the assessment of higher education students. And yet the comparability of learning outcome achievement between different universities would depend to a great degree on the clarity and usefulness of any standards developed. The need to allocate numerical scores to each category of standards was essential in applying the model, if the development of quantitative institutional performance indicators was to be possible.

The set of standards developed here represents one suggestion only but the general approach appears valid and implementation plausible. There would be considerable debate about whether the number of categories used is workable and whether the descriptions could provide sufficient guidance to raters to produce the level of

consistency of grading of assessment tasks required for any set of performance indicators to produce reliably differentiated values for different institutions. A range of other alternatives should also be investigated.

While the level of acceptance by universities of the principles of criterion-based assessment and the standards as outlined suggest that the second step of the model could possibly be implemented in this country but there would need to be further engagement with the academic community to settle on the precise descriptors of the grades. The standards as specified should therefore not be regarded as a conclusive solution to the issue of achievement measurement. The effectiveness of a four point scale for the standards should also be evaluated by applying the performance measures proposed, to see whether the scales allow sufficient differentiation between institutions' student achievement levels.

8.3.3 Individual student assessment

While there was a lot of commonality between graduate attributes proposed by universities at the institutional level, the specification of learning outcomes was rarely done for courses and most activity was located at the subject level, where they were also assessed. This absence of course learning outcomes meant that the top down model originally proposed for specifying and assessing achievement of the outcomes and which had been used overseas did not really relate to current practice in the Australian sector. Instead, a bottom-up model for measurement and a way to build up a picture of whether the learning outcomes specified for a university had been achieved, were needed. This was done by mapping student enrolments in subjects to courses and then to learning outcomes, and using subject assessment results derived from the standards as base data for a set of performance indicators. In order to measure achievement by individual students of 'course-based' learning outcomes required working back from the subject enrolment to the course and to the learning outcomes themselves. This bottom-up model is shown in Figure 6.6, and has the same steps as the approach contained in the model in Figure 5.1 but the steps are reversed in order. In the end the bottom-up model is potentially able to describe how the agreed learning outcomes could be measured but in a more complex way than for those countries and institutions able to specify general sector-, institution- or course-based learning outcomes.

The assessment of student learning outcomes is a critical success factor for the main objective of this study – institutional performance measurement. For this reason the discussion of different types of assessments, their characteristics, history and current regimes used locally and internationally has been presented at some length in Chapter 5 to provide a theoretical and policy context for the development of a potentially feasible approach to learning outcomes measurement. This was and remains the most challenging aspect of the study. In trying to form a view of possibilities based on experiences overseas and within Australia a conceptual model for the different forms of local assessment appropriate for cognitive and non-cognitive learning outcomes and some principles for assessment were developed that were relevant to the particular objective of developing indicators of institutional performance (see Sections 5.9.1 and 5.9.2). Both of these aspects can be questioned. Three options for an approach to individual assessment were considered, with two of them close to existing regimes, but the most controversial of the proposals is represented by Option 3. This option, that it might be possible to set a relatively small number of 'signature assessment' tasks appropriate to particular subset of the learning outcomes is untested in Australia. The signature assessments would be framed in terms of 'real world' or 'authentic' problems in a discipline context. A data repository could be established to contain these assessments which could be accessed by all institutions. Such an approach of storing exemplars of assessment for sharing between institutions was used within the AMAC project for medical schools assessment (Schuwirth et al., 2014; Wilkinson, Canny, Pearce, Coates, & Edwards, 2014). The success of a similar but more general approach would depend on this being able to be done consistently and at a similar level of complexity for all disciplines and for the results to be graded according to the set of common standards.

It was envisaged that these signature assessments would be a small core part of the full selection of assessment tasks for the subjects in a course of study contributing to the learning outcomes and so would not force rigid uniformity on the sector or the teacher. This addresses concerns expressed when particular assessment approaches or tests are mandated. Examples of possible signature assignments for ten of the core cognitive learning outcomes are given in Table 6.4. The trials underway in Europe and the US of related approaches of specifying some generalised assessment tasks at the course level

claim to be able to demonstrate comparability of measurement between disciplines and institutions, and are showing promising results.

This research has shown that currently in Australian higher education assessment of course or institutional based learning outcomes as envisaged in this component of the original model does not often occur even though some studies have stated that it would be desirable for universities and the sector to develop such assessment approaches (e.g.(Barrie et al., 2014)).

In summary, the approach outlined to assessment of individual student learning using local achievement assessment appears to have face validity but the concept of signature assessments has not yet been proven. It has not been fully tested due to lack of data.

8.3.4 Institutional performance indicators

Provided the base individual assessment data discussed above could be sourced from universities, the formulation of the speculative set of performance indicators for the final building block of the model is possible. The indicators for the cognitive learning outcomes are mainly related to the distribution of achievement data obtained by applying the standards to the grading of the signature assessment tasks.

Because the value of the performance indicators developed for the cognitive learning outcomes is so dependent on the distribution of the grades awarded using such a scheme, it seemed appropriate to define some basic indicators in terms of the parameters of those distributions. A new indicator is also proposed for the achievement of the cognitive learning outcomes. This is the *student achievement score* which is a weighted average (by student load in EFTSL) of the grades achieved from the standards in assessment tasks associated with the learning outcomes. Based on past experiences with student progress rates defined in the late 1980s in terms of four categories of academic outcomes, such an indicator could be expected to differentiate well between the performance of universities provided the standards were applied honestly and rigorously by the institutions in the assessment of student learning. Again because of lack of availability of test data no definitive position on the robustness of student achievement score as a differentiator for institutional performance could be determined.

While obvious proxy learning outcomes indicators such as attrition rates, student opinion about the quality of teaching and graduation rates have earlier been discounted

in this dissertation because they are not outcome measures, the cognitive indicators developed can in one sense also be regarded as surrogates for what this model is trying to measure. This is because they aim to use student performance data to draw an inference about teaching quality and performance for the university as a whole and in this sense could be regarded as proxy performance measures for teaching quality. Nevertheless this measure would provide information not previously collected on a large scale about what has been learned by students and how well it has been learnt during their courses of study.

The approach proposed for determining possible performance indicators for the affective learning outcomes and the use of the University Experience Survey data from the QILT project was based on the research work done on assessment of learning outcomes by Shavelson, and Astin *et al.* The survey questions used to calculate this part of the performance indicators set are those included in the student experience survey, and they have not been specifically tailored to the affective outcomes identified in this study. They also rely entirely on what students say about their own experience rather than any outside perspective on the attainment of the three affective outcomes considered for use of the Survey. While student peer and self-assessment is becoming more common the UES data is not the total answer for assessing these affective learning outcomes. A more rounded solution would be to do what Shavelson or Astin both propose and devise a customised survey which may include teacher, student and peer input for some of the affective learning outcomes. The indicators included in the speculative set relating to this category of learning outcomes represent a starting point for further discussion in the context of the range of student surveys currently administered in universities and what might be possible in the future.

This review of the model to measure learning outcomes and develop a set of institutional performance indicators has discussed what appears to be possible for the Australian higher education sector. The analysis presented is a form of proof of the concept of measurement embedded in the version of the model shown in Figure 6.6. With the adjustments made to specify and assess course and sector learning outcomes through the assessment of subjects, it seems to be robust enough to be implemented in the sector and to produce reliable performance data.

Hence it appears that much of what was proposed in the model is theoretically possible to implement in Australia's universities. But the application of the model also raises some unresolved issues and limitations.

8.4 Issues and limitations of the model

The greatest limitation of this study is that it has not been possible to adequately test the model on real achievement data because of the absence of such data in institutional data repositories. This is the current situation but it is possible that recent developments in data mining of digital information on assessments used for student portfolios or certification of studies undertaken through MOOCs will be a future game changer for storage and retrieval of information about individual student assessment of learning outcomes. In any case face validity of the model can been established and attempts have been made to prove the concepts embedded in the model by outlining practical outputs that could be derived for Australian universities for each of its building blocks.

Each step of the model is dependent on the feasibility of the application of the preceding ones. The model will therefore fail if any of the building blocks cannot be applied appropriately. Adoption of some of the proposals presented could require significant changes to the way subjects of study would be taught and assessed in the universities. The implementation of standards statements and grading rubrics of the type developed in particular have potential to cut across the way assessment has traditionally been carried out in many universities.

8.4.1 Discipline context

The main areas of contention relating to the model are the use of generalised approaches to both learning outcomes specification and their assessment. The application of the first part of the model on specifying a set of learning outcomes for Australian undergraduate education made use of the findings of a number of discipline based projects funded by the Australian Learning and Teaching Committee (ALTC). Comparison of the learning outcomes and associated standards produced in these projects suggested a great deal of commonality between the learning outcomes derived by the different project groups, and so there seemed to be some basis for developing a generalised learning outcomes set which might apply to all disciplines. These similarities in expression and content of the learning outcomes were highlighted by using a generic descriptor of 'discipline' in place of the specific field of study label.

When this was done there were few differences between the outcomes specified across disciplinary groups. But this does not guarantee acceptance by the academic community in the sector to this semantic approach to the statements of outcomes. Further discussion about this needs to take place and the proposed generalised learning outcome set trialled in universities. It is said that academic staff associate with a discipline as the primary driver of their approach to pedagogy, although this seems to be less the case for younger academics who are more involved in multi-disciplinary studies. The issue of what differences would emerge in the generalised learning outcome set if it was formulated from the discipline perspective or how the elements would differ for specific disciplines should be more fully explored. This remains a potential vulnerability of the model.

8.4.2 Practical application of the standards

Pedagogical issues arise in the approach to standards formulation in the second component of the model. These include whether the grades of 0 to 3 are meaningfully described and sufficient to allow assessment which is useful for both student learning purposes and for certification. There is a risk in having only a small number of grading categories and broad statements of attainment in the standards which are to be used in performance measurement in that it is easier to 'game' the indicator values by placing a relatively high proportion of the class in the top achievement category, as noted in the previous chapter. The design of the institutional performance measures could also mitigate against this.

On the other hand, questions can be raised about whether the proposed standards are too normative reducing the benefits of using criterion-based assessment and grading, and whether the descriptions are clear enough to enable consistent marking of assessment tasks against the criteria of the learning outcomes to which they refer. Continuing academic concerns about these issues would mean that it may not be possible to implement the model successfully on a national scale as framed even though they appear to have potential for implementation.

What could well be seen as an imposition of a set of rubrics or marking standards through the development of a set aligned with the learning outcomes raises issues about intrusion on the pedagogical autonomy of individual teachers in designing assessment and deciding how to mark and grade any assignments set. Such concerns could be partly addressed by teachers being able to use a sufficiently diverse range of assessment

tasks, some of which would be the tasks which the teacher would normally set specific to their particular subjects and institutions, but which also would include a core set of tasks such as the signature assessments which could be graded against the standards specified. In this way the generalisable signature assignments would only be a small part of the whole assessment students undertake in a subject of study but they would be an important common core for measurement across disciplines.

8.4.3 Assessment issues

A limitation in the application of the model to Australian higher education is that not all of the 38 learning outcomes identified in Chapter 4 are included for assessment. So at best this model would produce data on student achievement of a subset of 15 of the 38 learning outcomes derived by application of the model. This raises the question of whether the smaller common set of fifteen learning outcomes selected is sufficiently comprehensive to provide an accurate measure of achievement of learning outcomes by all institutions. Information provided by universities responding to the data collection survey conducted during the study supported a view that the subset was quite comprehensively used but nevertheless it does not represent a complete exploration of the learning outcomes derived using the conceptual framework discussed in the first building block of the model.

The issue of comparability of local assessment tasks done by individual students, which may be idiosyncratic to a particular university, is central to its success. The way it can best be used to allow valid comparison between students' collective performance in different universities is critical to the formulation of the set of performance indicators. Three options for types of assessment regimes in the application of the model were considered and all have advantages and disadvantages as discussed in Chapter 6. The feasibility of using local assessment achievement of the generalisable type represented by the signature assignments (Option 3), which could provide greater consistency of approach, remains a contested issue with no clear conclusions in the study. This issue is likely to have the greatest impact on the feasibility of a model like the one proposed in this study being accepted and implemented. The acceptability of the signature assessment concept should be research-driven by the academic community and any actual assignments developed cooperatively by communities of scholars whose expertise ranges across broad disciplines of study. There is no obvious way forward for this proposed method other than by getting universities to work together to build on

what has been done in other learning outcomes projects in a range of countries, to resolve some of the remaining issues and to design a set of assessment tasks collaboratively across disciplines which would test students' knowledge and skills. Such a collaborative approach would also build academic capacity in relation to assessment and through a trial of the model would also provide an opportunity for teachers to interact and build approaches with colleagues from other disciplines and universities.

8.4.4 Performance measurement issues

Finally, issues associated with the development and use of performance indicators in part four of the model can relate to technical aspects of the indicator design, social responses by those whose performance is being measured, as well as political issues related to how such indicators would be used. Some comments are made on these aspects in Chapter 7 but the extent of such reactions to the implementation of a set of performance indicators like these can have a significant impact on the effectiveness of their use.

Unlike the learning outcome statements and the standards themselves the proposed performance indicators have not been canvassed at all in the university community to date and so remain untested both in terms of trialling with real data or for acceptability by teachers and administrators in the sector. Both of these would need to occur before any implementation of such a set could be attempted. The formulation of the speculative set of performance indicators assumes that individual student achievement data would be available from institutions. The cognitive indicators could be populated by achievement data using any one of the three assessment options canvassed in the dissertation provided that the assessment grades applied were criterion-based and used the set of standards for comparative marking. Hence the indicators developed are not dependent on a particular choice of assessment option even though the results of applying them may be more coherent under Option 3. The indicators could therefore contribute to practice in measurement of institutional performance for student learning achievement on their own, in a similar way to student progress rate but referencing the learning outcomes and standards statements.

In relation to the affective performance measures included in the set, more work should be done on identifying the best type of data collection for such indicators which relate more to students' interactions with others and engagement with their communities. The

QILT data currently used provides a start but some of the surveys used overseas could be explored further and the data augmented with other local information about these specific interactions to provide a broader view of student learning for these outcomes.

The inclusion of some broad surrogates in the complete set of indicators shown in Table 7.4 seems contradictory given comments made in Chapter 2 of the dissertation. However there is often a trade-off between technical and practical matters in implementation. The chosen indicators were added because they do round out outcome measures in terms of employment options for graduates even though they do not relate directly to the specific knowledge or skills learned. Several of the learning outcomes which were included in the 38 outcomes specified through part 1 of the model but not in the core set of 15, related specifically to employment skills and so employment outcomes are broadly relevant. The proxy indicators proposed are graduation rates, employment outcomes and the proportion of graduands going on to further studies. They augment the new information on actual achievement of student learning provided through the cognitive and affective learning performance indicators and are already available annually from publication of graduate careers survey data on the QILT website.

8.4.5 Implementation issues

Finally, there are some practical issues which would need to be addressed if such an approach to assessment were to be implemented. These include how the model could be sustained with consistency over a period of years, including how the signature assignments might need to change to support changes in curriculum and how they could be adjusted for teaching at different levels in courses of study. Addressing these matters may have significant cost implications for the sector. The educational issues of teachers teaching to the test and students being selective about what they choose to learn if the assignments are not varied from year to year would require attention. The level of administration and management required for such a system to work, which would be considerable, would be critical to its implementation.

8.5 Implications for theory and practice

The application of the model developed in this study outlines one possible way to define, assess and measure achievement of learning outcomes for Australian universities using local achievement assessment and to use those measurements in a set of

quantitative institutional performance indicators, which appear to have some validity based on previous experience with similar measures.

Whether such a measurement model could ever be fully implemented in the sector because of the doubts about generalisability of the approaches used depends on the importance placed on the study's objectives by the sector and government and whether there is the collective 'will' of the higher education institutions in Australia and the Government to address the issues canvassed above. Based on debate within the sector it appears that the problem of getting a better understanding of students' learning outcomes by the end of a bachelor degree from Australian universities remains a concern of government and the desire to define outcome performance indicators is even stronger now than it was four years ago when this work commenced. This is evidenced by the work done on QILT and government statements about the need to maintain high quality in teaching and excellent employment outcomes for graduates. Without some policy development and research supported by associated funding which signals a commitment to improving the assessment of learning outcomes from Australian higher education, the nation risks being left behind other countries which are pursuing various approaches to address some of the problems identified here (e.g.US, Canada, China, India and several European countries). Apart from the practical aspects of getting the model used in this study to work effectively in learning outcomes measurement there are some larger sector-wide implications for the nature of teaching, the policy framework for higher education and research and scholarship at both theoretical and practical levels. Some of these are discussed below.

8.5.1 Policy

In order for the approach outlined in this study to be seriously considered it needs to be framed within appropriate policy within government and the institutions themselves. When quality was first discussed in higher education in the 1990s and the first institutional performance indicators were formalised in Australia, these issues were placed in the context of major new government policies and reforms. Each university in Australia was required to formulate a set of graduate attributes, to be assessed against a suite of performance indicators and develop administrative systems following adoption of these new policies. The more recent Review of Higher Education conducted in 2008 also contained recommendations which were adopted by government about the quality of teaching and the student experience and the need to focus on learning outcomes

rather than input measures. The policy framework has not advanced much since that time. To restimulate institutions some of the initiatives undertaken under the auspices of the Office of Learning and Teaching could be extended or refocussed but this also requires renewal of the government agenda in relation to teaching and learning. The policy implications of this study include that there needs to be government endorsement of the ongoing importance of teaching and learning to the development of employable rounded graduates to the future economic and innovation agendas of the nation, as has been done in the United States and in some parts of Europe. A clearly stated policy position on this and how any performance indicators developed under a model like that suggested here might be used in the sector from a funding or quality assurance perspective would be a first step to serious consideration about the measurement issues. Action on the policy front needs to replace the general rhetoric of universities and politicians on this matter and large scale research based initiatives funded as in other countries to explore how learning outcomes assessment could be practically achieved.

8.5.2 Pedagogy

The pedagogical implications of this model were it to be implemented in the sector would be significant, as the proposals for shifting to an objective based outcomes approach through the specification of the learning outcomes, standards and signature assignments show. Closer linkages between the specification of learning outcomes and the way material is taught would be needed, and as has been experienced in some colleges in the United States where learning outcomes have been placed at the centre of the approach to teaching (such as Alveno College), the nature of teaching itself is transformed.

If adopted at a sector level, the changes outlined for assessment have potential to impact on the way course materials are developed and combined together in the curriculum, and are assessed. The assessment component has the potential to have the greatest impact on the way teaching is carried out. For the model to work would require greater consistency in approach than exists at present in Australian higher education. The issue of course based assessment rather than focussing on subjects for alignment with the stated learning outcomes would also cause courses to be taught in a holistic way in order to ensure course learning outcomes could be achieved by the students. This only emphasises the need for a strong policy framework within universities and sector-wide

relating to learning outcomes assessment which ensures that innovation is not stifled, and new approaches to teaching and assessment are introduced.

The route taken in Australian higher education to more commonly define learning outcomes at the subject level than for a course has made the measurement task more complex. This begs the question as to whether some of the problems about comparability of measurement would be avoided if, as recommended in the AAGLO project (see section 5.8.2), the focus were to be changed from subjects to looking at course based assessment instead, as has been done in many other countries.

This study has grappled with the need for the assessment of learning outcomes in higher education to address both the need for the sector to be able to address accountability to stakeholders as well as to serve the key function of assisting students to learn. The development of performance indicators addresses the former. There is perhaps a role in the future for a model like the one presented to address the dominant component of accountability and to keep local assessment tasks like tests and examinations formative and entirely for the purposes of student development.

8.5.3 Scholarship and research

Understanding the nature of the impact of discipline on learning outcomes and their assessment is central to the model if the more generalisable approach proposed to these constructs is to be accepted. In particular research about the appropriateness and feasibility of the signature assessment concept as presented here couched in discipline contexts is needed. There are several opportunities for further research following on from this study to establish whether the proposals are educationally sound or whether alternative approaches would be more satisfactory and acceptable to the sector. This research would require engagement of a community of scholars from a range of disciplines to be able to advance the scholarship of teaching and learning in relation to learning outcomes assessment and measurement.

8.5.4 Practice

Much of what has been discussed relates to the practical aspects of implementation of the model at an institutional and sector level for new proposals for assessment and measurement. The implications for practice in the sector include changes to the way assessment results are stored and used as well as the introduction of additional specifications for an annual data collection on student learning outcomes attainment.

These are large scale operational issues and have cost and practice implications for each university as well as central regulators.

8.6 Further steps required to advance the model

What would need to be done to make this model work as a national approach to obtaining valid inter-institutional comparisons of the achievement of learning outcomes? The way Australian higher education has chosen a more institutionally centred (rather than sector based) approach to the measurement of learning outcomes than a number of other countries has meant that projects funded in recent years by government agencies have commonly focussed on institutional proposals rather than addressing issues more generally as this study has attempted to do. In summary, to advance this model and address the key issues and limitations documented above there would be a need for work to be done on:

- Further testing of acceptability of the conceptual model for learning outcomes measurement using routine assessment outlined here;
- Further development of the concept of generalisable common signature assessments and their feasibility from an academic and pedagogical perspective;
- Projects which involve collaborative development of particular assessment approaches which address the generalisable learning outcomes developed in this study in a consistent way;
- Collection of real data from a small sample of Australian universities which have been identified through this study as pursuing the goal of consistent measurement of learning outcomes across different disciplines and institutions; and
- Continued pursuit of approaches for more top down measurement of learning outcomes at the course and institutional levels which make use of routine local achievement assessment conducted in universities.

8.7 Costs and benefits

The costs associated with the implementation of the model outlined in this dissertation for developing more useful measures of performance in relation to achievement of learning outcomes, would be significant both for the universities and the government. They would include the development of a new component of the national student data

collection on achievement, changes to individual universities' student or learning management systems, collaboration costs associated with the development of new assessment assignments on which to base the measurement, and overlaid quality assurance processes to ensure the validity and reliability of any comparisons made between the achievement levels of individual students and collectively.

In addition within the institutions themselves, in capturing base assessment data for the performance indicators, the application of the principles identified may require a reshaping of curriculum and modes of assessment which could be more labour intensive than at present.

These costs could only be accommodated in the sector if there was a strong commitment to changes required and clear benefits arising from the introduction of such measures.

Certainly there would be significant benefits to several groups of stakeholders from successful implementation of the model. It is considered that there is sufficient validity of the logic used and the approach that, with the cooperation and commitment of the higher education sector, the model could possibly be implemented successfully in Australia. If this could be achieved once remaining issues are addressed, then each of the key stakeholder groups to higher education would reap a number of benefits.

Students would:

- have clear information about what the expectations of their university are for their learning outcomes;
- have improved learning experiences;
- experience greater engagement in their study because of the explicit identification of both cognitive and affective learning outcomes as the outcomes they are expected to achieve; and
- receive statements of graduate attainment that would state in greater detail what they actually know and can do by the end of their studies rather than just lists of grades and subjects completed.

Teachers would:

- improve their teaching and assessment by having greater clarity about what needs to be taught to achieve the learning objectives of the courses they teach and how that curriculum needs to align with the nature of assessment used;
- gain a greater sense of teacher engagement because of the need to work with colleagues across disciplines as well as within their own; and
- assist students to become internationally competitive in their achievements.

Employers would:

- be able to compare the knowledge and skills of graduates from a range of institutions more easily; and
- identify the broad level of achievement attained by potential employees framed in terms of their knowledge and skill set.

Governments and funding agencies would:

- ensure institutional accountability for funding of teaching and learning outcomes;
- be able to examine relative institutional performance in teaching and learning; measured by achievement of learning outcomes through benchmarking based on achievement of learning outcomes;
- fund on outcomes rather than inputs or outputs; and
- incentivise universities to improve their quality of teaching and learning.

These benefits are substantial and seem worth the cost of having a go at implementation of such a model.

Bibliography

Access Economics Pty Ltd. (2005). *Review of Higher Education Outcome Performance Indicators*.

Adelfio, G, Boscaino, G, & Capursi, V. (2014). A new indicator for higher education student performance. *Higher Education, 68*, 653-668.

Association of American Colleges and Universities. (2010). VALUE:Valid Assessment of Learning in Undergraduate Education, from https://www.aacu.org/value/rubrics/index.cfm

Association of American Colleges and Universities. (2013a). Liberal Education and America's Promise (LEAP), from www.aacu.org/LEAP

Association of American Colleges and Universities. (2013b). Using the VALUE Rubrics for improvement of learning and authentic assessment. AACU website: Association of American Colleges and Universities.

Association of American Colleges and Universities. (2013c). VALUE - original project description Retrieved 24 August 2013, from http://www.aacu.org/value/project_description.cfm

Astin, A W, & Antonio, A L. (2012). *Assessment for Excellence - the philosophy and practice of assessment and evaluation in higher education* (2nd ed.). United Kingdom: Rowman and Littlefield Publishers, Inc.

Australian Council for Educational Research. (2001). *Graduate Skills Assessment*. Canberra, Australia: Department of Education, Training and Youth Affairs.

Australian Learning and Teaching Council. (2010a). *ALTC Assessment - completed and continuing projects and fellowships related to assessment*. Canberra, Australia: Department of Education, Employment and Workplace Relations.

Australian Learning and Teaching Council. (2010b). Learning and Teaching Academic Standards Project - Learning and Teaching Standards Statements (Vol. Accounting, Creative and performing arts, Engineering and ICT, Health, medicine and veterinary science, Geography, Law, Available from www.altc.edu.au/standards/disciplines/BME;

www.altc.edu.au/standards/disciplines/CPA;
www.altc.edu.au/standards/disciplines/EICT;
www.altc.edu.au/standards/disciplines/HMVS;
www.altc.edu.au/standards/disciplines/ASSH;
www.altc.edu.au/standards/disciplines/law

Australian Learning and Teaching Council. (2011a). *Learning and Teaching Academic Standards Project Final Report*.

Australian Learning and Teaching Council. (2011b). *Learning and Teaching Standards Project - Learning and Teaching Standards Statements*.

Australian Learning and Teaching Council. (2011c). *Resources to Assist Discipline Communities to Define Threshold Learning Outcomes (TLOs) Project*. Retrieved 1 September 2013, from http://www.olt.gov.au/resource-resources-assist-discipline-communities-define-tlos-2011.

Australian Qualifications Framework Council. (2013). *Australian Qualifications Framework*, from http://www.aqf.edu.au/wp-content/uploads/2013/05/AQF-2nd-Edition-January-2013.pdf.

Australian Universities Quality Agency. (2006). *Setting and monitoring academic standards for Australian higher Education - a discussion paper*. Canberra, Australia: AUQA.

Baik, C, & James, R. (2014). Seven strategic areas for advancing the assessment of learning in higher education. In Hamish Coates (Ed.), *Higher Education Learning Outcomes Assessment - International Perspectives* (Vol. 6, pp. 279-294). Frankfurt, Germany: Peter Lang.

Baker, G R, Jankowski, N A, Provesis, S, & Kinzie, J. (2012). *Using assessment results: Promising practices of institutions that do it well*. Illinois, United States: NILOA.

Baldwin, Peter. (1991). *Higher Education : Quality and Diversity in the 1990s*. Canberra, Australia: Australian Government Publishing Service.

Banta, T, Griffin, M., Flateby, T.L., & Kahn, S. (2009). Three promising alternatives for assessing college students' knowledge and skills. *Occasional paper 2*. Retrieved from http://www.learningoutcomesassessment.org/NILOAReports.htm

Barrie, Simon. (2007). A conceptual framework for the teaching and learning of generic graduate attributes. *Studies in higher education, 32*(4), 439-458.

Barrie, Simon. (2009). *Introduction to mapped universities' statements of graduate attributes*. Sydney, Australia: Australian Teaching and Learning Council.

Barrie, Simon, Hughes, Claire, Crisp, Geoffrey, & Bennison, Anne. (2014). *Assessing and assuring Australian graduate learning outcomes:principles and practices within and across disciplines*. Canberra, Australia: Office for Learning and Teaching.

Barrie, Simon, Hughes, Claire, Smith, Calvin, & Thomson, Kate. (2009). *The national Graduate Attributes Project - key issues to consider in the renewal of learning and teaching experiences to foster Graduate Attributes*, from www.itl.usyd.edu.au/projects/nationalgap/introduction.htm.

Blaich, C, & Wise, K. (2011). From gathering to using assessment results : Lessons from the Wabash National Study. *Occasional papers, Occasional paper 8*. Retrieved from www.learningoutcomesassessment.org

Bloom, B.S., Engelhart, M.D., Furst, E.J. Hill, W.H.and Krathwohl, D.R. (1956). *Taxonomy of educational objectives: Handbook 1 Cognitive Domain*. London, UK: Longmans Green.

Bloxham, Susan, & Boyd, Peter. (2007). *Developing Effective Assessment in Higher Education : A practical guide*. Maidenhead, United Kingdom: McGraw Hill Education.

Bloxham, Susan, den-Outer, Birgit, Hudson, Jane, & Price, Margaret. (2015). Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assessment and Evaluation in Higher Education*. Retrieved from http://dx.doi.org/10.1080/02602938.2015.1024607

Bloxham, Susan, Hughes, Claire, & Adie, Leonore. (2015). What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices. *Assessment and evaluation in higher education*, 16

Boud, D. (2010a). Assessment 2020 : Seven propositions for assessment reform in higher education. Retrieved from www.assessmentfutures.com

Boud, D. (2010b). *Student Assessment for Learning in and after Courses - Final report of senior fellowship*. Canberra, Australia: Office of Learning and Teaching.

Bourke, Paul. (1986). *Quality Measures in Universities*. Canberra, Australia: Australian Government Publishing Service.

Bradley, D, Scales, W, Noonan, P, & Nugent, H. (2008). *Review of Australian Higher Education*. Canberra, Australia: Commonwealth of Australia.

Carless, D, Joughin, G, & Liu, Ngar-Fun. (2006). *How assessment supports learning - Learning-oriented assessment in action*. Hong Kong: Hong Kong University Press.

Carter, R. (1985). A taxonomy of educational objectives for professional education. *Studies in higher education, 10*(no. 2).

Cave, Martin, Hanney, Stephen, & Kogan, Maurice. (1991). *The Use of Performance Indicators in Higher Education - a critical analysis of developing practice* (2nd ed.). London, United Kingdom: Jessica Kingsley Publishers Ltd.

Cerych, L, Furth, D, & Papadopoulos, G. (1974). *Overall issues in the Development of Future Structures of Post-secondary Education*. Paris: OECD.

Chalmers, Denise. (2008). *Review of Australian and National Performance Indicators and Measures of the Quality of Teaching and Learning in Higher Education*. Canberra, Australia: Australian Learning and Teaching Council.

Coates, Hamish. (2008). *What's the difference?: models for assessing quality and value added in higher education* (Conference Paper): Australian Universities Quality Agency.

Coates , Hamish. (2010). Defining and monitoring academic standards in Australian higher education. *Higher Education Management and Policy*, *22*(1), 1-17.

Coates, Hamish. (2011). The datafication of Higher Education:Evaluating tests that measure learning outcomes. *Joining the dots Research Briefing, 1*, 7. Retrieved from www.acer.edu.au/jtd

Coates, Hamish. (2014). *Higher education learning outcomes assessment - International perspectives* (Vol. 6). Frankfurt, Germany: Peter Lang.

Coates, Hamish, & Richardson, Sarah. (2011). An international assessment of bachelor degree graduates' learning outcomes. *Organisation for Economic Cooperation & Development*, *23*(3), 51-69.

Commonwealth of Australia. (2009). *Transforming Australia's Higher Education System*. Canberra, Australia: Australian Government Publishing Service

Council for National Academic Awards. (1990). *The measurement of value added in higher education*. London, United Kingdom: Council for National Academic Awards and Polytechnics and Colleges Funding Council.

Cunha, J.M., & Miller, T. (2014). Measuring value-added in higher education: Possibilities and limitations in the use of administrative data. *Economics of education review*, *42*, 64-77.

Dawkins, John. (1987). *Higher Education : A policy discussion paper*. Canberra, Australia.

Dawkins, John. (1988). *Higher Education : A policy statement*. Canberra, Australia.

Department of Education Employment and Workplace Relations. (2011). *Assessment of Generic Skills, Discussion Paper*. Canberra, Australia: Commonwealth of Australia

Department of Education Science and Training. (2007). *Technical notes 1 and 2 : Student Outcome Indicators of Australian Higher Education Institutions, 2002 and 2003*, from http://www.dest.gov.au/higher_education/policy_issues_reviews/key_issues/assuring_quality.

Department of Education Training and Youth Affairs. (2001). *Higher Education - annual report for the 2002-2004 triennium*. Canberra, Australia: Commonwealth of Australia.

Department of Employment Education and Training, & Organisation for Economic and Cultural Development. (1993). *The Transition from Elite to Mass Higher Education*. Canberra, Australia: Commonwealth of Australia.

Dochy, F, Segers, M, & Wijnen, H. (1990). Selecting Performance Indicators - a proposal as a result of research. In Goedegebuure L Maassen P and Westerheidjen D (ed) (Ed.), *Peer Review and Performance Indicators* (pp. 135-154). Utrecht: Lemma.

Eber, P A, & Parker, T S. (2008). Assessing student learning : Applying Bloom's taxonomy. *Human Service Education, Indiana University-Purdue University, Fort Wayne*.

Edwards, D, & Pearce, J. (2014). Outcomes assessment in practice. In Hamish Coates (Ed.), *Higher education learning outcomes assessment - International perspectives* (Vol. HERP 6, pp. 69-88). Frankfurt, Germany: Peter Lang.

Entwhistle, Noel J, & Ramsden, Paul. (1983). *Understanding Student Learning*. Kent, United Kingdom: Croom Helm Ltd.

Ewell, P.T. (2010). Conversations about Standards and Institutional Funding, *L H Martin Professional Development Session*. Melbourne, Australia.

Ewell, P.T., Gaston, P., & Schneider, C.G. (2011). *The Degree Qualifications Profile*. Indianapolis, United States: Lumina Foundation.

Ewell, P.T., Kuh, G, & Ikenberry, S. (2013). *The Degree Qualifications Profile - Implications for Assessment*. Indianapolis, United States: Lumina Foundation.

Ewell, P.T., Mandell, C., Martin, E., & Hutchings, P. (2013). Mapping the curriculum: learning outcomes and related assessments. Retrieved from www.learningoutcomeassessment.org

Freeman, Mark. (2009). *Embedding the development of intercultural competence in business education*. Canberra, Australia: Australian Learning and Teaching Council.

Freeman, Mark. (2011). *Learning and teaching academic standards resourcing for Accounting*. Canberra, Australia: Australian Learning and Teaching Council.

Goedegebuure, L , Maassen, P, & Westerheidjen, D (1990). *Peer Review and Performance Indicators - quality assessment in British and Dutch higher education*. Utrecht: Lemma.

Graduate Careers Australia, & The Social Research Centre. (2015). *2014 University Experience Survey National Report*. Canberra, Australia.

Griffin, P. , & Nix, P. (1991). *Educational Assessment and Reporting, a New Approach*. New South Wales, Australia: Harcourt Brace Jovanovich Ltd.

Group of Eight. (2011). *Go8 Update - Go8 pilots a new quality verification system*. Canberra, Australia: Group of Eight.

Henman, P., James, R, Lingard, R., Gable, A., Foster, M., & Power, M. (2015). *The rise of performance governance in Australian higher education : Comparing the influences of performance measurement in higher education, schooling and primary health care*. Paper presented at the Seminar, University of Melbourne.

Higher Education Council. (1992a). *Achieving Quality*. Canberra, Australia: Australian Government Publishing Service.

Higher Education Council. (1992b). *The Quality of Higher Education - Discussion Papers*. Canberra, Australia: National Board of Employment, Education and Training.

Higher Education Quality Council of Ontario. (2013a). Defining and measuring learning outcomes Retrieved 27 August 2013, from www.heqco.ca/en-CA/Research/LearningOutcomes/Pages/default.aspx

Higher Education Quality Council of Ontario. (2013b). Research Plan 2013-14 Retrieved 24 August 2013, from http://www.heqco.ca/SiteCollectionDocuments/HEQCO%20Research%20Plan%202013-14.pdf

Higher Education Quality Council of Ontario. (2013c). Tuning: Identifying and Measuring Sector-based Learning Outcomes in Post-secondary Education. In Personal communication from H Coates (Ed.). draft paper for comment.

Immerwahr, J. (2000). *Great expectations: How New Yorkers view higher education* (No. R-00-2f). New York, USA: National Center for Public Policy and Higher Education, CA.

International Tuning Academy. (2016). Measuring and Comparing Achievements of Learning Outcomes in Higher Education in Europe' (CALOHEE) Retrieved March 2016, from http://news.tuningacademy.org/calohee-project/

James, R, McInnis, C, & Devlin, M. (2002). *Assessing Learning in Australian Universities*. Melbourne, Australia: Centre for the Study of Higher Education, University of Melbourne and The Australian Universities Teaching Committee,.

Jones, A. (2006). *Re-disciplining generic skills: An examination of the relationship between the disciplinary context and generic skills in higher education.* University of Melbourne, Melbourne.

Keshavarz, Mohsen. (2011). Measuring Course Learning Outcomes. *Journal of Learning Design, 4*(4), 1-9.

Kleinberg, Jon. (2003). An impossibility theorem for clustering. *Department of Computer Science, Cornell University*.

Kuh, G, Jankowski, N, & Ikenberry, S. (2013). *From Provosts' lips to NILOA's ear:what we know about institutional assessment practice in 2013*. Paper presented at the Making Learning Outcomes Usable and Transparent. Retrieved 2 April, 2014, from www.learnigoutcomesassessment.org

Kuh, G, Jankowski, N, Ikenberry, S, & Kinzie, J. (2014). *Knowing what students know and can do - The current state of student learning outcome assessment in US colleges and universities*. United States: NILOA.

Lederman, Doug. (2015). New effort to standardize faculty-driven review of student work. *Inside Higher Education*, 2015, from https://www.insidehighered.com

Lennon, Mary Catharine, & Frank, Brian. (2014). Decentralised learning outcomes assessments: Developments in Ontario, Canada. In Hamish Coates (Ed.), *Higher education learning outcomes assessment* (pp. 89-110). Frankfurt, Germany: Peter Lang.

Lennon, Mary Catharine, Frank, Brian, Humphreys, James, Lenton, Rhonda, Madsen, Kirsten, Momri, Abdelwahab, et al. (2014). *Tuning:Idenifying and measuring sector-based learnign outcomes in post-secondary education*. Toronto, Canada: Higher Education Quality Council of Ontario (HEQCO).

Linke et al. (1991). *Performance Indicators in Higher Education - Report of a trial evaluation study commissioned by the Commonwealth Department of Employment, Education and Training*. Canberra, Australia.

Lokhoff, J., Wegewijs, B., Durkin, K., Wagenaar, R., Gonzales, J., Isaacs, K., et al. (2010). *A Tuning guide to formulating degree programme profiles*. Spain: Competences in Education and Recognition Project (CoRe).

Melton, Reginald. (1996). Learning Outcomes for Higher Education: Some Key Issues. *British Journal of Educational Studies, 44*(4), 409-425.

Nusche, Deborah, & Organisation for Economic Cooperation and Development. (2008). *Assessment of Learning Outcomes in Higher Education: A Comparative Review of Selected Practices. OECD Education Working Papers, No. 15*: OECD Publishing.

Oliver, Beverley. (2011a). *Assuring graduate outcomes*. Sydney, Australia: Australian Learning and Teaching Council.

Oliver, Beverley. (2011b). *Building course team capacity to enhance graduate employability*. Canberra, Australia: Office of Learning and Teaching.

Organisation for Economic Cooperation and Development. (1988). *Performance Indicators in Higher Education : A study of their development and use in 15 OECD countries*. Paris, France: IMHE, OECD.

Otter, Sue. (1992). *Learning outcomes in higher education*. United Kingdom: Department of Employment, UK Government.

Ou, Lydia Liu. (2009). Measuring Learning Outcomes in Higher Education. *ETS (Educational Testing Service) R&D connections, No 10*(June 2009), 6.

Ou, Lydia Liu. (2011a). Value-added assessment in higher education: a comparison of two methods. *Higher Education, 61*(4), 445-461.

Ou, Lydia Lui. (2011b). Measuring value-added in higher education: conditions and caveats - results from using the Measure of Academic Proficiency and Progress (MAPP™). [Article]. *Assessment & Evaluation in Higher Education, 36*(1), 81-94.

Rhodes, T., & Finley, A. (2013a). *Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment*. United States: Association of American Colleges and Universities.

Rhodes, T., & Finley, A. (2013b). Using the VALUE Rubrics for improvement of learning and authentic assessment. (Foreword). Retrieved from www.aacu.org/value/casestudies

Richardson, S., & Coates, Hamish. (2014). Essential foundations for establishing equivalence in cross-national higher education assessment. *Higher Education, 68*(6), 11.

Sadler, D R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education, Vol 13*(No 2), 191-209.

Sadler, D R. (2005). Interpretation of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education, Vol 30*(No 2), 175-194.

Sadler, D R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education, 34*, 159-179.

Sadler, D R. (2010). Fidelity as a precondition for integrity in grading academic achievement. *Assessment and Evaluation in Higher Education, 35*(6), 727-743.

Sadler, D R. (2011). Academic freedom, achievement standards and professional identity. *Quality in Higher Education, Vol 17*(No 1), 85-100.

Sadler, D R. (2014). The futility of attempting to codify academic achievement standards. *Higher Education*(67), 273-288.

Sanders, W.L. , & Horn, S.P. (1994). The Tennessee Value-Added Assessment System : Mixed - Model Methodology in Educational Assessment. *Journal of Personnel Evaluation in Education, 8*, 299-311.

Schneider, C.G, & Ewell, P.T. (2013). *The degree qualifications Profile (DQP) and implications for assessment*. Paper presented at the webinar conducted by NILOA.

Schuwirth, L., Pearce, J., Australian Medical Assessment Collaboration, & ACER. (2014). *Determining the Quality of Assessment Items in Collaborations : Aspects to Discuss to Reach Agreement* Canberra, Australia.

Shavelson, R. (2007). Assessing student learning responsibly: from history to audacious proposal. *Change, January/February 2007*, 27-33.

Shavelson, R. (2009). *Measuring college learning responsibly : accountability in a new era*. Palo Alto: Stanford University Press.

Shavelson, R., & Huang, L. (2003). Responding responsibly to the frenzy to assess learning in higher education. *Change, 35*(1), 10-19.

Spellings Commission. (2006). *A Test of Leadership:Charting the future of US Higher Education*, from http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf.

State Higher Education Executive Officers. (2015). MSC: A multi-state collaborative to advance learning outcomes assessment. Retrieved March 2016, from http://www.sheeo.org/projects/msc-multi-state-collaborative-advance-learning-outcomes-assessment#What

Stella, A, & Woodhouse, D. (2007). Benchmarking in Australian higher education: A thematic analysis of AUQA audit reports. In Australian Universities Quality Agency (Ed.). Melbourne, VIC.

Teichler, Ulrich. (1988). *Changing Patterns of the Higher Education System. The Experience of Three Decades*. London: Jessica Kingsley.

Tertiary Education Quality and Standards Agency. (2011). Higher Education Standards Framework (Threshold Standards). Canberra, Australia: Australian Government.

Trow, Martin. (1974). Problems in the transition from elite to mass higher education. *Policies for Higher Education, Paris OECD* 51-101.

Trow, Martin. (1976). Elite higher education : An endangered species? *Minerva, 1*.

Tuning process. (2010). Tuning Methodology Retrieved 23 August 2013, from http://www.unideusto.org/tuningeu/tuning-methodology.html

Ursin, J. (2014). Learning outcomes in Finnish higher education from the perspective of comprehensive curriculum framework. In Hamish Coates (Ed.), *Higher Education learning outcomes assessment - International perspectives* (pp. 159-174). Frankfurt, Germany: Peter Lang.

van Dooren, Wouter, Bouckaert, Geert, & Halligan, John. (2015). *Performance management in the public sector*. Electronic book: Hoboken: Taylor and Francis.

Volkwein, J Fredericks. (2011). *Gaining ground: The role of institutional research in assessing student outcomes and demonstrating institutional effectiveness*. United States: NILOA.

Wilkinson, D., Canny, B., Pearce, J., Coates, H., & Edwards, D. (2014). Collaborative assessment of medical student learning outcomes : the Australian Medical Assessment Collaboration. *Medical Journal of Australia, 199*(9), 578-580.

Wran, N. (1988). *Report of the Committee on Higher Education Funding*. Canberra, Australia: Commonwealth of Australia.

Zlatkin-Troitschanskaia, O., Kuhn, C., & Toepper, M. (2014). Modelling and assessing higher education learning outcomes in Germany. In Hamish Coates (Ed.), *Higher education learning outcomes assessment - International perspectives* (pp. 213-237). Frankfurt, Germany: Peter Lang

APPENDICES

APPENDIX 3.1-Comparison of approaches to specifying learning outcomes

Table A – Characteristics by approach

Characteristic of the system	Otter	HEQCO Tuning matrix	AACU (VALUE system of rubrics)	Lumina Degree Qualifications Profile(DQP)/NILOA	Australian Qualifications Framework
Number of tier 1 clusters of learning outcomes	4	6	13	5	3
Names of top level clusters	• Core • Subject based • General • Other	• Knowledge • Critical and creative thinking • Communications • Social responsibility • Personal & interpersonal • Practice & methods	• Civic engagement • Creative thinking • Critical thinking • Ethical reasoning • Information literacy • Inquiry and analysis • Integrative learning • Intercultural knowledge and competence • Foundations and skills for lifelong learning • Oral communication • Problem solving • Quantitative literacy • Reading • Teamwork • Written communication • Global learning	• Broad integrative knowledge • Specialised knowledge • Applied learning • Global learning	• Knowledge • Skills • Application of skills and knowledge

Characteristic of the system	Otter	HEQCO Tuning matrix	AACU (VALUE system of rubrics)	Lumina Degree Qualifications Profile(DQP)/NILOA	Australian Qualifications Framework
Number of second level learning outcomes	Variable depending on source For employers-17 For students- 19	41	88	No specific sub-categorisation Detailed descriptions which are more a set of threshold standards	0
Applicable across all disciplines/ discipline specific	Core, General & Other learning outcomes across all disciplines. Subject based outcomes by discipline	Yes based on discipline specific learning outcomes and merged	Yes – but matrix does not address discipline knowledge at all	Yes	Yes
Outcomes defined by level of study	No	Yes, four levels, 2 and 3 year diplomas, Bachelor, and masters levels)	Only for bachelor degrees	Yes	Yes including other education sectors
Associated standards	No	Yes- single standard for each learning outcome	Intercultural knowledge and competence	See above comment relating to number of learning outcomes	Yes -very general statements

Characteristic of the system	Otter	HEQCO Tuning matrix	AACU (VALUE system of rubrics)	Lumina Degree Qualifications Profile(DQP)/NILOA	Australian Qualifications Framework
Assessability	Not explored But a discussion of use of taxonomies for achievement such as Bloom, Carter, Kolb, Klemp	Not very focused on assessment of the outcome. Now the focus of further HEQCO work	Lifelong learning	Not directly – statements of what is expected to be demonstrated on general terms	Not assessable in broad form
In use at present	No	No-draft only completed in 2013	Yes	Yes	Yes but mainly as an accreditation tool
Strength	Bottom up approach involving the academic community and students	Inter-institutional application	Sector-wide learning outcomes Detailed rubrics which lead to design of assessment tasks Large take-up of the system in US colleges and universities.	Aimed at defining a qualifications framework which is applicable across. Links into VALUE	Good course accreditation tool
Weaknesses	No specific outcomes defined	No links to assessment or rubrics	Focuses on US liberal education courses – less applicable to professional education	No rubrics No focus on assessment in the profile itself	No specific learning outcomes or assessment tools defined

Table B– Second level learning outcomes for approaches with specification at this level

Otter	HEQCO Tuning matrix	AACU (VALUE system of rubrics)
• Communication	• Reading comprehension • Effective writing • Listening comprehension • Presentation skills • Oral communication skills • Graphical communication	• Oral communication organization • Language • Delivery • Supporting material • Central message • Reading comprehension • Reading genres • Relationship to text • Reading analysis • Reading interpretation • Readers' voice • Context and purpose of writing • Content development • Genre and disciplinary conventions • Sources and evidence • Control of syntax and mechanics
• Problem solving	• Problem identification • Problem solving	• Define problem • Identify strategies • Propose solutions/ hypotheses • Evaluate potential solutions • Implement solution • Evaluate outcomes
• Critical thinking	• Critical thinking	• Explanation of issues • Evidence • Influence of context and assumptions • Student's position • Conclusions and related outcomes • Creative thinking – acquiring competencies • Taking risks • Solving problems • Embracing contradictions • Innovative thinking • Connecting, synthesizing, transforming
• Analytical skills	• Analysis skills	• Topic of inquiry selection • Existing knowledge, research and views • Design process • Analysis • Conclusions • Limitations and implications
• Planning and organisation	• Evaluation	

Otter	HEQCO Tuning matrix	AACU (VALUE system of rubrics)
• Self discipline/self sufficient • Self-critical • Increased self-esteem	• Increased independence/self-motivated • Self-direction • Lifelong learning • Personal reflection	• Lifelong learning-curiosity • Initiative • Independence • Transfer • Reflection and self-awareness
• Enhanced career options/ Increased job satisfaction • Time management skills • Flexible	• Resource management Professional/legal responsibilities • Health & safety • Understanding risk and benefits	Integrative learning • Connections to experience • Connections to discipline • Transfer • Integrated communication
• Subject knowledge in field	• Breadth of knowledge • Limits of knowledge • Multi-disciplinary • Theory and concepts • Practice • Inter-disciplinary practice	
• Library skills	• Information management	• Information literacy – determination of information needs • Access information needs • Evaluate information and sources critically • Use information effectively • Access and use information ethically and legally
• Data collection skills/numeracy • Basic IT skills	• Numeracy	• Quantitative interpretation • Representation • Calculation • Application/analysis • Assumptions • Quantitative communication
	• Teamwork	• Contribution to team meetings • Facilitates the contributions of others in a team • Individual contributions outside of team meetings • Fosters constructive team climate • Responds to conflict
• Broader attitudes/ political & cultural sensitivity	• Diversity & respect • Environment & sustainability	• Knowledge – cultural self-awareness • Knowledge- cultural worldview frameworks • Empathy • Communication skills • Attitudes - Curiosity • Attitudes - Openness

Otter	HEQCO Tuning matrix	AACU (VALUE system of rubrics)
		• Global self-awareness • Global perspective taking • Cultural diversity • Personal and social responsibility • Understanding global systems • Applying knowledge in global contexts
• Social skills • Socially tolerant	• Diversity & respect • Environment & sustainability • Social awareness	• Civic engagement – Diversity of communities and cultures • Analysis of knowledge • Civic identity and commitment • Civic communication • Civic action and reflection • Civic contexts
	• Ethical principles and guidelines	• Ethical self-awareness • Understanding different ethical perspectives • Ethical issue recognition • Application of ethical perspectives • Evaluation of different ethical perspectives
	• Research • Research methods • Ethics of research • Formatting/ referencing • Creativity	

APPENDIX 4.1 – Tier 2 Graduate Attributes by institution

	Citizen-ship	Communic-ation	Critical analysis	Employmt skills	Equity & ethics	Flexibil-ity	Global context	Indept Intellect enquiry	Info and IT	Innovn	Knowledge of disc	Lifelong learning	Plan & Organisn	Profess skills	Research in disc	Scholar-ship	Self-manage-ment	Skills-disc related	Sustain-ability	Team-work	Total
ACU		1	2	0	3		0	1	2	0	1	0			1					1	12
Adelaide		1	0	0	1		1	0	2	1	1	1					1	1		1	11
ANU		1	1		0		1	1	0	0	1	0								1	6
Ballarat		1	1	1	1		1	0	0	0	1	1	1				1			1	10
Bond		1	1		1		1	0	0	0	2	0								1	7
Canberra		1	1		1		0	0	1	0	0	0	1					1		1	7
CDU		1	1		1		1	1	1	1	1	1	1				1		1	1	13
Charles Sturt		1	1	1	1	1	2	1	0	0	1	0							1	1	11
CQU		1	1		0		0	0	0	0	1	1								1	5
Curtin		1	1		1		1	0	2	0	1	1			1						9
Deakin		1	1		1		1	0	2	0	1	0			3			2	1	1	14
Edith Cowan		1	1		0		1	0	0	0	0	0				1				1	5
Flinders		1	0		1		1	1	0	0	1	0						1		1	7
Griffith		1	1		1		1	0	0	1	1	0								1	7
James Cook		3	4		4		1	2	6	0	1	2			1				1	1	26
La Trobe		0	0		0		0	0	0	0	1	1			1			1		1	5
Macquarie		1	2		1		0	1	0	1	1	0									7
Melbourne	2	1	3		5		0	0	1	0	1	1				1			1	1	17
Monash		1	0	3	2		1	0	0	1	0	0			1						9
Murdoch		1	1		2		1	1	1	0	1	1			1					1	11
Newcastle		1	1		1		1	0	0	1	1	1		2			1		1	1	12
Notre Dame	1	1	1		2		1	0	0	0	1	1				1				1	10
Queensland		1	1		1		0	0	0	1	1	0					1				6
QUT		1	1		2		1	1	0	0	1	1					1			1	10
RMIT		1	5	1	9	2	3	0	4	3	0	3	1				1		4	2	39
Southern Cross		1	0		1		1	1	1	1	2	1							1	1	11
Sunshine Coast		1	2		2		0	0	2	1	1	0	1				1		1	1	13
Swinburne		1	2		1	1	1	0	0	1	1	1		1						1	11
Sydney		1	0		1		1	1	1	0	0	1			1	1					8
Tasmania		1	1		1		1	0	0	0	1	0									5
UNE		1	1		1		1	0	1	0	1	1								1	8
UniSA		1	1		1		1	1	0	0	1	1								1	8
UNSW		1	2		2		0	2	2	2	3	1							1	1	17
USQ		2	1	1	1		1	0	1	1	1	1	1		1	1			1	1	15
UTS		1	1		2		1	1	1	2	0	1		2	1		1		1	1	15
UWS		0	0		0		0	0	0	0	0	0									0
UWA		1	1		1		0	0	0	0	1	0			1		1	1		1	8
VU		0	1		1		0	1	1	0	1	1						1		1	8
Wollongong		1	1		1		0	1	0	0	1	0									5
Total attributes	3	39	46	7	58	4	29	18	32	18	36	25	6	5	13	5	10	8	14	32	408

APPENDIX 4.2 – ALTC learning outcomes and threshold standards by discipline

Discipline/ learning outcomes	Knowledge	Ethics and professional responsibility	Thinking skills	Research skills	Communication, collaboration and leadership	Self-management	Application of skills
Law (bachelor)	Understand the fundamental areas of legal knowledge, the Australian legal system and underlying principles and concepts, including international and comparative contexts; Broader contexts in which legal issues arise; principles and values of justice and ethical practice in the law.	Demonstrate: Understanding of approaches to ethical decision-making; an ability to recognise and reflect on and develop an ability to respond to ethical issues likely to arise in professional contexts; ability to recognise and reflect upon professional responsibilities of lawyers in promoting justice and service to the community; Developing ability to exercise professional judgement.	Identify and articulate legal issues; Able to apply legal reasoning and research to generate responses to legal issues; Engage in critical analysis and reasoned choice amongst alternatives; Think creatively in approaching legal issues and generating responses	Demonstrate intellectual and practical skills needed to identify, evaluate and synthesise relevant factual, legal and policy issues	Communicate in ways that are effective, appropriate and persuasive for legal and non-legal audiences; Collaborate effectively	Learn and work independently; Reflect on and assess their own capabilities and performance and make use of feedback to support personal and professional development	
Education (Masters)	Possess an in-depth understanding of the field of education and an advanced understanding of at least one specialist area; Engage with current topics and issues in an area of specialisation; Demonstrate knowledge of current research findings and methods in their area of specialisation; Understand Indigenous education including the history and cultural development of Indigenous Australians	Have an advanced understanding of the values and ethical principles that underpin the broad field of education; Conduct of practice in an ethical, collaborative, professional and accountable manner while valuing social and environmental sustainability; Engage with broader community issues of ethical conduct, equity and social justice; Reflect upon sensitive and complex issues;	Achieve advanced cognitive, creative, analytical and practical skills essential to their area of practice in education; Interpret, apply and /or conduct research		Have advanced speaking, reading, writing, listening, interpretive and advocacy skills to equip them for leadership in their field; Demonstrate advanced knowledge and understanding of the discourses, debates and issues in education; Communicate complex educational issues to diverse social and cultural groups; Work collaboratively with others in education or related fields; Act constructively in collaboration with others	Demonstrate a high level of critical self-knowledge as a foundation for personal autonomy; Have the capacity for independent, autonomous, self-directed learning; Pursue further learning opportunities for continuing professional development	Synthesise theoretical understanding and practical skills to effectively plan, analyse, present and implement complex activities in their specialisation; Lead innovation and change in their area of study; Apply current research findings and methods to an area of educational practice
Architecture (Masters)	Identify, explain and work with knowledge of architecture, its history and precedents and with knowledge of people, environments, culture , history and ideas pertinent to architectural propositions	Demonstrate their understanding of architecture's status as an ethical service-oriented profession committed to responsible care for the inhabited environment; Engage proactively in the effective procurement of architectural propositions	Propositional, imaginative iterative, integrated thinking to synthesise complex architectural designs; Support decision-making using evidence-based reasoned argument and judgement pertaining to architectural propositions (described as Design in Standards document)	Research and evaluate emergent knowledge to fulfil the profession's role in society	Communicate with a variety of audiences in appropriate ways		
Creative Arts (bachelor)	Demonstrate skills and knowledge of the practices , languages, forms, materials, technologies and techniques in the Creative and Performing Arts discipline	Recognise and reflect on social, cultural and ethical issues, and apply local and international perspectives to practice in the discipline		Develop, research and evaluate ideas, concepts and processes through creative, critical and reflective thinking and practice	Interpret, communicate and present ideas, problems and arguments in modes suited to a range of audiences; Work independently and collaboratively in the discipline in response to project demands		Apply relevant skills and knowledge to produce and realize works, artefacts and forms of creative expression
Engineering (no level specified)		Identify, interpret and analyse stakeholder needs, establish priorities and the goals, constraints and uncertainties of the system(social, cultural, legislative environmental, business etc) using systems thinking while recognising ethical implications of professional practice	Apply problem solving, design and decision-making methodologies to develop components, systems and/or processes to meet specified requirements, including innovative approaches to synthesise alternative solutions, concepts and procedures, while demonstrating information, skills and research methods		Communicate and coordinate proficiently by listening, speaking, reading and writing English for professional practice, working as an effective member or leader of diverse teams, using basic tools and practices of formal project management	Manage own time and processes effectively by prioritising competing demands to achieve personal and team goals, with regular review of personal performance as a means of managing continual professional development	Apply abstraction, mathematics and discipline fundamentals to analysis, design and operation, using appropriate computer software, laboratory equipment and other devices, ensuring model applicability, accuracy and limitations

Discipline/ learning outcomes	Knowledge	Ethics and professional responsibility	Thinking skills	Research skills	Communication, collaboration and leadership	Self-management	Application of skills
Arts, Humanities and Social Sciences - Geography (bachelor)	Demonstrate a coherent understanding of trends, processes and impacts that shape Australian environments and/or societies; Demonstrate an understanding of the academic discipline, including awareness of its concepts, history and principal subfields, whilst acknowledging the contested, provisional and situated nature of geographical understanding		Apply geographical thought creatively, critically and appropriately to specific spaces, places and/or environments; Recognise, evaluate and synthesise various views, arguments and sources of knowledge pertinent to solving environmental and social problems	Resolve geographical questions by ethical means, applying evidence-based knowledge and appropriate research techniques, including those associated with field-work	Communicate geographical perspectives and knowledge effectively to specialist and non-specialist audiences using appropriately selected written, oral and visual means. Contribute effectively as a member or leader of diverse teams working in geographical or multi-disciplinary contexts.	Reflect on and direct their intellectual and professional development as geographers	
Health, Medicine and Veterinary Science (professional entry level)		Demonstrate professional behaviours; Promote and optimise the health of individuals and/or populations and welfare		Retrieve, critically evaluate, and apply evidence in the performance of health-related activities	Deliver safe and effective collaborative healthcare	Reflect on current skills, knowledge and attitudes, and pan ongoing personal and professional development	Assess individual and/or population health status and, where necessary, formulate, implement and monitor management plans in consultation with patients/clients/carers and communities
Science (bachelor)	Demonstrate well-developed knowledge in at least one disciplinary area and one other disciplinary area	Demonstrate knowledge of the regulatory frameworks relevant to their disciplinary area and personally practise ethical conduct	Gather, synthesise and critically evaluate information from a range of sources; Collect, accurately record, interpret and draw conclusions from scientific data	Design and plan an investigation; select and apply practical and/or theoretical techniques or tools in order to conduct an investigation	Communicate scientific results, information, or arguments, to a range of audiences, for a range of purposes, and using a variety of modes. Work effectively, responsibly and safely as an individual or team context	Be an independent and self-directed learner	Demonstrate a coherent understanding of science by: articulating the methods of science and explaining why current scientific knowledge is both contestable and testable by further enquiry; Explaining the role and relevance of science in society
Accounting (bachelor)	Integrate theoretical and technical accounting knowledge which includes a section of auditing and assurance, finance, economics, quantitative methods, information systems, commercial law, corporation law and taxation law	Exercise judgement under supervision to solve routine accounting problems in straigtforward contexts using social, ethical, economic, regulatory and global perspectives			Justify and communicate accounting advice and ideas in straightforward collaborative contexts involving both accountants and non-accountants	Reflect on performance feedback to identify and action learning opportunities and self-improvements	Critically apply theoretical and technical accounting knowledge and skills to solve routine accounting problems
Building and Construction (bachelor)	Integrate and evaluate the fundamental principles and technical knowledge of building and construction technology, management, economics and law	Demonstrate an integrated understanding of both the theory and practice of building and construction based on experience	Identify and resolve typical building challenges with limited guidance, employing appropriate evidence based problem-solving and decision-making methodologies	Research and develop methods and strategies for the procurement and delivery of contemporary construction work.		Critically and creatively reflect on personal behaviours and capabilities in the context of entry to professional practice	Interpret and negotiate building and construction information, instructions and ideas with various project stakeholders

APPENDIX 6.1 - Data collection proforma

PhD study - Using assessment of student learning outcomes to measure university performance

Proforma

1. Please enter the name of your University

2. If you do not wish to participate in this collection of data please tick the following box and return the proforma to martl@student.unimelb.edu.au.

Nil Response	

3. Does your university use common learning outcomes for all undergraduate bachelor degrees? Please tick box and indicate how many learning outcomes are used

Yes		Number	
No			

3a. If yes, for what purpose are the common outcomes used? Please tick relevant boxes

Internal accreditation	
External accreditation	
Curriculum modification	
Course review	
Quality assurance of teaching	
Institutional benchmarking	
Academic policy development	
Resource allocation	
Other (please specify)	

3b. And how many of these common learning outcomes are regularly assessed for student attainment? Please enter number below

Number of outcomes assessed at the course level	

The following 38 common learning outcomes have been derived earlier in this research study following an analysis of graduate attributes for all Australian universities and of work done through the Office of Teaching and Learning. They are clustered into 12 categories associated with a set of traits expected of a graduate from an undergraduate degree on completion of study. Questions 4 and 6 below list these outcomes and seek information about use and assessment approaches for them in your university.

4. Please indicate by ticking the relevant boxes which of the following common course learning outcomes, or something very similar, are used or are relevant to your university and the extent of their use.

Undergraduate learning outcome	Used across the whole university	Used only in some disciplines	Not used but relevant	Not relevant
Knowledge of discipline				
Understand, describe and apply theories and information relating to one or more disciplines				
Develop and demonstrate skills related to the discipline and apply them to professional practice				
Research and Scholarship				
Understand the theoretical basis of the discipline and apply its principles to professional practice				
Research, develop, and evaluate emergent knowledge in the discipline				
Produce innovative solutions to problems				
Be capable of initiating and embracing change				
Collaboration				
Work and learn collaboratively				
Work in a team to achieve joint goals and contribute effectively to the team's outcomes				
Lead, manage and contribute effectively to a team or project				
Communication				
Demonstrate effective oral communication in English				
Write clearly, coherently and creatively appropriate to audience needs				
Generate, calculate, interpret and communicate numerical information in ways appropriate to a discipline				
Equity and social justice				
Demonstrate respect for dignity of others and for human diversity				
Recognise and respect the role of cultural difference and diversity in work and social contexts				
Understand Indigenous Australian issues and cultures				

Undergraduate learning outcome	Used across the whole university	Used only in some disciplines	Not used but relevant	Not relevant
Ethics				
Understand and demonstrate professional ethical responsibilities				
Recognise ethical issues and apply ethical principles in complex situations				
Self-awareness and self-discipline				
Learn independently in a self-directed manner				
Demonstrate initiative in setting goals and completing learning tasks				
Reflect on and direct own intellectual and professional development				
Exercise initiative and responsibility				
Manage own time and meet deadlines for learning tasks				
Exhibit openness, intellectual humility, spirit of enquiry				
Able to reflect on and evaluate learning and performance In tasks				
Use feedback on performance and learning for improvement				
Thinking and analysis				
Demonstrate ability to think critically, to analyse and evaluate claims, evidence and arguments and to reason and deploy evidence clearly and logically				
Able to apply problem solving processes in novel situations				
Skills and their application in employment				
Locate, organize and evaluate information with emphasis on primary sources				
Utilise information and communication and other relevant technologies to solve problems and in learning				
Apply sound planning and organisational skills in learning tasks				
Demonstrate knowledge of regulatory frameworks and political influences for professional area of study				

Undergraduate learning outcome	Used across the whole university	Used only in some disciplines	Not used but relevant	Not relevant
Demonstrate capability to solve real world problems by applying learning for discipline(s)				
Adapt learning approach to suit different tasks				
Adaptable and able to manage change				
Civic and social understanding				
Be a well informed citizen able to contribute to their communities wherever they work and live				
Global understanding				
Able to develop and apply international perspectives in their discipline				
Demonstrate competence in culturally diverse and international environments				
Social and environmental sustainability				
Understand financial, social and environmental sustainability				

5. If you use common learning outcomes, in what ways does your university ensure that its learning outcomes are aligned with curriculum and assessment in a course of study? Please tick all relevant boxes. If not, please go to Question 6.

Curriculum mapping of subject content to course learning outcomes	
Assessment mapping for subjects to course learning outcomes	
Whole of course assessments such as capstone subjects, or portfolio submissions Please specify:	
Rely on Academic Quality assurance and accreditation scrutiny	
Other, please specify	

6. **For those learning outcomes that you identified in Question 2 as <u>in use</u> in your university, please list the three main key assessment approaches used to evaluate attainment of each of the outcomes?**

Undergraduate learning outcome	Assessment type		Brief description of task (eg essay, presentation, project, portfolio in particular subject, team task, capstone subject, standardised external test, student surveys, rubrics)
	S-Summative F-Formative B-both	C - Criterion based N - Normative	
Knowledge of discipline			
Understand, describe and apply theories and information relating to one or more disciplines			
Develop and demonstrate skills related to the discipline and apply them to professional practice			
Research and Scholarship			
Understand the theoretical basis of the discipline and apply its principles to professional practice			
Research, develop, and evaluate emergent knowledge in the discipline			
Produce innovative solutions to problems			
Be capable of initiating and embracing change			
Collaboration			
Work and learn collaboratively			
Work in a team to achieve joint goals and contribute effectively to the team's outcomes			
Lead, manage and contribute effectively to a team or project			

Undergraduate learning outcome	Assessment type		Brief description of task (eg essay, presentation, project, portfolio in particular subject, team task, capstone subject, standardised external test, student surveys, rubrics etc)
	S-Summ-ative F-Formative B-both	C - Criterion based N - Normative	
Communication			
Demonstrate effective oral communication in English			
Write clearly, coherently and creatively appropriate to audience needs			
Generate, calculate, interpret and communicate numerical information in ways appropriate to a discipline			
Equity and social justice			
Demonstrate respect for dignity of others and for human diversity			
Recognise and respect the role of cultural difference and diversity in work and social contexts			
Understand Indigenous Australian issues and cultures			
Ethics			
Understand and demonstrate professional ethical responsibilities			
Recognise ethical issues and apply ethical principles in complex situations			
Self-awareness and self-discipline			
Learn independently in a self-directed manner			
Demonstrate initiative in setting goals and completing learning tasks			

Undergraduate learning outcome	Assessment type		Brief description of task (eg essay, presentation, project, portfolio in particular subject, team task, capstone subject, standardised external test, student surveys, rubrics etc)
	S-Summ-ative F-Formative B-both	C - Criterion based N - Normative	
Reflect on and direct own intellectual and professional development			
Exercise initiative and responsibility			
Manage own time and meet deadlines for learning tasks			
Exhibit openness, intellectual humility, spirit of enquiry			
Able to reflect on and evaluate learning and performance In tasks			
Use feedback on performance and learning for improvement			
Thinking and analysis			
Demonstrate ability to think critically, to analyse and evaluate claims, evidence and arguments and to reason and deploy evidence clearly and logically			
Able to apply problem solving processes in novel situations			
Skills and their application in employment			
Locate, organize and evaluate information with emphasis on primary sources			
Utilise information and communication and other relevant technologies to solve problems and in learning			

Undergraduate learning outcome	Assessment type		Brief description of task (eg essay, presentation, project, portfolio in particular subject, team task, capstone subject, standardised external test, student surveys, rubrics etc)
	S-Summ-ative F-Formative B-both	C - Criterion based N - Normative	
Apply sound planning and organisational skills in learning tasks			
Demonstrate knowledge of regulatory frameworks and political influences for professional area of study			
Demonstrate capability to solve real world problems by applying learning for discipline(s)			
Adapt learning approach to suit different tasks			
Adaptable and able to manage change			
Civic and social understanding			
Be a well-informed citizen able to contribute to their communities wherever they work and live			
Global understanding			
Able to develop and apply international perspectives in their discipline			
Demonstrate competence in culturally diverse and international environments			
Social and environmental sustainability			
Understand financial, social and environmental sustainability			

Information sought in questions 7-9 is needed in order to understand the feasibility of using data from some of these assessment practices from your university in a later trial of a set of performance indicators.

7. **How and where is information stored on these individual assessment results in your university for these learning outcomes?**

Location	**Format (eg course level, subject level, part of a subject, etc)**
Central Student Management System	
Central Learning Management System	
Local data base records not held centrally	
Local paper based information not held centrally.	
Other (please specify)	

8. **How transferrable are the assessment approaches? Please tick all which apply**

Between courses and disciplines in your university?	
To other universities?	
Not at all transferrable	

9. **Does your university engage with other universities in peer review of assessment tasks and standards of attainment?**

Yes	
No	

If yes, please indicate name of other institution(s) and what practices are used.

10. **Is your university interested in participating in a trial of the performance indicators developed through this study?**

Yes	
No	

11. **Would you be willing to provide non-identified data on student results for some of these assessment tasks in electronic format for trialling the indicators?**

Yes	
No	

12. **Name and details of contact person for follow-up**

Name :
Telephone:
Email:

Thank you for providing this information
Please return completed form to martl@student.unimelb.edu.au

APPENDIX 6.2– Assessment characteristics and possible types of assessment tasks by learning outcome

Attribute cluster (Tier 2)	Learning outcome (Tier 3)	Level 1 standard (Developing)	Level 3 standard (Exemplary)	Cognitive or affective outcome/type of assessment	Is it routinely assessable?	Assessment task
Trait 1: Knowledge of discipline						
Discipline knowledge	Understand, describe and apply theories and information relating to one or more disciplines	Demonstrates a developing understanding of how theories work and information can be used to solve a problem in a discipline	Demonstrates comprehensive understanding of theories and their application and provides a well-argued solution to a problem in the discipline	Cognitive/ Psychological	Yes	• Short answer quiz/tests • Multiple-choice questions • Essay • Examination • Solve a technical problem in the discipline
Discipline –related skills	Understand and apply skills related to the discipline to professional practice	Demonstrates developing capability in skills and their application in an area of professional practice	Demonstrates excellence skills and their application to professional practice in the discipline	Cognitive/ Psychological	Yes	• Short answer quiz/test • Multiple-choice questions • Practical exercise relating to skills • Practicum • Simulation • Practical examination

Attribute cluster (Tier 2)	Learning outcome (Tier 3)	Level 1 standard (Developing)	Level 3 standard (Exemplary)	Cognitive or affective outcome/type of assessment	Is it routinely assessable?	Assessment task
Trait 2: Research and Scholarship						
Scholarship	Understand the theoretical basis of the discipline and apply its principles to professional practice	Demonstrates a developing understanding of the theoretical basis of the discipline and has applied it successfully to a problem from professional practice	Demonstrates excellent understanding of the theoretical basis of the discipline and has applied it in a transformative way to a problem from professional practice	Cognitive/ Psychological	Yes, but similar to Knowledge and Skills in the discipline	• Poster presentation • Essay • Presentation and paper
Research in discipline	Understand, create and evaluate emergent knowledge in the discipline	Demonstrates developing knowledge of new ideas and understanding of their impact on the discipline	Demonstrates excellence in knowledge of new ideas and understanding of their impact on the discipline	Cognitive/ Psychological	Yes	• Research project/paper on new discipline area • Thesis • Abstract for a paper • Presentation of paper
Innovation	Create innovative solutions to problems	Demonstrates developing capacity to look at problems in a new way and devise solutions based on this approach	Demonstrates excellence in developing innovative solutions to problems.	Cognitive/ Psychological	Yes, but not very differentiated from research	• Poster presentation • Essay • Oral presentation • Report on solving a new problem

Attribute cluster (Tier 2)	Learning outcome (Tier 3)	Level 1 standard (Developing)	Level 3 standard (Exemplary)	Cognitive or affective outcome/type of assessment	Is it routinely assessable?	Assessment task
	Understand the need for change and apply knowledge and skills to bring it about	Demonstrates a developing understanding of new ways of addressing problems in the discipline	Demonstrates excellence in addressing complex problems using knowledge and skills to solve them in the discipline	Cognitive/ Psychological	Difficult to assess	• Simulation of a real world situation • Report on solution to problem • Critique
Trait 3: Collaboration						
Teamwork	Learn collaboratively	Demonstrates a developing capacity to work in a team undertaking a task	Demonstrates an excellent capacity to contribute to the work in a team undertaking a joint task	Affective/ Behavioural	May not be. Possibly use other measures such as engagement surveys	• Team project • Portfolio of own learnings from joint project • Report/reflection on what has been learned from/with others

Attribute cluster (Tier 2)	Learning outcome (Tier 3)	Level 1 standard (Developing)	Level 3 standard (Exemplary)	Cognitive or affective outcome/type of assessment	Is it routinely assessable?	Assessment task
Teamwork	Participate as a member of a team to achieve joint goals and contribute to the team's outcomes	Demonstrates developing capability to contribute actively to work and discussions of a team to achieve goals	Demonstrates excellent and active contributions to work and discussions of a team and shows leadership of the team in achieving goals	Affective/ Behavioural	May not be. Possibly use other measures such as engagement surveys	• Team project • Report/reflection on own and others' roles (self-assessment) • Oral or written reflection on own contribution
Leadership	Lead, plan and play the main role in a team project	Demonstrates a developing capacity to lead and plan a project and to play a main role	Demonstrates excellent planning and leadership of a team engaging members and facilitating their contributions to maximise outcomes for the team project	Affective/ Behavioural	May not be. Possibly use other measures such as engagement surveys	• Plan for a team project including allocation of roles • Reflection on own and others' leadership in a project • Peer review by others
Trait 4 : Communication						
Communicat-ion (oral, written, graphical)	Communicate orally in English	Demonstrates a developing capacity to convey ideas orally to different groups in a presentation	Demonstrates excellence in convey ideas orally to different groups in a presentation clearly explaining concepts and ideas	Affective/ Behavioural	Yes	• Presentation to two differently constituted groups • Oral examination • Workshop • Seminar presentation

Attribute cluster (Tier 2)	Learning outcome (Tier 3)	Level 1 standard (Developing)	Level 3 standard (Exemplary)	Cognitive or affective outcome/type of assessment	Is it routinely assessable?	Assessment task
Communication (oral, written, graphical)	Generate, analyse and communicate numerical information in ways appropriate to a discipline	Demonstrates a developing capacity to interpret numerical information adequately for the discipline	Demonstrates excellence in calculation and interpretation of numerical information to enhance arguments by its use in the discipline	Cognitive/ Psychological	Yes	• Quantitative test • Problem requiring interpretation of data • Problem in discipline which requires quantitative understanding • Problem which requires information to be extracted from data
Trait 5 :Equity/social justice						
Equity/social justice	Demonstrate respect for dignity of others and for human diversity	Demonstrates developing respect for dignity of others and for human diversity in addressing issues in the discipline	Demonstrates complex understanding of human diversity and respect for the dignity of others in addressing issues in the discipline	Affective/ Psychological	Yes	• Essay • Real-world problem requiring acknowledgement of human diversity • Ethical problem
Equity/social justice	Recognise and respect the role of cultural difference and diversity in work and social contexts	Demonstrates developing understanding of cultural difference and the importance of it to work in the discipline	Demonstrates excellent understanding of cultural difference and is able to adapt work response to take account of this diversity	Affective/ Psychological	Yes	• Simulation • Essay • Real world problem requiring understanding of cultural difference

Attribute cluster (Tier 2)	Learning outcome (Tier 3)	Level 1 standard (Developing)	Level 3 standard (Exemplary)	Cognitive or affective outcome/type of assessment	Is it routinely assessable?	Assessment task
Indigenous cultural understand-ing	Understand Indigenous Australian issues and cultures	Demonstrates developing understanding of Indigenous issues and cultures and their importance to professional practice	Demonstrates excellent understanding of Indigenous issues and cultures and their importance to professional practice	Affective/ Psychological	Yes	• Essay • Real-world problem requiring Indigenous cultural competence • Ethical problem • Real world problem relating to an Indigenous issue
Trait 6 : Ethics						
Professiona l ethics related to discipline	Understand and demonstrate professional ethical responsibilities	Demonstrates developing understanding of ethical values and responsibilities in professional practice	Demonstrates excellent understanding of ethical responsibilities in professional practice	Affective/ Psychological	Yes	• Simulation of an ethical problem in discipline • Essay • Examination • Practicum
Personal ethics	Understand ethical issues and apply ethical principles in complex situations	Demonstrates developing capacity to understand ethical issues and apply ethical principles appropriately to practice in complex situations	Demonstrates excellent understanding of ethical issues and capacity to identify and apply ethical principles in complex situations	Affective/ Psychological	Yes	• Problem in personal ethics • Examination • Report • Simulation of personal ethical situation

Attribute cluster (Tier 2)	Learning outcome (Tier 3)	Level 1 standard (Developing)	Level 3 standard (Exemplary)	Cognitive or affective outcome/type of assessment	Is it routinely assessable?	Assessment task
Trait 7 : Self-awareness and self-discipline						
Independent intellectual enquiry	Learn independently in a self-directed manner	Demonstrates developing capacity to work independently and to plan own learning approach to the discipline	Demonstrates excellence in working independently and demonstrates sophisticated planning of own learning approach	Cognitive/ Psychological	Not in a comparable way	• Portfolio of achievements • Project • Plan for a project • Description of approach to be taken to a complex piece of work
Independent intellectual enquiry ctd	Apply initiative in setting goals and completing learning tasks	Demonstrates developing capacity to apply initiative in goal setting and satisfactorily completion of set learning tasks	Demonstrates excellent initiative in goal setting and satisfactorily completion of set learning tasks	Cognitive/ Psychological Also behavioural	Partly but also use other data	• Track record of timely completion of complex tasks Plan and completion of a real world task
Lifelong learning	Analyse own intellectual capabilities and create opportunities for own professional development	Demonstrates developing capacity to analyse own intellectual capabilities and understand the need for ongoing learning	Demonstrates excellence in analysing own intellectual capabilities and understanding the need for ongoing learning	Affective/ Psychological	Yes, except for measuring ongoing commitment	• Portfolio achievements • Tutorial • Presentation and responses to questions Self-assessment on a topic

Attribute cluster (Tier 2)	**Learning outcome (Tier 3)**	**Level 1 standard (Developing)**	**Level 3 standard (Exemplary)**	**Cognitive or affective outcome/type of assessment**	**Is it routinely assessable?**	**Assessment task**
Self-manage-ment	Demonstrate initiative and responsibility	Demonstrates developing initiative in learning tasks and takes responsibility for own learning	Demonstrates excellent levels of initiative and self-motivation in learning tasks and responsibility for own learning	Affective/ Psychological	Yes in part, but not long term.	• Research project • Real world problem which requires investigation and development of an approach
	Manage own time and meet deadlines for learning tasks	Meets all deadlines for learning tasks and developing management of time and priorities between several tasks	Meets all deadlines for tasks and demonstrates excellent use of own time to achieve goals	Affective/ Psychological	Yes, but difficult to measure other than as achieved/ not achieved	• Records of time for submission of multiple tasks • Map of quality vs timeliness of submission
	Demonstrate openness, intellectual humility, spirit of enquiry	Demonstrates developing intellectual humility, spirit of enquiry	Demonstrates excellent self-awareness and openness and explores solutions to problems in depth, showing detailed interest in the subject	Affective/ Behavioural	Yes, by behaviour	• Demonstration • Tutorial • Oral examination • Presentation on new issue

Attribute cluster (Tier 2)	Learning outcome (Tier 3)	Level 1 standard (Developing)	Level 3 standard (Exemplary)	Cognitive or affective outcome/type of assessment	Is it routinely assessable?	Assessment task
Self-manage-ment	Apply feedback on own performance and learning for own improvement	Demonstrates a developing capacity to respond well to feedback and to use it in improving responses in future tasks	Demonstrates excellence in responding to feedback, using it in improving response and applying it more widely in future tasks	Affective/ Behavioural	Not easy to test generally in the absolute	• Retesting and a concept on which feedback has been given (formative assessment task)
Self-awareness	Evaluate learning and performance in tasks	Demonstrates developing capacity to reflect on the significance of a learning task and accurately assess own performance	Demonstrates excellence in reflecting on significance of a learning task, accurately assess own performance and understand own strengths and weaknesses	Affective/ Psychological	Yes, but not in a comparable way	• Reflection on challenging learning task • Self-assessment • Report on experience
Trait 8 : Thinking and analysis						
Critical analysis	Demonstrate ability to think critically, to analyse and evaluate claims, evidence and arguments clearly and logically	Demonstrates developing capability to think critically, evaluate and analyse data and present logical and clear arguments	Demonstrates excellence in thinking critically, evaluating and analysing data and evidence and presenting logical and clear arguments	Cognitive/ Psychological	Yes	• New or real world problem which requires analysis and deduction rather than specific knowledge • Reviewing publication and critiquing findings • Develop argument for a position

Attribute cluster (Tier 2)	Learning outcome (Tier 3)	Level 1 standard (Developing)	Level 3 standard (Exemplary)	Cognitive or affective outcome/type of assessment	Is it routinely assessable?	Assessment task
Trait 9 : Skills and their application in employment						
Information and IT skills	Locate, organise and evaluate information with emphasis on primary sources	Demonstrates developing capacity to locate and organise new information from primary sources and use it to support an argument.	Demonstrates excellent capacity to locate and organise new information from primary sources and use it to support an argument.	Cognitive/ Psychological	Yes	• Research paper • Examination • Annotated bibliography • Essay with references
	Apply information and communication and other relevant technologies to solve problems and in learning tasks	Demonstrates developing capacity to utilise ICT to solve problems and apply it to own learning	Demonstrates excellence in utilising ICT to solve problems and applying it to own learning	Cognitive/ Psychological	Yes	• Problem involving searching for and utilizing a new technology • Note books/workbooks • Modelling tasks • Online study

Attribute cluster (Tier 2)	**Learning outcome (Tier 3)**	**Level 1 standard (Developing)**	**Level 3 standard (Exemplary)**	**Cognitive or affective outcome/type of assessment**	**Is it routinely assessable?**	**Assessment task**
Employ-ment skills	Demonstrate knowledge of regulatory frameworks and political influences for professional area of study	Demonstrates developing knowledge of regulatory frameworks and political influences on professional practice	Demonstrates excellent knowledge of regulatory frameworks and political influences on professional practice	Cognitive/ Psychological and Behavioural	Yes	• Essay • Quiz • Short answer test • Problem relevant to the discipline • Practicum

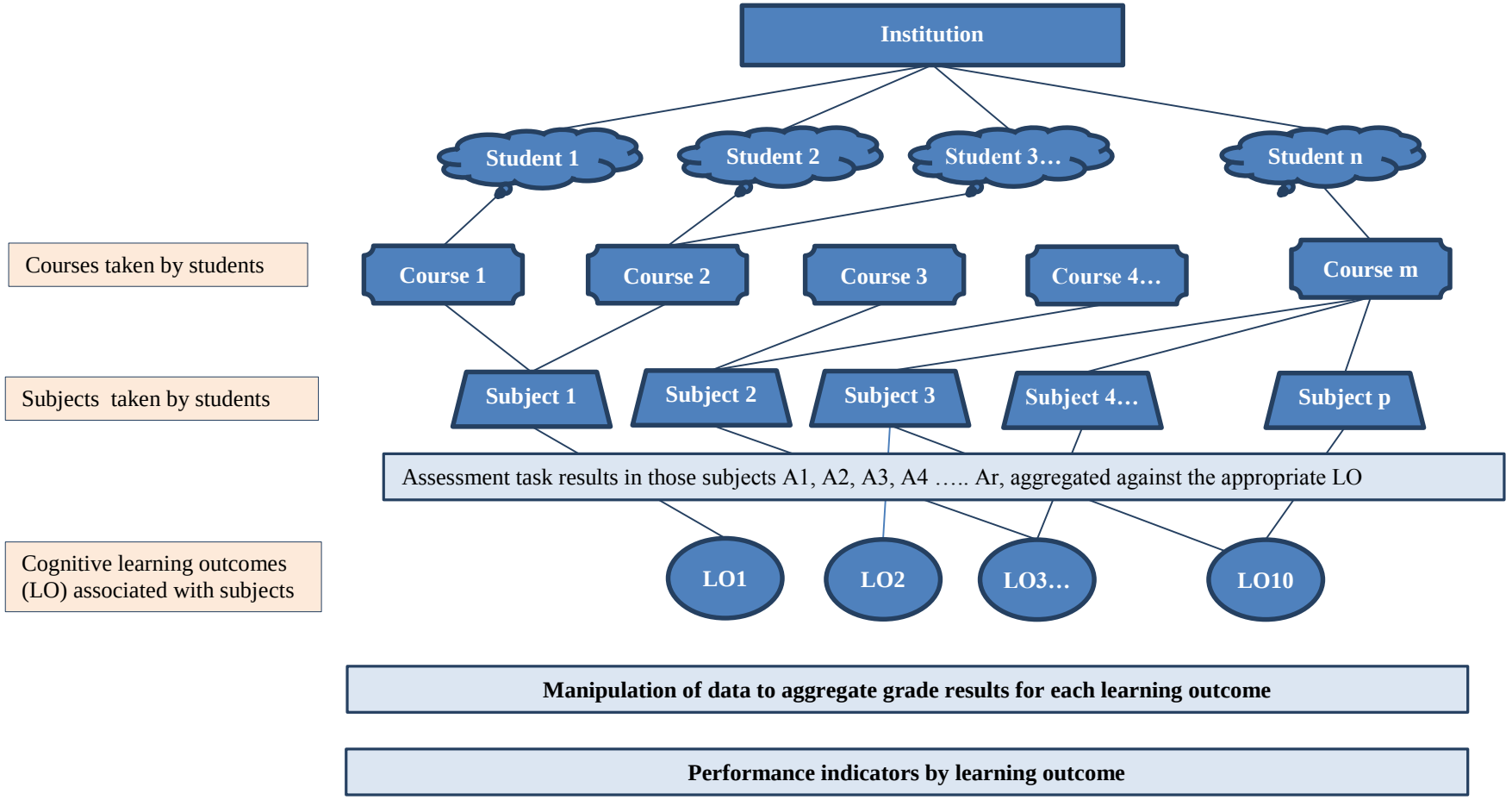
APPENDIX 7.1 – Technical details on compilation of data on student assessment for performance indicators

Data structure diagram

Figure A.7.1 shows the complexity of the networks which govern how the data to enable the capture of the necessary grades result for use in a performance indicator might be structured. It highlights how in the model proposed for furthering the measurement of the cognitive learning outcomes shown in Figure 6.6, a single learning outcome might be linked to a number of different subjects and from that to several different courses and then to individual student enrolment records.

This shows the need to be able to trace back through this networks of linkages to get a clear picture of the relationship between the student's result in subject based assessment tasks and achievement of each learning outcome.

Figure A7.1 – Data structure and flows for results records for students by cognitive learning outcome



Cognitive learning outcomes - Structure of file

The results data file for student learning outcomes achievement would contain elements and a structure similar to the following:

Table A7.1 – Possible structure of a data file to measure achievement of cognitive learning outcomes

File element	Data specification
Teaching period and year of data	Year followed by identifier for teaching period used in other data collections
Institution code	Unique identifier 1-40
Institution name	Text
Student ID	Student identifier used in other student data collections such as enrolment and completions
Course code	Course in which student is enrolled Text
Course name	Course in which student is enrolled Text
Subject in course contributing to learning outcome	Subject code
Assessment item used in the subject	Code linked to subject
Discipline code of subject contributing to learning outcome	6 digit identifier used in other Government statistical collections (field of education)
Learning outcome code	Unique identifier 1-10
Learning outcome name	Text
Student assessment task grade result	Value of 0-3 for this student in this subject and for the assessment item for the learning outcome
EFTSL assessment task grade result	EFTSL value for this student in this subject for the learning outcome

Such a structure is flexible and would allow the data file to be processed in each year (for example results aggregated) by:

- Sector

- Institution
- Discipline within and between institutions
- Course
- Institution and discipline; and
- Learning outcomes across institutions or disciplines.

Affective learning outcomes - Structure of file

The data obtained from the QILT surveys would consist of a file of unit records at the individual student level with the following structure:

Table A7.2 – Possible structure of a data file to measure achievement of affective learning outcomes

File element	Data specification
Year of data	Year survey conducted
Institution code	Unique identifier 1-40
Institution name	Text
Student ID	Unique student identifier
Course code and name	Text
Discipline code of course student is undertaking	6 digit identifier used in other Government statistical collections (field of education)
Affective learning outcome code	Unique identifier AF1 or AF2 or AF3
Affective learning outcome name	Text
Relevant Student survey response	One of (AF1 –collaborative learning, AF2 – teamwork , AF3 - independent learning)
Value of response of student	Score 0 (not applicable), 1-5