

June 2008

Final Report

STUDENT SURVEYS ON TEACHING AND LEARNING

The University of Sydney

Associate Professor Simon Barrie

Dr Paul Ginns

Rachel Symons

Acknowledgement:

This project would not have been possible without the generous engagement with the study by colleagues working with student surveys of teaching in many Australian universities. These colleagues kindly contributed to the study by completing surveys, commenting on drafts of this report and through their participation in workshops. Their suggestions and insights have proved invaluable. Their scholarship and their willingness to work together provide the necessary platform on which the next generation of student evaluation of teaching systems will be developed in Australian universities.

On behalf of the project team – Thank You.

Support for this project has been provided by the Australian Learning and Teaching Council, an initiative of the Australian Government Department of Education, Employment and Workplace Relations. The views expressed in this report do not necessarily reflect the views of the Australian Learning and Teaching Council Ltd.

This work is published under the terms of the Creative Commons Attribution-Noncommercial-ShareAlike 2.5 Australia Licence. Under this Licence you are free to copy, distribute, display and perform the work and to make derivative works.

Attribution: You must attribute the work to the original authors.

Noncommercial: You may not use this work for commercial purposes.

Share Alike. If you alter, transform, or build on this work, you may distribute the resulting work only under a licence identical to this one.

For any reuse or distribution, you must make clear to others the licence terms of this work.

Any of these conditions can be waived if you get permission from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/2.5/au/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

Requests and inquiries concerning these rights should be addressed to the Australian Learning and Teaching Council.

Level 14.300 Elizabeth Street, Surry Hills NSW 2010 Australia

PO Box 2375 Strawberry Hills NSW 2012 Australia

Telephone 02 8667 8500

Facsimile 02 8667 8515

www.altc.edu.au

ABN 30 109 826 628

2010

Contents

This report replaces the interim report on this project published in October 2007.
It is presented as 9 chapters in three sections, representing the different stages of the study.

This report scopes Australian universities' use of student evaluation of teaching surveys in the context of a larger study which is developing and piloting Teaching Quality Indicators.

Section One: The study context

1. THE STUDY CONTEXT	7
Background and Information	2
Aim	2
Report	8
2. OVERVIEW OF NATIONAL AND INTERNATIONAL STUDENT EVALUATION OF TEACHING SURVEY USE.....	4
Australia	5
England, Wales and Northern Ireland.....	7
United States of America	8
Canada.....	9
Taiwan.....	10
Hong Kong.....	10
Other non-national higher education groups.....	16

Section Two: Exploring Australian universities' student evaluation of teaching surveys

3. EXPLORING AUSTRALIAN UNIVERSITIES' STUDENT EVALUATION OF TEACHING SURVEYS : RESEARCH APPROACH AND STUDY DESIGN.....	13
4. OVERVIEW OF STUDENT EVALUATION OF TEACHING SURVEY USE IN AUSTRALIAN UNIVERSITIES	21
Profile of respondents	21
Overview of current survey practice in Australian universities	21



Contents

Level 1: Individual	22
Level 2: Department/School/Faculty.....	23
Level 3: Institution	25
Level 4: Across the Sector	26
Overview of dimensions of teaching probed in relation to organisational level...	27
5. ISSUES AND IMPLICATIONS ARISING FROM CURRENT SURVEY PRACTICE	42
Commonalities across SET surveys	42
Commonalities across SET practices	44
 Section Three: Exploring institutional SET practice in relation to the proposed Teaching Quality framework	
6. MOVING FORWARDS: OUTLINE OF STAGE TWO	48
Context: A new agenda for teaching and learning in Australian universities	49
Mapping the TQI dimensions	52
Engaging institutional SET experts: National workshop.....	62
7. DEVELOPING BENCHMARKING SCALES AND ITEMS FOR THE TQI FRAMEWORK.....	64
8. KEY ISSUES FOR PILOT UNIVERSITIES SEEKING TO USE SET DATA IN RELATION TO THE TEACHING QUALITY INDICATOR DIMENSIONS.....	94
Using existing SET data	94
Developing a new generation of SET data.....	94
A framework for scale & item development and testing	95
Guidelines for aggregating SET data.....	98
Developing new SET processes.....	98

Contents

Suggestions for appropriate alternative and complementary data sources to be considered	100
Suggestions for fostering a new approach to benchmarking using SET data...	101
9. CONCLUDING COMMENTS AND SUGGESTIONS	102
10. REFERENCES	104
11. APPENDICES	109
Teacher Level Items	110
Subject Level Items	111
Course/Degree Level Items	113
University Level Items	114
Additional Dimensions and exemplar items	114
Learning Environment:	115
Student Motivation and Effort.....	116

Section One

The study context



1. THE STUDY CONTEXT

Background and introduction

Data on students' experiences of teaching is a central component of most Australian and many international sets of university teaching performance indicators. As with any quality assurance indicator, the Student Evaluation of Teaching (SET) indicators reflect a range of variables including implicit and explicit beliefs about what constitutes quality teaching or learning in particular contexts, and hence *what* is important to be measured, beliefs about *who* should do the measurement, and what the measurement might *mean*.

Student evaluation of teaching data plays a more prominent role in Australian sets of teaching performance indicators than it does in most other countries. In Australia the centrality of student evaluations of teaching to both institutional and national quality assurance strategies reflects in part the pervasive influence of the student focused learning perspective on conceptions of university teaching quality in this country over the past twenty years, and in part the shift in the sector towards seeing students as 'clients' and 'consumers' of higher education 'services'.

While Australia has, in the Course Experience Questionnaire (CEQ), one of the most theoretically sound, psychometrically validated and comprehensive national SET systems of any country, the same cannot always be said of universities' own SET practices. To a large extent these have remained idiosyncratic institutional practices, developed within universities and operating independently of any national system and usually without reference to each other. The variations between universities are often also reflected within institutions with different disciplines and faculties often having their own independent SET surveys and systems. Until recently, the SET systems that had evolved within Australian universities rarely had an explicit theoretical basis or published psychometric data; however, they did have a degree of face-validity in their ongoing use.

This situation has now changed somewhat and in recent years most universities have developed internally coherent institutional SET systems. Often this has occurred within the context of new institutional and government quality assurance processes. Recent national imperatives, such as the requirement of the first stage of the Learning and Teaching Performance Fund (LTPF) for publicly available institutional SET results have also contributed to changes in universities' SET practices in terms of survey tools and procedures. The development and piloting of new sector-wide surveys such as the Australasian Survey of Student Engagement (AUSSE) (see p.12), will no doubt also prompt further changes in Australian universities' internal SET practices.

Aim

The aim of this study was to explore current institutional SET practices in Australian universities with a view towards developing a framework that would assist in making sense of the vast array of SET data already collected by universities. From the perspective offered by this overview and framework, the study sought to provide some

initial resources and support for those universities looking to use their own SET data in the context of the emerging teaching quality framework being developed by the Teaching Quality Indicators (TQI) project funded by the Carrick Institute.

The first stage of the study provided an overview of national SET surveys used in Australia and selected other countries and a description and preliminary analysis of current SET practice in Australian universities in relation to the findings of a recent major review of the research literature on SET multi-section validity studies by Abrami, P.C., d'Apollonia, S., & Rosenfield, S. 'The dimensionality of student ratings of instruction: What we know and what we do not.' (2007). The second stage of the study linked the descriptive framework derived from the analysis based on the Abrami et al. dimensions to the dimensions of teaching quality proposed in the overarching Teaching Quality Indicators project. The study then provided some initial suggestions for potential sources of SET data in relation to the TQI dimensions. The second stage of the study provided further suggestions for how universities might move forwards in establishing SET systems and procedures which would enable institutional benchmarking using internal SET data.

Report

This study provides an environmental scan and initial analysis of within-university SET practices across 29 Australian universities. The report is based on data collected in the first half of 2007 and as such represents a 'snapshot' of institutional SET practice which may well have changed since. The report analyses the core items drawn from internal student evaluation of teaching surveys currently in use in these Australian universities using a framework derived from a key review of the major multi-section validity studies of SET. It considers patterns of use of SET items in Australian universities and suggests how the analytic framework might be developed for use in the Australian context. Based on the findings of the analysis the report provides some preliminary conclusions and suggestions to better enable institutions to use internal SET data for benchmarking and quality assurance. The report also identifies a range of validated survey scales which could be used to gather SET data in relation to the proposed TQI dimensions (Chalmers 2007) and identifies a network of SET experts with the potential to collaborate in developing and validating shared SET items, scales and procedures, for those levels of the proposed TQI dimensions where no suitable validated SET scales were identified.

Follow-up

This report does not provide prescriptive recommendations as its intention is to support discussion on the current SET use. However in the context of the overarching Teaching Quality Indicators project a provisional set of survey items will be identified for each of the proposed TQI dimensions for use at the level of individual subjects. The items will be posted on this website shortly and will provide the basis for further consultative development and validation by universities participating in the Teaching Quality Indicators project.

Comments on this report are welcomed. Please contact A/Prof Simon Barrie at The University of Sydney (S.Barrie@usyd.edu.au).

2. OVERVIEW OF NATIONAL AND INTERNATIONAL STUDENT TEACHING SURVEY USE

This section provides a summary and critical analysis of the features of the SET tools currently used in Australia and other countries at ‘whole of sector level’ as a backdrop to the use of SET within Australian universities.

Collecting student feedback on teaching, focusing either on a teacher or subject, has been a common practice for several decades in many higher education institutions around the world. In contrast, the development of surveys to collect such data at the level of the whole degree or institution has been a more recent occurrence. The systematic use of such surveys to gather data across several institutions or a whole sector is a relatively recent development in many countries.

Australia has a longer history of national data collection in relation to students’ experiences of university teaching and learning than most countries, having administered the Course Experience Questionnaire since 1993. The original survey was extended with additional scales in 2001 and a new survey, the Australasian Survey of Student Engagement (AUSSE), based on an American survey, the National Survey of Student Engagement (NSSE), is currently being trialled in around 30 Australian and New Zealand universities. The UK has only recently introduced a national survey, the National Student Survey (NSS) in 2005. There is no equivalent survey currently used across the European Union; although some member states may have such surveys, there is no evidence of the systematic use of these at a national level. To date, New Zealand has only piloted their Learner Opinion Survey. North American universities typically collect such data on a state by state basis though there are some surveys, such as the National Survey of Student Engagement (NSSE) which are used across several states and also in other countries such as Canada.

National uses of such student evaluation surveys for quality assurance purposes presuppose certain features.

1. The surveys should have demonstrated psychometric reliability and validity.
2. The surveys explicitly articulate a particular perspective on what constitutes ‘quality’ teaching and learning. In some cases this perspective is simply an agreed set of values about what is ‘good’ teaching, in others it is an empirically derived theoretical perspective on teaching and learning. The nature of the underlying perspective on teaching and learning has implications for how the results of such surveys can be used to drive evidence-based policy development and teaching improvements.

The following section describes, in brief, the recent history and state of play in whole sector use of student feedback surveys in several countries.

Australia

The Australian Graduate Survey (AGS). This comprises two parts, the Graduate Destination Survey and the Course Experience Questionnaire.

Graduate Destination Survey

Australian universities have administered the Graduate Destination Survey (GDS) since 1972, under the guidance of Graduate Careers Australia. The GDS is sent to all students who complete requirements for a degree in Australian universities. It focuses on details of current employment or study, as well as questions related to job search strategies. Traditionally, this data has been used by universities to advise both prospective and current students, and staff, about employment opportunities in different fields of education. More recently, two GDS variables, percentage of (Australian citizen/permanent resident bachelors) respondents in full-time work and further study, have been used in the National Learning and Teaching Performance Fund. As of 2007, the GDS has been renamed the Australian Graduate Survey (AGS).

Course Experience Questionnaire

Australian universities have administered the Course Experience Questionnaire (CEQ) since 1993, as part of the AGS. The CEQ was developed by Professor Paul Ramsden (Ramsden, 1991; Wilson, Lizzio, & Ramsden, 1997) as a teaching performance indicator, focusing on aspects of the classroom teaching environment which previous research had found are linked to deep and surface approaches to learning, and higher quality learning. These scales include Good Teaching; Clear Goals and Standards; Appropriate Assessment; and Appropriate Workload. The CEQ also includes an outcome scale, Generic Skills, and an “Overall Satisfaction with Course Quality” item.

Arguing that the CEQ is somewhat limited by its focus on in-classroom experience, Griffin, Coates, McInnis, and James (2003; p.260) have argued “...the original CEQ was based on a theory of learning which emphasises the primary forces in the undergraduate experience as located within the classroom setting”. In order to expand the range of performance indicators available to institutions, Griffin et al. developed an expanded range of CEQ scales, reflecting features of contemporary higher education settings beyond classroom settings. The expanded scales focus on Student Support, Learning Resources, Course Organisation, Learning Community, Graduate Qualities, and Intellectual Motivation.

Since 2002, Australian universities have been required, at a minimum, to collect graduate responses on the Good Teaching and Generic Skills scales, and the Overall Satisfaction Item. In addition, universities have the choice to also collect data using either the additional core CEQ scales, the extended scales, or a combination of both, subject to the limitation that the selected items take up no more than one page of the AGS.

Traditionally, components of the AGS, and the CEQ scales, have been intended for benchmarking teaching quality primarily at the degree level, allowing tracking over time of the quality of a specific degree, as well as benchmarking similar programmes at different institutions. The development by the Australian Commonwealth Government of the

National Learning and Teaching Performance Fund (NLTPF) has seen components of the AGS and CEQ being used for performance-based funding of institutions, and more recently, cognate disciplines within institutions. This use of the AGS/CEQ has prompted intense discussion within the Australian higher education sector, given concerns over differential survey practices and response rates between institutions. To address these concerns, the Department of Education, Science and Training (DEST) commissioned the Graduate Destination Survey Enhancement Project (Graduate Careers Australia, 2006). The broad goals of this project were to "...design and develop the processes, resources, and ideas needed to generate a new era of research into Australian student experiences and graduate outcomes" (Graduate Careers Australia, 2006; p.xxi), and thus improve both the quality of responses to such surveys and confidence in their findings and usage (particularly with regards to the NLTPF). At present, the sector is still debating the precise form of AGS standardisation.

The CEQ is one of the most extensively validated student feedback surveys of its type in the world. Unlike some surveys it is explicitly based on a well-researched theoretical model of learning. This strength is also a potential weakness. The model of learning on which the survey is based recognises that learning is a complex process, and the CEQ focuses on student perceptions as a key indicator of this process. Many of the uses made of the CEQ data ignore this complexity. In addition, while student perceptions are important, they are not the only aspect of a quality teaching and learning experience. However, there are few surveys in use around the world that allow such clear connections to be made between student perceptions data and an extensive and evolving body of research on the student experience. As such, it provides a sound basis for evidence based policy and teaching enhancement activities. Moreover, the presence of fifteen years of national data provides Australia with unrivalled trend data for comparison purposes.

First Year Experience Questionnaire

The First Year Experience Questionnaire (FYE) has been administered at five-year intervals since 1994, by the University of Melbourne's Centre for the Study of Higher Education (Krause, Hartley, James, & McInnis, 2005). Surveying a stratified sample of first-year students of 7 universities in 1994 and 1999, and 9 universities in 2004, its goal is to "assemble a unique database on the changing character of first year students' attitudes, expectations, study patterns and overall experience on campus" (Krause et al., 2005; p.1). It draws on the CEQ for much of its content. In addition, the 2004 FYE included items and scales focussing on student engagement, and the role information and communication technologies can play in student engagement. Unfortunately, the response rate for the 2004 survey was only 24%, raising concerns about the representativeness of the most recent findings.

This survey is important as it has provided the evidence base for many universities' strategies to improve university transition and first year retention and progression. Most Australian universities gather data from first year students using a variation of this survey or an internal adaptation of the CEQ.

Australasian Survey of Student Engagement (AUSSE)

The Australasian Survey of Student Engagement (AUSSE) is currently being piloted by the Australian Council for Educational Research (ACER). Drawing on the National Survey of Student Engagement (NSSE) described below, the AUSSE is based on a definition of engagement as *“students’ involvement with activities and conditions likely to generate high quality learning”* and is based on the assumption that *“...learning is influenced by how an individual participates in educationally purposeful activities. While students are seen to be responsible for constructing their knowledge, learning is also seen to depend on institutions and staff generating conditions which stimulate and encourage student involvement”* (ACER, 2007). The AUSSE includes items loading on 5 scales on the “Student Engagement Questionnaire”: Active Learning, Academic Challenge, Student and Staff Interactions, Enriching Educational Experiences, and Supportive Learning Environment.

England, Wales and Northern Ireland

National Student Survey

The National Student Survey (NSS) has been used by universities in England, Wales and Northern Ireland since 2005 to assist prospective students in making choices; provide a source of data for public accountability; and assist institutions in quality enhancement activities (Sharpe, 2007). It is administered to students in their final year of study. The NSS drew on the CEQ for its conceptual foundation, in emphasising student perceptions of the learning environment and subsequent impacts on learning outcomes. The first iteration of the NSS included 6 scales: quality of teaching, assessment and feedback, academic support, organisation and management, learning resources, and personal development, as well as an overall satisfaction item. The second iteration tested two additional scales, learning community and intellectual motivation, adapted from the CEQ. Lastly, in 2007, individual institutions will be able to pilot test scales a subset of 10 additional scales, e.g. careers, course content/structure, workload, the physical environment.

Sharpe (2007) notes that “Being based on the CEQ, the theory-base of the NSS is the same as for the CEQ, i.e., it emphasises the importance of students’ perceptions of their learning context and the impact of this upon their learning outcomes.” (p.10). However, at present, unlike the CEQ, these theoretical linkages have not been empirically tested.

Destinations of Leavers from Higher Education (DLHE).

Used to collect information on the activities of students following departure from a higher education institution, the Destinations of Leavers from Higher Education (DLHE) has been used since 2002/03, and replaces the former First Destinations Supplement (FDS) used between 1994/95 and 2001/02. It is managed by HESA (Higher Education Statistics Agency), and is carried out roughly 6 months after the student completes his/her degree. The data collected and disseminated is similar in kind to that collected by Australia’s AGS.

First Year Experience Survey

Yorke et al (1997) investigated the experiences of first year students, and reasons for discontinuation, in six institutions in England's north-west. Building on this study, Yorke & Longden (2007) report a two-phase survey project investigating the experience of first year students in 25 universities in the United Kingdom, spanning a range of institution types and nine broad fields of study. A sampling frame was used which ensured no institution was asked to survey more than 3 fields of education. The first phase investigated the experiences of first year full-time students following their first semester, while the second phase (beginning in January 2007) surveys ex-first year students about their reasons for discontinuing.

Survey items (Likert format) used in the first phase were not designed to necessarily reflect specific scales. However, principal components analysis of the data suggested at least 5 scales with adequate psychometric properties. These were labelled; understanding the academic demand; supportive teaching; stimulating learning experience; feedback; and coping with academic work. While sub-group descriptive statistics are presented in the report, inferential statistics are not, due to concerns about the adequacy of the sampling frame and substantial variations in response rates across institutions. The exclusion of part-time first year students from the sampling frame also limits the potential conclusions that can be drawn from the survey.

United States of America

Cooperative Institutional Research Program (CIRP)

The Cooperative Institutional Research Program (CIRP), a national longitudinal study of the American higher education system, was established in 1966 by the American Council on Education, and is now administered by the Higher Education Research Institute. CIRP provides several complementary surveys, two focused on freshmen (CIRP Freshman Survey, and Your First College Year – YFCY) and one for seniors (College Senior Survey - CSS), which focuses on students' academic and personal development.

The CIRP Freshman Survey gathers data on a wide range of demographic characteristics, as well as receipt of financial aid; secondary school achievement and activities; educational and career plans; and values, attitudes, beliefs, and self-concepts (CIRP, 2007). The broad goal of the YFCY survey is to *"identify features of the first year that encourage student learning, involvement, satisfaction, retention and success, thereby enhancing first-year programs and retention strategies at campuses across the country"* (Your First College Year, 2007). The CSS acts as an "exit" survey for senior students, gathering data on student academic and campus life experiences, as well as post-college plans; its alignment with either of the above surveys allows for longitudinal analyses, particularly student cognitive and affective change.

National Survey of Student Engagement

The National Survey of Student Engagement (NSSE) was developed in the late 1990s, and is now used by over 1100 US 4-year colleges and universities. The NSSE:

“is designed to obtain, on an annual basis, information from scores of colleges and universities nationwide about student participation in programs and activities that institutions provide for their learning and personal development. The results will provide an estimate of how undergraduates spend their time and what they gain from attending college. Survey items on The National Survey of Student Engagement represent empirically confirmed “good practices” in undergraduate education. That is, they reflect behaviors by students and institutions that are associated with desired outcomes of college” (NSSE, 2007a).

Kuh (2001) describes 4 main factors underlying students responses to the core 22 items representing activities in which students engage inside and outside the classroom: student-faculty activities, student-student activities, activities reflecting diversity, and classwork activities. Three factors were found to underlie student responses concerning educational and personal growth: personal-social, practical competence, and general education. Three factors were found to underlie responses to items tracking opinions about the school: quality of relations, the social climate of campus, and the academic quality of the campus.

Several other surveys have been designed within the same theoretical framework as the NSSE. These include the Beginning College National Survey of Student Engagement (BCSSE), which measures “entering first-year students’ pre-college academic and co-curricular experiences, as well as their interest in and expectations for participating in educationally purposeful activities during college” (NSSE, 2007b); the LSSSE, for students of Law schools; the HSSE, for High School students; and the CCSSE, for students of community colleges. The NSSE was developed from the College Student Experience Questionnaire (CSEQ). The CSEQ explicitly links students’ ratings of their participation in teaching and learning activities (student engagement) with self-report of progress towards development of key outcomes.

While there is a steadily increasing amount of research related to the measurement properties and institutional use of the above Student Engagement surveys (see NSSE, 2007c for a current list), at present, the theoretical and empirical connections between the Student Engagement and Student Learning (e.g. Ramsden, 1991) approaches to student feedback are just beginning to be elucidated. For instance, Laird, Shoup, and Kuh (2005) used NSSE data to develop a measure of “deep learning” activities (with “higher order learning”, “integrative learning”, and “reflective learning” sub-scales), and investigated field of study differences in relations between deep learning and outcomes (student gains in personal and intellectual development, self-reported grades, and satisfaction with the college experience).

Canada

NSSE

The NSSE, described above, was also used in 17 Canadian universities or colleges in the 2007 round.

British Columbia College and Institute Student Outcomes Survey

Graduates of British Columbia's public colleges, university colleges, and institutes are contacted by telephone between 9 and 20 months after they complete their programmes, and invited to respond to the British Columbia College and Institute Student Outcomes (CISO) Survey. The survey commenced in 1988. BC Stats manages the collection of student outcomes information on behalf of the Outcomes Working Group (OWG), representing the Ministry of Advanced Education (AVED). Uses of the Survey include supporting institutions in evaluating and improving programmes and services; assisting prospective students in their programme choices; and enhancing understanding of the education and labour markets.

Beginning in 2001, the Survey has been modified to include scales and items focusing on "learner-centred practice". The framework used consists of 5 factors: Learner and Learning Support Services; Teaching and Learning Processes; Curriculum; Campus Life; and Learning Gains.

Taiwan

A standardised instrument for investigating Taiwanese students' perceptions of their institutions' learning environments has been developed by Huang (2006). The College and University Environments Inventory (CUES-I) consists of 7 scales: student cohesiveness, faculty-student relations, administrative support, language abilities, emotional development, library resources, and student services. Huang (in review) describes the use of the survey to explore relations between the above dimensions and student academic aspirations and satisfaction in a random sample of 12,423 students in 42 Taiwanese universities. The study reported statistically reliable relations between the above sets of variables at both the individual level and the aggregated university level.

At present, the above instrument has not progressed beyond use as a research instrument. The scope of the field testing to date suggests the CUES-I might be useful for benchmarking in the Taiwanese higher education sector in future.

Hong Kong

A survey of employers of students of three Hong Kong higher education providers (City University of Hong Kong, the Hong Kong Polytechnic University and the Vocational Training Council) was conducted in 2002 and 2003 by the Education and Manpower Bureau. The goals of the survey were to obtain the opinions of employers about full-time publicly funded sub-degree graduates, regarding graduate attributes such as language proficiency, numeracy, IT literacy, analytic and problem-solving ability, work attitude, interpersonal and management skills, and technical job-related skills. Employers were also invited to suggest ways of improving the quality of graduates.

This survey is focussed only on outcomes. It uses employer ratings rather than self-assessment by graduates. The survey therefore presents a view of student learning in

terms of employability and ignores the role of higher education in developing individuals capable of contributing as agents of social good.

Other non-national higher education groups:

There are now several consortia of universities that represent affiliations of universities based on type of institution rather than national boundaries. The two most familiar of these in Australia are the AC21 and Universitas 21. The AC21 has begun to explore the possibility of using a common student feedback instrument across member institutions to gather data on student experiences for benchmarking purposes. The survey being used for this is The University of Sydney's 'Student Course Experience Questionnaire' (SCEQ), which is an adaptation of the CEQ for use with currently enrolled students. The SCEQ has been trialled at Nagoya University in 2005 and 2006 and there has been discussion about trialling the survey in a North American university next. The SCEQ has also been used at other universities outside Australia such as Oxford in the UK.

The surveys discussed above are the source of the validated scales suggested in chapter seven of this report. Reviews of additional surveys are available in Chalmers (2007).

Having considered the broad international and national backdrop against which SET practices in Australian universities have emerged, we will now turn to a consideration of the uses of student evaluation of teaching in different Australian universities and an analysis of the SET survey items currently in use.

Section Two

Exploring Australian universities' student evaluation of teaching surveys

3. EXPLORING AUSTRALIAN UNIVERSITIES' STUDENT EVALUATION OF TEACHING SURVEYS: RESEARCH APPROACH AND STUDY DESIGN

This section describes the research approach adopted in investigating Australian universities' current uses of Student Evaluation of Teaching (SET) surveys and in the framework applied in conducting the preliminary analysis of the items used in these surveys.

The research approach sought to recognise that student evaluation of teaching (SET) practices and student evaluation of teaching surveys are complex cultural activities embedded in institutional history and politics. These SET systems have typically evolved over many years, being shaped by different individuals' explicit and implicit conceptions of teaching and learning as well as institutional factors that range from staff performance management policies to university responses to external factors such as the AUQA audits and accreditation.

As the culture of the university was considered to be an important factor in relation to the overall TQI project's goals of developing an approach to teaching quality that was based within the diverse institutional cultures of Australia's universities, it was agreed that an analysis of the student evaluation of teaching policies alone would not necessarily convey the nuances of distributed practice across the university community nor provide a basis for engagement by these communities with the products of the research. Pragmatic issues also shaped the research approach such as the time available to provide a preliminary report, and the challenges of dealing with the large volume of qualitative data represented by item banks of SET survey items (up to 500 items in some cases) and potentially detailed descriptions of the multiple layers at which SET operates in different institutions.

As such, the researchers adopted a strategy which sought to engage key individuals in each university as collaborative participant researchers. A network of these individuals existed through institutional participation in the annual Evaluation forum, an Australasian conference on SET. These individuals were known to be familiar with both the overall SET systems in their institutions and the actual surveys used and represented an existing community of practice that would readily engage with the project. Universities were contacted through the relevant DVC(A) and asked to nominate an individual to act as liaison person to provide information about the university's SET practices, and in most cases the individuals active in the Evaluation network were the people nominated. The study design aimed to engage the nominees who chose to participate in the study in an iterative process of describing, reflecting and analysing current SET practice. The process was iterative as the initial requirement was for an overview within a short time frame. To meet this requirement while still laying the groundwork for a more detailed analysis, a two stage data collection process was developed. The aim of the first stage of this process was to provide an overview of current SET practice and an initial analysis



of the SET survey items in current use at different levels of a representative sample of Australian universities. The results of this initial data collection were used as the basis for the second stage of additional data collection and analysis. The approach to data collection and analysis was intended to provide a basis for engaging participants more fully as co-researchers in any future studies.

The initial phase of the data collection asked respondents to complete two online surveys. The first of these surveys asked respondents to provide a brief overview of current practice in relation to SET at four different levels of their institution.

1. The first level asked how **individuals** use SET (for example lecturers tutors etc)
2. The second level asked how **departments / schools / faculties** use SET
3. The third level asked how SET was used at the **whole institution** level and
4. The fourth level asked about how the university used SET **across the sector** (national and international benchmarking) or part of the sector (for instance, collaborations with other ATN universities).

At each level the open response sought by the survey was semi-structured under two questions with a series of associated trigger prompts:

1. **Purpose & use:** Please describe the sorts of surveys and survey practices as well as the ways student feedback data are used at this level.
2. **How is the data interpreted?** Please describe how the data is aggregated and how comparisons are made. Are standards applied? What are these? How are benchmarks or averages decided? Who has access to the results?

A model answer was provided to give an indication of the level of detail the survey sought to elicit.

Table 1: Illustrative survey 1 response provided as example to respondents:

Individuals	Purpose & use:	To provide voluntary and confidential feedback to a particular staff member (e.g. lecturer, tutor, demonstrator) for improving teaching. Used for promotion applications, teaching award applications, PMD. Only the staff member involved can order the survey, receive the data and that person is responsible for acting on feedback.
	How is the data interpreted	There is no aggregation of different individuals' results. We do not provide faculty means or any other comparison – there is scope to aggregate multiple instances of an individuals teaching on request – this is rare. HoDs / or any third party cannot request this data from a unless Where staff have multiple orders (e.g. teaching multiple tutorials within the same unit of study), an aggregate report is also provided
Departments / Schools / Faculties	Purpose & use:	Schools, departments and faculties use surveys at both the unit of study level and the degree level for quality assurance/improvement, informing curriculum design, and accreditation purposes. Schools, departments and faculties are responsible for ordering of unit of study evaluations, using either a standard central form or another faculty-approved form, and determining the frequency of evaluation of a given unit (at least once every 3 years). Degree-level surveys are administered centrally; the CEQ is administered every year and an internal version (for currently enrolled students) every 2 years.
	How is the data interpreted	In addition to individual unit results, unit of study-level data is aggregated for an entire school/department for a HoD, or an entire faculty for a Dean. Degree-level data is aggregated to faculty level, and to school level for one faculty. Schools, departments and faculties are responsible for setting desired standards for both unit and degree level surveys. Comparisons are often made on unit of study-level data within schools, departments and faculties to identify excellent units and those that may be in need of improvement. Comparisons are often made on faculty-aggregated degree level results with cognate faculties. The unit of study coordinator receives a report for his/her unit as well as the survey forms. Heads of schools/departments, Deans of faculties, and the Deans' representatives on the University's Evaluation and Quality Assurance Working Group receive reports detailing average results for each unit surveyed in their school/department/faculty. Degree-level results are publicly accessible, given at least 5 respondents.
Whole Institution	Purpose & use:	Primarily degree-level surveys (e.g. Course Experience Questionnaire, Student Course Experience Questionnaire) are used for evaluation of success of whole institution initiatives (e.g. Special Projects). Major uses of these surveys include competitive performance-based funding of teaching and quality audits (e.g. AUQA). Both faculty-aggregate unit of study level and degree-level data can be used for Academic Board reviews of faculties.
	How is the data interpreted	Data are aggregated at levels meaningful to the research question (e.g. across the whole university, have the experiences of commencing first year students changed?) and comparisons made over time. No institution-wide standards are used, because of field of study differences. Degree-level results are publicly accessible, given at least 5 respondents.
Across the Sector	Purpose & use:	Degree-level survey results (e.g. Course Experience Questionnaire, Student Course Experience Questionnaire) are used for internal competitive performance-based funding of teaching. Results of the CEQ are also used in the NLTPF.
	How is the data interpreted	For internal competitive performance-based funding of teaching, degree-level survey results are aggregated at the faculty level. Faculties are compared with Group of 8 cognate academic organisational units on scales from the above surveys and other teaching performance indicators. Deans are provided with a summary of results. The National Learning and Teaching Fund is administered by DEST



The second survey gathered preliminary data in relation to the nature of the SET items used by different Australian universities at different levels of the institution. The design of the second survey used a framework to structure data entry (responses) and analysis, to assist in making sense of the potentially vast volume of data represented by the national collection of the various SET items used by Australian universities.

The survey structured the data collection in relation to the SET items around four key institutional uses of SET suggested in a review of the research literature.

1. The first gathered SET survey items about teachers noting that these could be drawn from surveys about lecturers, tutors, demonstrators etc.
2. The second section gathered SET survey items about subjects noting that these might be known by other names such as 'units of study', modules, themes etc.
3. The third gathered SET survey items about whole degree courses noting that some items from surveys such as the Course Experience Questionnaire might be relevant here.
4. The fourth gathered SET items that addressed the whole university experience rather than just courses, noting that items from surveys such as the National Survey of Student Engagement (NSSE) and items from CEQ scales such as the Learning Community Scale might be considered here.

At each of the four levels, the survey structured collation and subsequent analysis of exemplar SET items by coding them against 42 key dimensions of teaching and learning. These 42 dimensions were structured under 6 categories, with 3 – 12 'specific' dimensions and one 'global' dimension under each category. The six categories were:

1. Teacher predispositions / personality
2. Course preparation and organisation
3. Approaches to teaching and teaching strategies
4. Quality of learning outcomes
5. Learning climate
6. Assessment

The 42 dimensions used in the survey were based on the findings of a review of the major multi-section validity studies by Abrami, d'Apollonia, & Rosenfeld (2007). These studies examined the available research literature on SET and, based on the analysis of these, identified the major dimensions of teaching and learning targeted in SET instruments. While the scope of Abrami et al.'s analysis was international, the majority of the research studies on which it was based were situated in the North American higher education context. In addition it is a retrospective analysis – it identifies the dimensions that are historically represented in SET practices. Given the impact of an academic community's conceptions of teaching and learning, as well as the impact of institutional culture, on SET practices, there was no expectation that the dimensions would all prove equally important or relevant in the context of Australian uses of SET. However they would still provide a research-derived starting point for the analysis and the study would also provide the means of validating this influential research-based SET framework in the Australian context. To capture items reflecting additional dimensions which might reflect

either new and emerging conceptions of teaching and learning or unique features of the Australian higher education context, a seventh category “Other” was included.

For each of the four institutional levels (teachers, subjects, courses, whole university), respondents were asked to provide one or two of the most representative items from their core set of SET survey items, against the 42 dimensions. There was no expectation that all 42 dimensions would be present in the SET items of any single institution. Nor was there an expectation that all universities would have discrete SET surveys at each of the four institutional levels. It was also recognised that some survey items might cover more than one dimension and might also be interpreted as applying to more than one level (for instance an item about fair assessment might conceivably apply to both individual teachers and to subject level surveys). Respondents were asked to list these items once only in the most representative dimension and level. Where respondents had more than one or two items for a particular dimension they were able to include these.

Piloting of the second survey indicated it would take between 1 - 3 hours to complete. In recognising the limited time available to potential respondents, and with the aim of encouraging participation in the study, an alternative means of responding to the second survey was offered. In cases where individuals felt unable to complete the longer survey they were invited to complete the first overview survey and to provide the researchers with a copy of their SET surveys or item banks. The research team then took on the task of allocating the items across the dimensions of teaching. This did not provide information as to the respondents' perceptions of the dimensions probed by their university's items. However, there remains potential to explore this in the focus groups in the second phase of the study and it did ensure a representative sample of SET items were included in the first analysis.

The above methodology differs somewhat from most other Australian research into the nature and use of student evaluations of teaching in Australian universities in that it uses a categorisation of SET dimensions derived from a meta-analysis of extensive empirical research. A similarly research based review is that of Marsh (2007) in which he reviews previous research on the multidimensionality of SET surveys, concluding that more such research-based approaches to studying SET are required. A recent interesting Australian study by Davies, Hirschberg, Johnstone, and Lye (2007) surveyed 38 Australian universities regarding their survey practices, and gathered survey items from university websites. Davies et al's description of survey practices classifies practices in terms of features identified by the authors such as online / paper, number of items and helpfully identifies some of the features on which such practices vary. The authors also classified SET survey items based on their own classification taxonomy. This taxonomy encompasses two broad categories – questions about the lecturer and the subject, versus questions about the student and their learning – and 18 more specific categories. The 18 descriptive categories identified by Davies et al. are amongst the dimensions identified by Abrami et al. (2007).

Table 2: The 42 SET dimensions identified by Abrami et al. (2007)

1	Predispositions (Teacher presage – what the teacher brings to the teaching situation)
1.1	<i>General Attitudes</i> : The students are evaluating the instructor's general attitudes. (An attempt is first made to fit items into the other, more specific dimensions. Only if they do not fit elsewhere are they classified here.)
1.2	<i>General Knowledge and Cultural Attainment</i> : The students are evaluating the instructor's general knowledge and cultural attainment beyond the course.
1.3	<i>Knowledge of Domain</i> : The students are assessing the instructor's knowledge of the specific course subject matter and its applications.
1.4	<i>Personal Appearance, Health, and Attire</i> : The students are evaluating the instructor's personal appearance, health and attire.
1.5	<i>Personality Characteristics and Peculiarities</i> : The students are evaluating the instructor's general personality characteristics and peculiarities not directly related to teaching (e.g., maturity, irritability, confidence, paranoia, cynicism, etc.).
1.6	<i>Preparation and Organization</i> : The students are evaluating the extent to which the instructor prepared himself/herself for instruction.
1.7	<i>Research Productivity and Reputation</i> : The students are evaluating the instructor's research productivity and reputation.
1.8	<i>Overall Teacher Predispositions</i> : Students are evaluating the Teacher's overall professionalism*
2	Course preparation and organisation
2.1	<i>Appropriate Use of Methods/Materials</i> : The students are evaluating the extent to which the instructor uses appropriate instructional methods and materials in class, including appropriate use of textbook and tests for learning.
2.2	<i>Choice of Required Materials</i> : The students are evaluating the qualities of the required course materials including text books, assignments etc.
2.3	<i>Choice of Supplementary Materials</i> : The students are evaluating the qualities of the supplementary materials (e.g., film, audio visuals, etc.). That is, they are evaluating whether they were interesting, valuable, or personally relevant. Unless explicitly labelled "supplementary" such materials are considered to be required.
2.4	<i>Workload</i> : The students are evaluating the performance standards and the workload (amount, difficulty) of the course and assignments.
2.5	<i>Overall Course</i> : The students are evaluating the overall worth and quality of the course.
3	Approaches to teaching and teaching strategies
3.1	<i>Answering Questions</i> : The students are evaluating the extent to which the instructor encouraged students to ask questions and responded to students' questions appropriately.
3.2	<i>Vocal Delivery</i> : The extent to which the instructor demonstrated skill in vocal delivery.
3.3	<i>Clarity of Instruction</i> : The students are evaluating the extent to which the instructor delivers clear, concise, understandable and accurate instruction (e.g., lectures, laboratories, etc.).
3.4	<i>Dramatic Delivery</i> : The students are evaluating the extent to which the instructor delivered instruction in an expressive, dynamic, dramatic or exaggerated manner.
3.5	<i>Interaction and Discussion</i> : The students are evaluating the extent to which the



	instructor modelled, encouraged and achieved interactive classes in which both students and instructor contributed to the class.
3.6	<i>Knowledge of Teaching and of Students:</i> The students are evaluating the instructor's knowledge of pedagogy (e.g., knowledge of students, student learning, and/or of instructional methods).
3.7	<i>Monitoring Learning:</i> The students are evaluating the extent to which the instructor monitored students' reactions and taught at the appropriate individual and class level.
3.8	<i>Stimulation of Interest in the Course:</i> The students are evaluating the extent to which the instructor stimulated their interest in the course by using a variety of activities, manifested by the extent to which good attendance, increased interest, outside reading, and liking/enjoyment for the subject matter were exhibited.
3.9	<i>Overall Instructor:</i> The students are evaluating the overall effectiveness of the instructor.
4	Quality of learning outcomes
4.1	<i>High level Cognitive Outcomes:</i> The students are evaluating the extent to which the instructor is promoting high level cognitive outcomes such as writing skills, reasoning, meta cognition, problem solving, etc.
4.2	<i>Low level Cognitive Outcomes:</i> The students are evaluating the extent to which the instructor is promoting low level cognitive outcomes (e.g., recall, recognition, knowledge, etc.).
4.3	<i>Relevance of Instruction:</i> The students are evaluating the extent to which the instructor emphasizes the relevance of the provided information, including recent research.
4.4	<i>Overall Learning:</i> The students are evaluating the overall quality and relevance of the perceived learning that took place including the achievement of short and long term objectives.
5	Learning climate
5.1	<i>Availability:</i> The students are evaluating the extent to which the instructor was available outside of the classroom for assistance or extra curricular activities.
5.2	<i>Concern for Students:</i> The students are evaluating the extent to which the instructor was concerned and helpful about student difficulties
5.3	<i>Enthusiasm for Students:</i> The students are evaluating the extent to which the instructor communicates his/her enthusiasm, interest or liking for students as people.
5.4	<i>Enthusiasm for Subject:</i> The students are evaluating the extent to which the instructor communicates his/her enthusiasm, interest or liking for the subject.
5.5	<i>Enthusiasm for Teaching:</i> The students are evaluating the extent to which the instructor communicates his/her enthusiasm, interest or liking for teaching.
5.6	<i>Friendly Classroom Climate:</i> The students are evaluating the extent to which the instructor modelled, encouraged and achieved a friendly and safe classroom.
5.7	<i>Management Style:</i> The students are evaluating the instructor's management style (e.g., authoritarian/participatory, formal/informal) and method of handling issues of classroom control (e.g., noise, order, seating, calling on students).
5.8	<i>Time Management:</i> The students are evaluating the extent to which the instructor handled class time.
5.9	<i>Motivating Students to Greater Effort:</i> The students are evaluating the extent to which



	the instructor motivated students to more effort, intellectual curiosity, love of learning, high academic aspirations, etc.
5.10	<i>Respect for Others</i> : The students are evaluating the extent to which the instructor modelled, encouraged and showed trust, respect, and consideration for others (e.g., listened without interruption, did not belittle or criticize others' criticism, treated others as equals, was punctual, etc.).
5.11	<i>Tolerance of Diversity</i> : The students are evaluating the extent to which the instructor modelled, encouraged and achieved tolerance for a diversity of opinions, ideas and viewpoints and an absence of prejudice in the classroom.
5.12	<i>Overall: Learning Climate</i> : The students are evaluating the overall extent to which the learning environment fostered their active engagement as valued members of a learning community.
6	Assessment
6.1	Assessment Tasks (labelled simply as 'Assessment' by Abrami et al.): The students are evaluating the extent to which the instructor's tests were appropriate in terms of content, frequency, time allocation, weight, difficulty, validity and learning opportunity. They are also evaluating the instructor's fairness and consistency in grading.
6.2	Feedback: The students are evaluating the instructor's use of review and feedback (frequency, positive/negative) and its effect on students.
6.3	Clear Criteria and Standards (labelled as 'Objectives' by Abrami et al.): The students are evaluating the extent to which the instructor communicated performance criteria and deadlines for assignments and tests.
6.4	Overall Assessment: The students are evaluating the overall extent to which the assessment supported their learning.

4. OVERVIEW OF STUDENT EVALUATION OF TEACHING SURVEY USE IN AUSTRALIAN UNIVERSITIES

Profile of respondents

All Australian universities were contacted through the Office of the DVC / PVC responsible for teaching and learning and invited to participate in the study and to nominate an individual to provide information on SET use in their institution. Thirty four universities provided contact details indicating their willingness to participate in this particular study. The thirty four designated individuals were then emailed and invited to complete the two online surveys. A follow-up email was sent to non-respondents four weeks later.

Of the thirty four individuals contacted, twenty seven responded on behalf of their university in time to be included in the data analyses (June 2007) and a further two institutions responded subsequently, and their responses have been included in the collated data set. Of the twenty nine respondents, twenty four respondents completed both survey one and survey two. Five universities took advantage of the invitation to complete survey one and provide the researchers with their survey items for entry by the research team in survey two. One university declined to complete either survey on the basis of insufficient time, and no further response was received from the remaining four institutions.

Of the twenty nine universities who contributed data, seven were Group of Eight universities; five were Australian Technology Network universities; five were Innovative Research universities; six were New Generation universities; and six were classified as non-aligned universities. A list of the university classifications used in the study is included in the appendices (Appendix 1) to this report.

It should be noted that SET practices in many Australian universities were undergoing change and development at the time of the study and the data reported by institutions in 2007 reflected current practice at the time of completing the surveys.

Overview of current survey practice in Australian universities

This section considers the general trends and patterns in relation to how student evaluation of teaching survey data is aggregated & used at different organisational levels of Australian universities and identifies issues to consider in Stage Two of the study.

The open response data gathered through the first online survey is summarized below for each level of surveying (Individual, Department/School/Faculty, Institutional and Across Sector – i.e. groups of institutions). The data is summarised in relation to the key issues identified in the survey trigger prompts.

1. The purpose and use of SET data at each level of practice
2. The way that SET data is interpreted at each level of practice.



The analysis identifies how survey practices vary in relation to the most commonly identified features in the responses. Figures in parentheses in the report represent the number of instances of practice reported in the qualitative comments.

In general the respondents provided significantly more detail in relation to the survey practices at the level of individuals and departments than at the institutional and sector level. This may reflect actual practices at these levels or it may reflect a lack of detail or transparency about institutions practices at these levels.

Accessing the collated survey responses

Institutional contacts were asked to review the data provided for their university and then asked to give approval for their de-identified survey responses to be included in the collated data set of responses. Where approval was granted the data has been included and is available as part of a dataset of items, scales and descriptions of institutional practice available at <http://www.itl.usyd.edu.au/projects/studentsurveys.htm>

Level 1: Individual

Purpose and use of the data?

Type of survey.

Most universities had a range of surveys for different types of teachers and some even had specific surveys for award applicants. These surveys were occasionally fully standardised (2) but typically included core and optional items. A small number of universities (2) allowed individuals to completely customise their surveys at this level.

Survey practices.

Evaluations of teaching practice were generally voluntary and initiated by the individual staff member. In four institutions surveys were initiated by a third party (HOS, Dean, course coordinator). The frequency of unit/ teaching practice evaluations is generally mandated by faculty and/or university practice, and closely linked to surveys of units of study.

Use of data.

Data is primarily used for individual improvement and to inform teaching practice (15). There was also a strong focus on the use of data as evidence for promotion (13), extension of contract and tenure (3), teaching award applications (7), and individual staff performance management and development (11). Data is typically confidential but some universities (7) also provide access to Deans, HOS, and other subject related staff. Some universities noted the expectation that individual staff members were responsible for closing the loop by reporting back to students and to HOS on their actions in response to feedback.

How is the data interpreted?

Aggregation.

Five institutions did not perform any aggregation, although one reported that aggregation was undertaken if there were multiple orders for the same staff member. Aggregation of results from four institutions was typically at the school level, with some aggregation across teaching type, faculty and entire university.

Basis for comparisons.

(standards/benchmarking/averages etc). In general, staff compared their own previous results, either with teaching in previous semesters (2), or with aggregated school or faculty reports (4). Three institutions either did not perform comparisons, or reported little formal use of comparative data at individual level. Where comparisons occurred it was typically based on scale scores, percentage agreement, means, standard deviation and frequency distribution. Standards were most commonly based on percentage agreement scores (e.g. 30% disagreement was the equivalent to unsatisfactory performance, or typified an area of concern). In five institutions, these were linked to performance development management and review practices.

Access to comparison results.

Results were confidential to the individual teacher in seven institutions, though one reported that comparative results by school and faculty were available with permission of the individual teacher. Access to results by individual and their line manager and above (Heads of School, Dean, academic supervisor) was permissible in eight institutions. Public access to the comparative results, either through public website or secure log on, was possible in seven institutions.

Level 2: Department/School/Faculty

Purpose and use of the data?

Type of survey.

Types of surveys reported included: unit of study (4); degree experience (3); first year experience (2), and a range of different surveys targeting different areas of the student experience (2). Surveys were both standardised (3), and customised for individual faculties or a combination of both (3), and were typically administered approximately every three years.

Survey practices.

Evaluations were received and ordered at a number of levels within the institutions: Course Coordinator, Department, School, but most commonly Faculty (7). Individual staff members were allowed access to all other individual results in one institution; though usually this was not the case.

Use of data.

In the majority of institutions who responded, this data was used primarily for strategic quality improvement and assurance processes (15). Specific uses included: internal course and program reviews (8); informing strategic planning processes (3); and in benchmarking (1). Eight institutions used the data to inform curriculum design and review processes; and three used it for accreditation purposes. Whilst use of SET data for promotion applications was reported at the individual level, this was not mentioned at the departmental/ school/ faculty. One university reported that the university wide rating system that had been developed through the use of SET data, had helped create more dialogue about teaching at both the school and the faculty level. The expectation that Deans were responsible for closing the loop by reporting back to students on faculty actions in response to feedback was noted by one institution.

How is the data interpreted?

Aggregation.

The majority of respondents aggregated data at school, and faculty level (8); whilst three of these also aggregated at university level, and one aggregated by teaching type. One university did not aggregate data but provided a course summary report which listed scores for each unit listed.

Basis for comparisons (standards/benchmarking/averages etc).

Comparisons were generally made at course level with cognate schools and faculties (6). Where external comparisons were made, it was on the basis of CEQ items adapted for use with current students (3). In these cases, comparison was against national field of study figures, or compared with other faculties in the university on the basis of adjustments using national field of study data to control for field of study effects. The basis for comparison is typically percentage agreement/ disagreement results, although one institution has instigated an electronic course performance reporting system which uses an annual electronic score card which compares results for courses across semesters and previous years. Some degree of benchmarking exists, with levels and averages set at university level (4).

Access to comparison results.

Typically, results from all unit of study evaluations within their faculty and/ or department were made available to course coordinators, Heads of School, and Deans (13). Data from course level comparisons was typically made available to all members of the university. Data was generally made available through public websites (8), although one institution restricted access through secure log on.

Level 3: Institution

Purpose and use of the data?

Type of survey.

At the institutional level, surveys were typically of three main types: annual national surveys on degree experience (CEQ) (8); internal surveys on degree experience (11); or surveys targeting different areas of the student experience e.g. library, student services, WebCT (9). One institution has combined its own survey with the CEQ as one graduate survey. Another is in the process of developing a common survey for implementation in Semester 2 2007. The frequency of internal institutional surveys is variable, ranging from a one to three year cycle.

Survey practices.

Little information was provided regarding whole institutional survey practice. In most cases, this was devolved to faculty level. Three institutions noted that results were considered by Academic Boards, University Councils, Learning and Teaching Committees, and PVCs (Learning and Teaching) or equivalent. One university provided CEQ results to the DVC Academic, highlighting areas of good and bad performance.

Use of data.

At an institutional level, the use of data from SETs falls into four main categories: evaluation, strategic planning and performance measurement (5); performance based funding (7); internal and external quality audits and accreditation (8); and course reviews (5). Comparison across faculties, field of education, and other institutions was also reported (6). In two institutions results are used to justify promotion and are considered in applications for Faculty and University teaching awards. Two universities have data warehouses which are used to inform and support strategic learning and teaching planning. Two universities use their own institutional data for research on survey practices and to improve the efficiency and effectiveness of their own practice. Institutional uses of SET data for benchmarking against the rest (or some) of the sector are emerging. Two universities reported that they were currently negotiating benchmarking projects; one was exploring learning and teaching benchmarking but this was not being undertaken systematically; and three universities undertook benchmarking with similar institutions.

How is the data interpreted?

Aggregation.

Aggregation generally takes place at faculty level (4), with some aggregation of specific items across the university, or at a level meaningful to research questions (e.g. First Year Experience)

Basis for comparisons (standards/benchmarking/averages etc).

Comparisons were, for the most part, made over time (9). Some comparison was made with CEQ data (2); or with institutionally based standards and performance targets (1). Three universities explicitly noted they did not use institution wide standards, due to field of study differences. Data was commonly benchmarked against national CEQ averages, or in the case of internal adaptation of the CEQ with adjusted national CEQ figures, or with selected other institutions (3).

Access to comparison results.

Five institutions made degree level comparisons publicly accessible on their websites. Other levels of accessibility included: senior management (6), internal and external review teams (2), and university councils (1). Two universities provided all staff with trend data on the institution's CEQ.

Level 4: Across the Sector

Purpose and use of the data?

Type of survey.

The main survey used at sector level was the GDS/ CEQ (9). Individual universities also participated in other national surveys - for example: First Year Experience and Attrition; NSSE; and the ATN online learning survey. One institution indicated its intent to take part in the University and Beyond Survey being conducted by GCA.

Survey practices.

The survey practices for the main survey – the AGS (CEQ & GDS) were noted to be becoming more standardised. The responsibility for these processes was noted to be different to many other SET processes.

Use of data.

The main sectoral use of the SET CEQ data was in the National Learning and Teaching Fund (8). One university noted plans for comparisons by a group of universities based on accumulated SET data in a particular discipline. Other sectoral SET data was used for specific cross-sectoral research and development, e.g. the First year experience data survey.

How is the data interpreted?

Basis for comparisons (standards/benchmarking/averages etc).

Most universities used the CEQ results as a basis for comparisons with other institutions (national competitors, Go8, ATN, other research focused universities etc), across field of study, or for performance monitoring (12) – see above - however there was little evidence of sectoral benchmarking activities using SET apart from the LTPF.

Access to comparison results.

Information on access to comparative results was sparse. While some universities reported various levels of comparison data between their own institution and the rest of the sector there was little evidence of collective uses of the data by groups within the sector except where this was for the purposes of research – for instance the collective use of cross sector data from the first year experience survey

Overview of dimensions of teaching probed in relation to organisational level

Responses to Survey 2 (coding of teaching evaluation items against the provided dimensions of teaching evaluation) were checked by the researchers for accuracy of coding. Where the initial coding was provided by the respondent the checking was by one of the research team. Where the coding was by one of the research team an independent check was made of this by another member of the team.

There was generally a high level of agreement; more than 85% for respondent coding and more than 95% for researcher coding. Where there was disagreement with the initial coding the coding was based on consensus reached through discussion amongst the research team.

Inappropriate coding typically reflected coding against a less representative dimension but in an appropriate level, or against a representative dimension but in an inappropriate level (e.g. subject level items included in a teacher level category). The complexity of the coding scheme – 42 dimensions, plus the potential to describe extra dimensions falling outside this taxonomy – meant that items were sometimes coded against earlier occurring dimensions in the taxonomy when they would more properly have been coded in later dimensions. This reflects an understandable lack of familiarity with the coding structure on the part of respondents.

There were some omissions noted in the checking process in relation to the omission by respondents of information relating to their university's use of CEQ scales and items at the course/degree and university levels. The use of CEQ scales by different universities was checked against Graduate Careers Australia records to ensure accuracy, and additional information relating to these SET items included in the analysis as necessary.

Most SET surveys are either standardized or use a combination of core and optional items, with only one university giving staff complete control over item choice at any level. Where the coding was completed by the research team from a database of items provided by the responding university, only core and standard items were used.

Overall Analyses

To gain an overview of the focus of teaching evaluation efforts at different levels across the sector, initial analyses focused on the major categories under which the dimensions of teaching and learning were grouped:

1. Predispositions
2. Course preparation and organisation
3. Approaches to teaching and teaching strategies
4. Quality of learning outcomes
5. Learning climate
6. Assessment
7. (Other)

The responses to survey two were considered first under these major categories and examined to determine if there were some categories that were more consistently focused on in SET surveys at any particular levels of surveying. A summary of the number of dimensions covered by the responding universities' SET surveys at each level of surveying, for each category of dimensions is given in Table 3. For instance, if 10 universities all allocated survey items to each of the 4 dimensions under 'Assessment' category at the 'Subject' level, the tally for that cell in the table would be 40.

Table 3: Tally of teaching dimensions in each category covered in SET surveys over the total number of dimensions reported at each level.

	Teacher Predispositions	Course preparation & organisation	Approaches to teaching & teaching strategies	Quality of learning outcomes	Learning climate	Assessment
	Dimensions 1.1 – 1.8	Dimensions 2.1 – 2.5	Dimensions 3.1 – 3.9	Dimensions 4.1 – 4.4	Dimensions 5.1 – 5.12	Dimensions 6.1 – 6.4
Individual teachers	30/245 or 12%	16/245 or 6.5%	83/245 or 34%	21/245 or 9%	60/245 or 24.5%	35/245 or 14%
Individual subjects	8/196 or 4%	59/196 or 30%	24/196 or 12%	33/196 or 17%	15/196 or 8%	57/196 or 29%
Courses/ Degrees	10/94 or 10% (10/373) (or 3%)	15/94 or 16% (58/373) (or 16%)	12/94 or 13% (99/373) (or 26.5%)	24/94 or 25.5% (57/373) (or 15%)	12/94 or 13% (81/373) (or 22%)	21/94 or 22% (68/373) (or 18%)
University experience	1/65 or 1.5% (1/126) (or <1%)	9/65 or 14% (9/126) (or 7%)	13/65 or 20% (23/126) (or 18%)	13/65 or 20% (30/126) (or 24%)	22/65 or 34% (56/126) (or 44%)	7/65 or 11% (7/126) (or 5%)

Figures in brackets in the final two rows of the table represent analysis with the added CEQ data from GCCA. As discussed in section 3.3, inspection of responses to survey 2 at the "Courses/Degrees" level and the "University" level found some institutions had not



included CEQ items in their responses, despite known use of the CEQ and in some cases an internal survey being used at these level of analysis. We used Table 3 in Graduate Careers Australia's 2005 Report of the Course Experience Questionnaire to identify usage of different dimensions at the Courses/Degrees and University levels, and appended these results to those provided by respondents. Results based on this recoding are given in brackets in Table 3, under the results based on original responses.

Inspection of Table 3 suggests a general lack of focus on whole university experience in SET which is notable given the contribution to quality teaching and learning that many institutions ascribe to their students' 'on-campus' experience. It is also a notable absence given the central role that whole university services such as libraries play in creating a supportive learning experience. The absence is also interesting in light of the part that independent study plays in quality learning and the contribution of 'out of class' experiences to the development of generic skills.

Across levels of analysis, several of the categories of dimensions appear to play a greater role in the sector than others. At the level of individual teachers there is a particular focus on dimensions included in the categories of 'Learning Climate', 'Teachers approaches to teaching and teaching strategies' and 'Assessment'. Interestingly, the emphasis on approaches to teaching is also present at the level of Course/Degree but the focus is absent at the subject level. The focus on assessment is maintained across SET instruments used for individuals, subjects and courses, perhaps indicating the centrality of assessment in any consideration of teaching and the idea that assessment is a microcosm of many aspects of the teaching and learning experience. The absence of assessment as a category in SET instruments used at the level of the whole university experience perhaps reflects the traditional view of assessment as being restricted to in-class learning; and while the idea of student portfolios of achievement representing a form of assessment that covers their broader university experience is one that is gathering strength, this is not reflected in such instruments. At the level of subjects the focus shifts to 'Course Preparation and Organisation', 'Quality of Learning Outcomes' and 'Assessment', a focus that is consistent with traditional views of curriculum. Interestingly 'Learning Climate' does not figure strongly in surveys at this level, which would be expected if these surveys fully reflected student focused perspectives on teaching and ideas around the importance of student engagement emerging in the more recent teaching and learning literature.

At the degree/course level, SET instruments focus on the full range of categories of dimensions. The importance of assessment items remains and, as with all levels, the focus on teacher predispositions is limited. The limited attention to SET items relating to teacher predispositions in the Australian context perhaps reflects the somewhat different theoretical perspective that informs university teaching in Australia to the North American research on which the dimensions were based. The dominance of the student learning perspective shifts attention from teacher input factors to students' experience of educationally important aspects of teaching. As such, this shifts attention away from the Northern American focus on who the teacher is and what s/he does. The focus on assessment at this level is interesting in that very little undergraduate assessment is focused at the course level. Instead assessment typically occurs in Australian higher education at the level of individual subjects that comprise a degree. The prevalence of



SET items relating to assessment at this level might again reflect the idea that assessment is a crucial aspect of any consideration of quality teaching and learning. Alternatively it might reflect the UK origins of many Australian surveys used at this level given that whole degree assessment is more of a feature of the UK system. The widespread use of both original and extended CEQ scales and items by responding universities plays a large role in the foci on 'Course Preparation and Organisation', 'Approaches to teaching and teaching strategies', and 'Quality of learning outcomes'.

SET items (and indeed SET instruments) probing the whole university experience were relatively rare across the institutions surveyed. Where items were present they tended to focus on Learning Climate and the quality of 'Learning Outcomes'. This perhaps reflects the idea that many of the higher-level generic attribute outcomes of higher education are the product of the whole university experience, not just formal classes. The focus on 'Learning Climate' dimensions reflects the focus of these dimensions on the quality of student-teacher interactions and the idea that these interactions go beyond the classroom. Given the importance of learning climate, in any consideration of students' experiences of teaching, at greater emphasis on learning climate might perhaps have been expected at other levels, especially the subject level.

Additional dimensions

Respondents were also asked to identify any items that were not able to be coded using the dimensions provided from Abrami et al's review of the multi-section validity studies and suggest additional dimensions these might represent.

In some cases items suggested did reflect existing dimensions and where this occurred the items were recoded under the existing dimensions and included in the preceding analyses. However other items in use in Australian SET surveys did not reflect the dimensions derived from the previous research.

The additional items in common use suggested two significant additional categories of dimensions.

The first of these categories relates to Learning Environment. Dimensions in this category reflected aspects relating to:

1. Facilities (classrooms, laboratories etc), which constituted the learning environment,
2. Information technology enabled learning environments,
3. Library and other learning resources present in the environment
4. Student & learning support services which create an effective learning environment
5. Administrative services which create an effective learning environment

Examples of these items at various levels are included in appendix 2 to this report. The second category relates to some aspects of **Student Motivation and Effort**. The following specific Student **Motivation and Effort** dimensions were suggested by the responses:

1. Student Motivation

2. Student Effort
3. Active Student Learning

The results also suggest that the dimensions within the existing Learning Climate category might benefit from broadening to encompass aspects of student motivation (rather than simply teacher engagement) that relate to the quality of interactions and classroom climate.

The majority of additional items and dimensions suggested were in the Learning Environment category with relatively few items suggested which represented Student Engagement dimensions. Across SET levels, the most common additional dimension related broadly to the quality of 'Information technology enabled learning environments' and 'Library and learning support services'. However, the increasing prominence of ideas about student engagement, both in the research literature and the recent development and national piloting of an Australian adaptation of the National Survey of Student Engagement lends additional importance to the indication that there are relatively few examples of student engagement SET items currently in use. Examples of the additional items are included at the end of appendix 2.

Analyses of Dimensions of Teaching Evaluation

The above analyses give a broad overview of the focus of teaching evaluation at different levels across the sector. However, closer inspection of responses showed considerable variation in the frequency of use of specific dimensions within the broad categories. The following analyses explore this variation with a particular focus on where overlap occurs in terms of dimensions commonly probed in SET instruments at different institutions.

The graph in Figure 1 depicts the frequency of presence of specific dimensions at different levels across the responding universities.

To assist in discerning the pattern of teaching dimension representation in SET instruments at different institutional levels (Individuals, Subjects, Degrees and Whole university), the same data is also presented separately for each level in the following four graphs.

Figure 1: Overall frequency of use of teaching evaluation dimensions

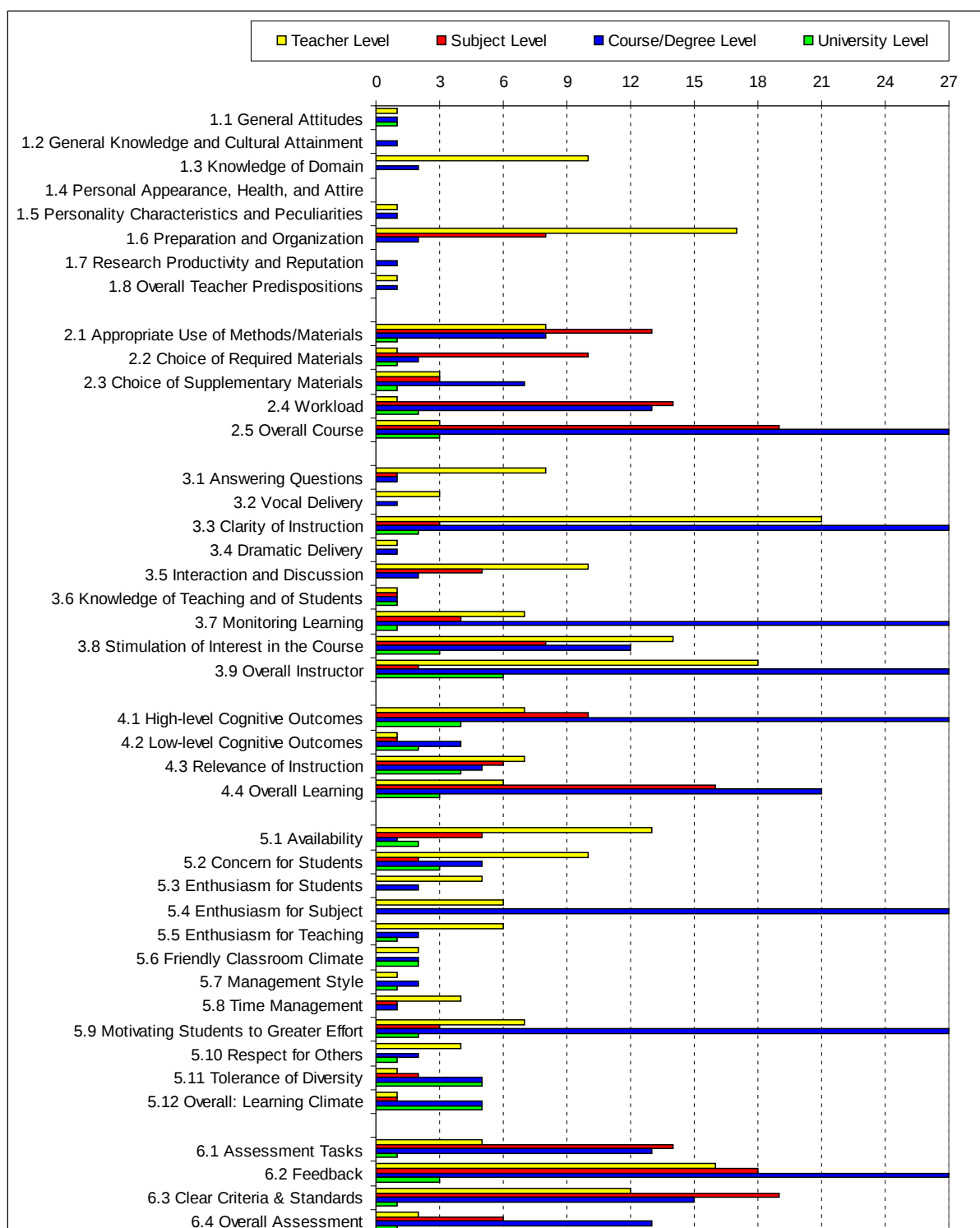


Figure 1.1: Individual teacher level: frequency of use of teaching evaluation dimensions

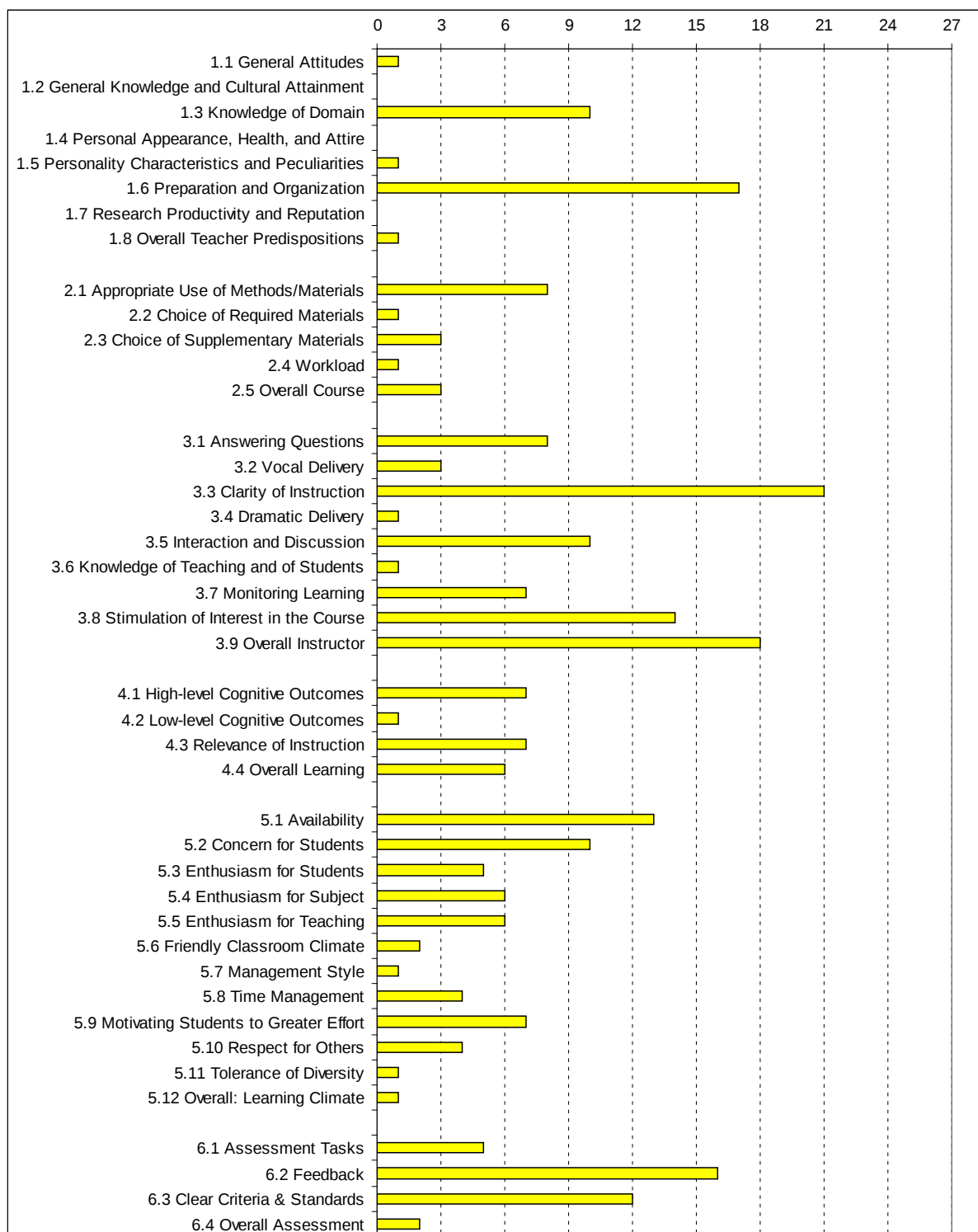


Figure 1.2: Subject level: frequency of use of teaching evaluation dimensions

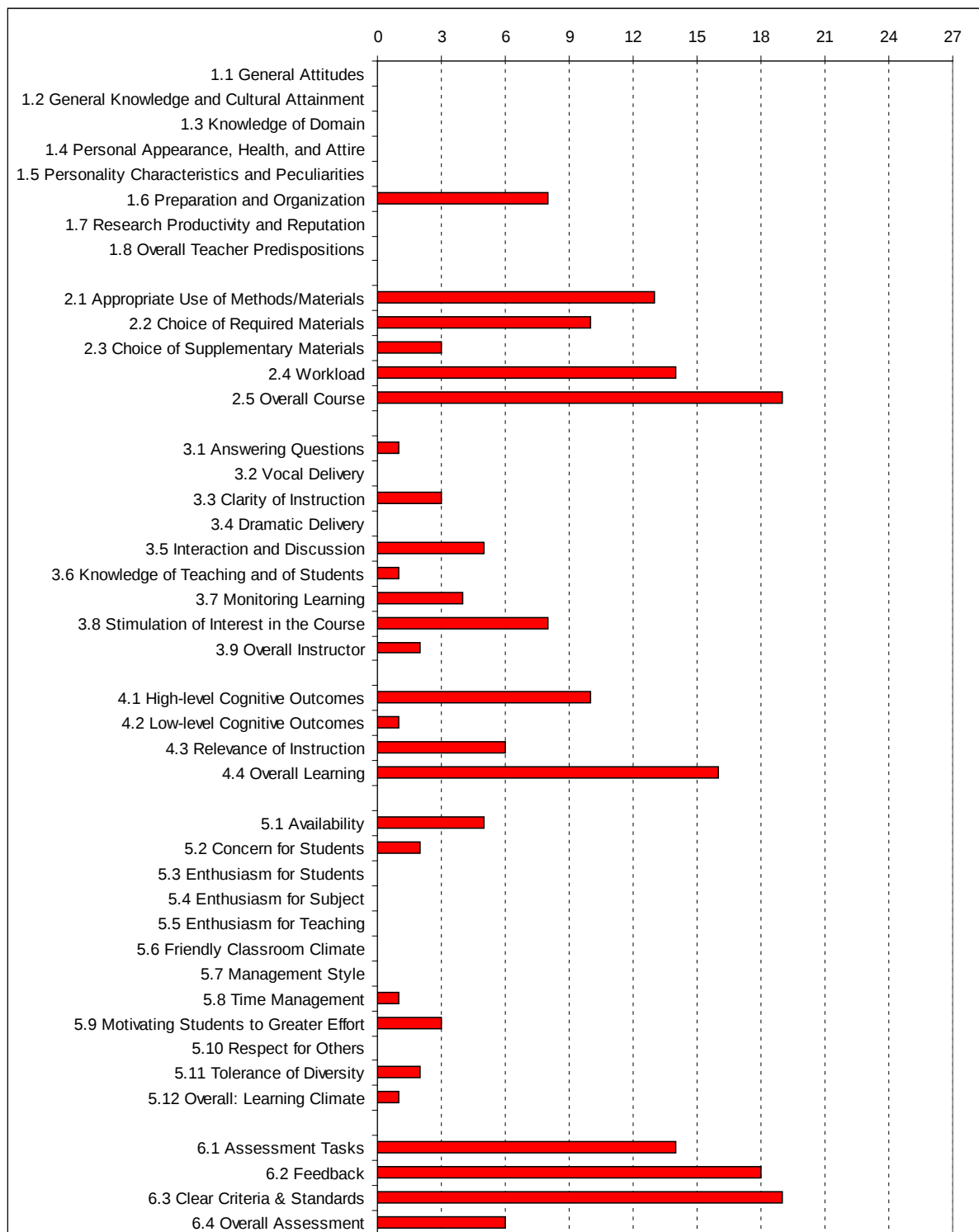


Figure 1.3: Course/degree level: frequency of use of teaching evaluation dimensions

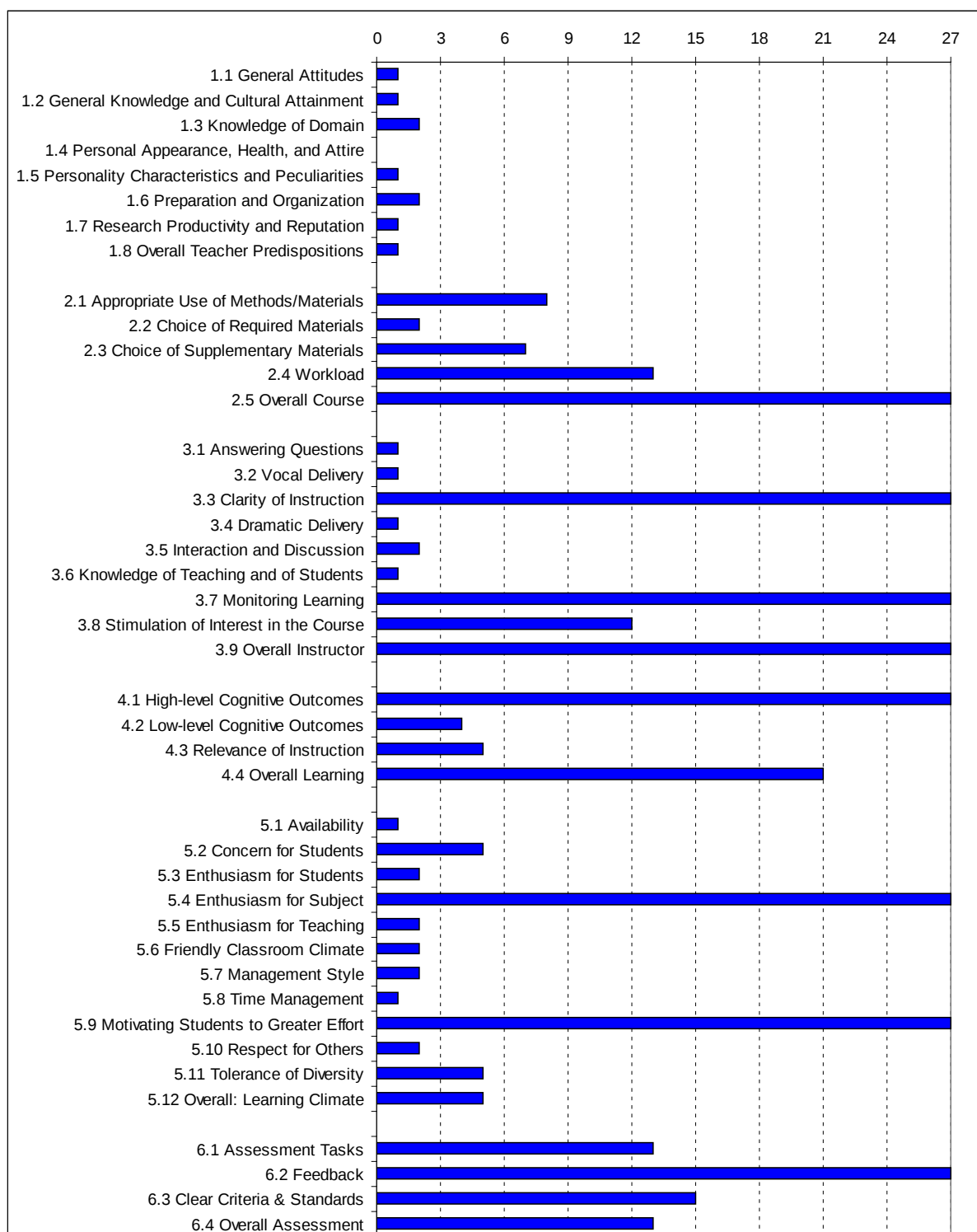
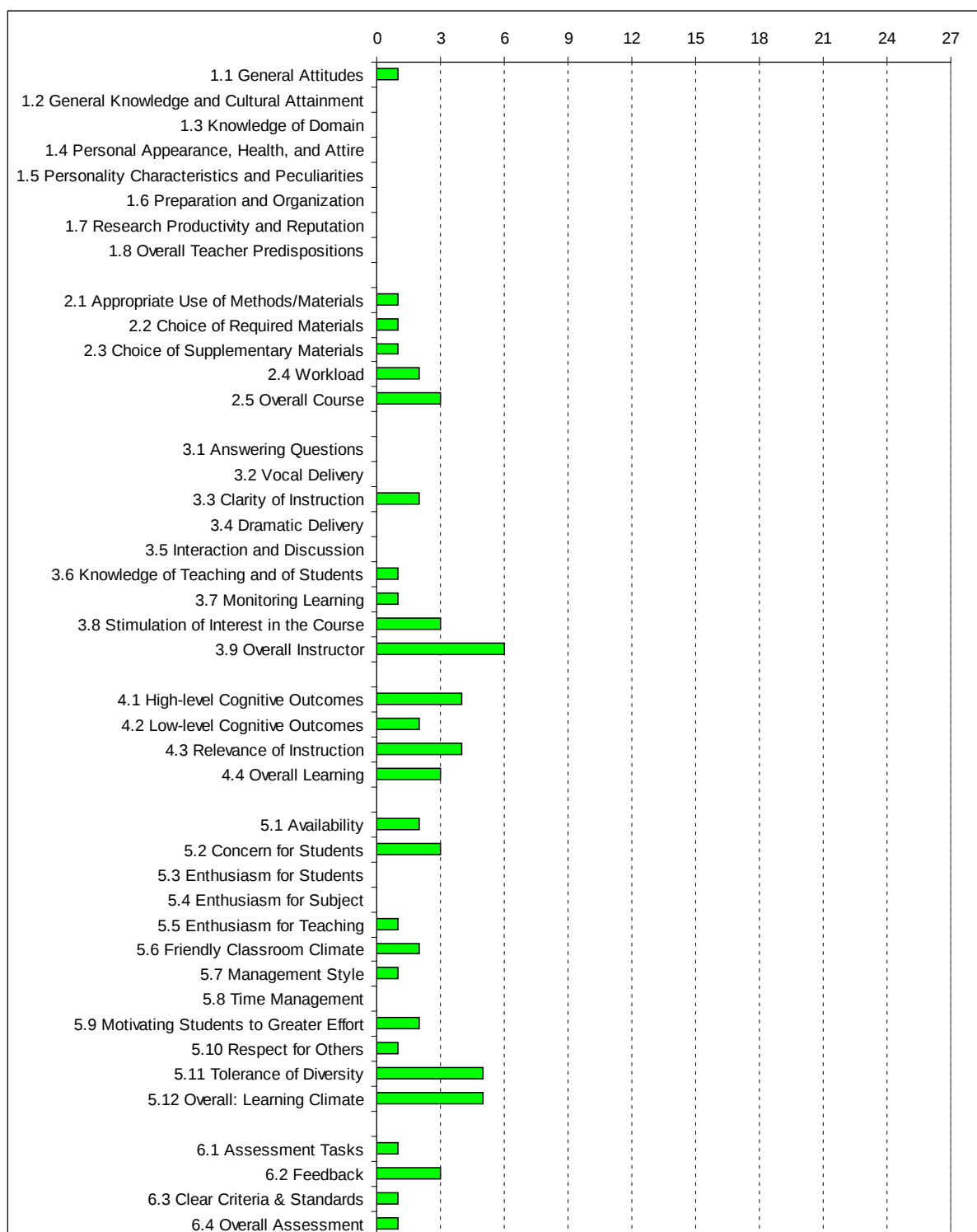


Figure 1.4: Whole university level: frequency of use of teaching evaluation dimensions



To aid interpretation the frequency of use of dimensions results have been reformulated in the next table to show the relative importance of a given dimension at a given level. The number of responses per dimension were re-categorised into 7 bands: 0-4 responses, 5-8 responses, 9-12 responses, 13-16 responses, 17-20 responses, 21-24 responses, and 25-27 responses. Shading of cells is used below to represent relative extent of common usage. The darker the shading the greater the level of use.

Table 4: Frequency of use of dimensions reformulated to show relative importance as seven bands (see previous page)

0-4 responses	5-8 responses	9-12 responses	13-16 responses	17-20 responses	21-24 responses	25-27 responses
Dimension						
Teacher Level						
Subject Level						
Course/Degree Level						
University level						
1. Teacher Predispositions						
1.1 General Attitudes	1	0	1	1		
1.2 General Knowledge and Cultural Attainment	0	0	1	0		
1.3 Knowledge of Domain	10	0	2	0		
1.4 Personal Appearance, Health, and Attire	0	0	0	0		
1.5 Personality Characteristics and Peculiarities	1	0	1	0		
1.6 Preparation and Organization	17	8	2	0		
1.7 Research Productivity and Reputation	0	0	2	0		
1.8 Overall Teacher Predispositions	1	0	1	0		
2. Course Preparation And Organisation						
2.1 Appropriate Use of Methods/Materials	8	13	9	1		
2.2 Choice of Required Materials	1	10	2	1		
2.3 Choice of Supplementary Materials	3	3	7	1		
2.4 Workload	1	14	13	2		
2.5 Overall Course	3	19	27	3		
3. Approaches to teaching and teaching strategies						
3.1 Answering Questions	8	1	1	0		
3.2 Vocal Delivery	3	0	1	0		
3.3 Clarity of Instruction	21	3	27	2		
3.4 Dramatic Delivery	1	0	1	0		
3.5 Interaction and Discussion	10	5	2	0		
3.6 Knowledge of Teaching and of Students	1	1	1	1		
3.7 Monitoring Learning	7	4	27	1		
3.8 Stimulation of Interest in the Course	14	8	12	3		
3.9 Overall Instructor	18	2	27	6		
4. Quality of Learning Outcomes						
4.1 High-level Cognitive Outcomes	7	10	27	4		



4.2 Low-level Cognitive Outcomes	1	1	4	2
4.3 Relevance of Instruction	7	6	5	5
4.4 Overall Learning	6	16	21	3
5. Learning Climate				
5.1 Availability	13	5	1	2
5.2 Concern for Students	10	2	5	3
5.3 Enthusiasm for Students	5	0	2	0
5.4 Enthusiasm for Subject	6	0	27	0
5.5 Enthusiasm for Teaching	6	0	2	1
5.6 Friendly Classroom Climate	2	0	2	2
5.7 Management Style	1	0	2	1
5.8 Time Management	4	1	1	0
5.9 Motivating Students to Greater Effort	7	3	27	2
5.10 Respect for Others	4	0	2	1
5.11 Tolerance of Diversity	1	2	5	5
5.12 Overall: Learning Climate	1	1	5	5
6. Assessment				
6.1 Assessment tasks	5	14	13	1
6.2 Feedback	16	18	27	3
6.3 Clear criteria & standards	12	19	15	1
6.4 Overall Assessment	2	6	13	1

Interpretation

While there are patterns of common focus on particular categories and to a lesser extent the dimensions within categories, which are shared across some institutions; there is not a high degree of commonality. The most commonly used dimensions in each category are listed in the following tables.

Individual teacher level

At the individual teacher level the most commonly used SET dimensions in each category are:

Category	Most commonly used SET dimensions in the category
Predispositions:	Personal preparation and organisation Knowledge of domain
Course preparation and organisation	Appropriate use of materials and methods
Approaches to teaching and teaching strategies	Clarity of instruction Overall instructor
Quality of learning outcomes	Relevance of instruction High Level outcomes
Learning climate	Availability Concern for students
Assessment	Feedback Objectives



Subject level

At the Subject level the most commonly used SET dimensions in each category are:

Category	Most commonly used SET dimensions in the category
Predispositions:	Personal preparation and organisation
Course preparation and organisation	Overall course Workload Appropriate use of materials and methods
Approaches to teaching and teaching strategies	Stimulation of interest Interaction and discussion Monitoring learning
Quality of learning outcomes	Overall learning High Level outcomes
Learning climate	Availability
Assessment	Objectives Feedback

Course level

At the Course level the most commonly used SET dimensions in each category are:

Category	Most commonly used SET dimensions in the category
Predispositions:	Personal preparation and organisation Knowledge of domain
Course preparation and organisation	Overall course
Approaches to teaching and teaching strategies	Clarity of instruction Monitoring learning Overall instructor
Quality of learning outcomes	High level outcomes Overall learning
Learning climate	Enthusiasm for subject Motivating students to greater effort
Assessment	Feedback

Whole University level

At the Whole University level the most commonly used SET dimensions are:

Category	Most commonly used SET dimensions in the category
Predispositions:	General attitudes
Course preparation and organisation	Overall course
Approaches to teaching and teaching strategies	Overall instructor
Quality of learning outcomes	High level outcomes Relevance Overall learning
Learning climate	Overall learning Tolerance of diversity
Assessment	Feedback

Usage across types of universities

Where over two-thirds of respondents used a given dimension at a given level of analysis (e.g. Degree/Course), cross-tabulations (using exact tests) were used to analyse whether the use of the most widely used dimensions varied across types of university (Group of Eight, Australian Technology Network, Innovative Research New Generation universities, and non-aligned universities).

- At the Teacher level, there was no association between type of university and use of “Clarity of Instruction” ratings, or use of “Overall Instructor” ratings.
- At the Subject level, there was no association between type of university and use of “Overall Course” ratings, use of feedback ratings, or use of subject objectives ratings.
- At the Course/Degree level, there was no association between type of university and use of “Overall Learning” ratings. (Analyses related to dimensions captured by the mandatory CEQ scales were not performed as the use of these dimensions are outside of universities’ control.)

These analyses suggest that for the most widespread dimensions, usage is largely uniform across the sector.

The following initial **general conclusions** can be drawn from the results:

- Across all levels of analysis, there are relatively few dimensions measured by the majority of responding universities.
- The Teacher, Subject and Degree/Course levels have considerably more measured dimensions than the Whole University level.
- At the Teacher level, the most common dimension across responding universities was Clarity of Instruction, followed by Overall Instructor ratings. A medium number of responding universities included ratings of Preparation and Organisation, Feedback, Stimulation of Interest in the Course, Availability, Objectives, Concern for Students, and Interaction and Discussion.



- At the Subject level, the most common dimension across responding universities were Overall Course ratings, Objectives, and Feedback. A medium number of responding universities included ratings of Overall Learning, Workload, Assessment, Appropriate Use of Methods/Materials, High-level Cognitive Outcomes, and Choice of Required Materials.
- At the Course/Degree level, because of the design of the Course Experience Questionnaire, all responding universities collected feedback on the Overall Course; Clarity of Instruction; Monitoring Learning; Overall Instructors (using the Good Teaching Scale as a global measure); Motivating Students to Greater Effort; and Higher Order Cognitive Outcomes. Over three-quarters of universities also collected feedback on students' Overall Learning.
- At the University level, there are presently few dimensions which are measured, and of those that are, only a few universities do so.

5. ISSUES AND IMPLICATIONS ARISING FROM CURRENT SURVEY PRACTICE

The findings of the first stage of this study suggested that there is considerable variation in individual universities' current SET practices. This variation does not reflect institutional type or variations in institutional SET policy. It appears more likely that the variations reflect the idiosyncratic and largely atheoretical historical basis for the development of many university SET systems. As universities' internal SET systems take on increasingly important roles within the institution, and particularly as internal systems begin to be linked to the more formal external SET systems used for national quality assurance, these historical features are changing.

There is considerable work currently being done within Australian universities in relation to psychometric validation of existing SET tools. However there appears to be less attention being paid to the theoretical basis for these tools and a consideration of how the tools relate to both the established theoretically sound models of teaching and learning and the new ideas emerging from the research on university teaching. In the absence of such a consideration many SET surveys remain based on the survey writers' implicit and sometimes unrecognized beliefs about what is good teaching, and hence what is important to measure. The framework developed by Abrami et al. and applied in the present study does not pretend to be a substitute for an examination of the theoretical basis of SET surveys, however it does provide an additional perspective from which to approach this reflection in the following ways:

Commonalities across SET surveys

1. Using the framework to reflect on the dimensions of teaching and learning covered by their existing SET surveys will help individual institutions to identify gaps in current practice. There may be important dimensions of teaching and learning that are absent from their existing quality assurance SET surveys.
2. The comparison of an individual institution's 'profile' on the dimensions, with the pattern of use of these dimensions across the sector, provides a benchmark for considering institutional SET practice.
3. The framework identifies dimensions of teaching that may represent shared foci of attention for particular institutions, and as such may provide the basis for universities selecting collaborative benchmarking partners.
4. Of the four levels of SET use, the greatest level of institutional overlap in coverage of the different dimensions is at the course level. This is partially because of the reported use by institutions of the national CEQ survey, but it is also the result of the adaptation of the CEQ for use as a current student survey by many institutions. Not surprisingly therefore, of all the SET items, the CEQ, and CEQ derived, scales and items provide the greatest commonality in terms of shared



dimensions and common scales and items between universities. This commonality is likely to increase as several other universities indicated they were also considering developing versions of the CEQ for use with current students. If the 2007 pilot of the AUSSE leads to institutions using this instrument in the future then the current commonality of CEQ data is likely to be replicated in AUSSE data. There is no equivalent across sector commonality in SET instruments at the levels of subject or individual teacher.

5. The CEQ and universities' adaptations of the CEQ for current student cohorts are primarily used at Course level, with some items providing information relevant to the whole university experience. However, universities do not rely on SET systems that operate at only one level.
6. Different dimensions of teaching and learning were prominent in SET surveys at different levels (see Table 4). These may provide grounds for developing common items based on this demonstrated shared focus at each level.
7. There is rarely an interconnection (theoretical or psychometric) between the levels of surveying in Australian universities. Current survey practice suggests that universities will continue to require surveys at different levels reflecting the different foci and different uses of data. The lack of internal coherence within SET systems suggests that SET items and data collected at one institutional level are likely to be largely independent of data collected at any other level. In the absence of this internal coherence, establishing a commonality at one level (for example 'course surveys') will not lead to commonalities at other levels and all levels of SET are likely to continue to need to be maintained. This may not be particularly efficient.
8. The areas of greatest overlap in SET items vary between levels of survey use. Across the range of levels, the 'Overall' SET items in the various categories often showed a reasonable amount of commonality across institutions and there may be potential for this commonality to be improved.
9. With the exception of the CEQ coverage described above, there are currently no SET items common to a sufficiently large number of institutions' SET surveys, which would provide a basis for sharing or comparison of data.
10. The range of SET items used in SET surveys appears at first glance to be highly diverse yet there is a degree of commonality in the dimensions of teaching and learning probed by the vast array of surveys and items. It would be possible for universities to approach benchmarking as a conversation about a shared teaching dimension – based on data about that dimension collected using different survey items. However this is a different approach to benchmarking which uses the 'comparison of scores' as a basis for a conversation.
11. There are relatively few SET items and systems that target the whole university experience. The emphasis remains on teaching in formal courses despite growing

awareness of the importance of students' engagement in the overall experience of higher education.

12. The Abrami et al. framework has proved useful in the context of analysing SET tools. It becomes more useful with the addition of a dimension described tentatively as 'Learning Environment' and the change in emphasis to the 'Learning Climate' dimensions to reflect student engagement rather than just teacher engagement. These additions suggest possible areas for further development in Australian universities' SET systems.

Commonalities across SET practices

1. There is considerable variation in how SET data is gathered, interpreted, used and acted up in Australian universities.
2. If the potential for sharing of information drawn from institutions' own SET systems is to be realized, then the variations between institutional (and even within institutional) uses of such data will need to be addressed as well as identifying commonalities in SET items. At present particular data is available to some people in some institutions (and parts of institutions) and not to the equivalent people in potential benchmarking universities.
3. Currently universities' own institutional uses of SET data are limited. There are very few examples of meaningful external benchmarking of SET data by universities beyond that imposed externally in the form of rankings based on CEQ data. Where this has happened using SET data it is typically not institution wide but occurring within existing discipline networks.
4. There have been recent initial efforts to establish benchmarking using SET data by some groups of universities.
5. Respondents have suggested that the internal uses of SET data are changing rapidly in Australian universities. There is increasing attention being paid by university communities to using the data to direct and monitor strategic change, rather than simply collecting the data for individuals' use in promotion and reward processes or for individual teaching improvement.
6. The internal processes used to set standards to support interpretation of SET data remain highly variable.

Overall, there is considerable variability across universities in the nature of SET scales and items used in the Australian higher education sector. In particular, there is insufficient commonality in currently used SET scales or items at the teacher and subject levels to support systematic benchmarking of core SET constructs. Some universities share an interest in some common dimensions; however, the items they use to probe these dimensions are very different. For example, at the "Teacher" level, examples of items evaluating "Clarity of Instruction" included the following:

- The lecturer explained important concepts and ideas in ways that I could understand
- The lecturer regularly summarised main points of the presentation to the class.
- The quality of this teacher's explanations was...
- During each lecture, the lecturer described what students were expected to learn from that lecture
- (name of teacher) gives clear explanations.

Another commonly evaluated dimension, “Feedback”, showed similar variability between items:

- During lectures, the lecturer provided useful feedback on student work.
- The tutor returned assessed work within a reasonable time.
- This teacher gave constructive feedback (in class, on assessment tasks, etc) that helped my learning.
- The lecturer provided timely and constructive feedback on my assessment tasks.
- The staff member put a lot of time into commenting on my work.

The most overlap was seen for the overall-type items; however, even this extent of overlap was minimal, with potential benchmarking partners framing such questions in a variety of ways. Examples of overall lecturer and overall unit of study items include the following:

- Overall my lecturer effectively supported my learning.
- Overall, I was satisfied with the performance of this lecturer.
- Overall, the quality of this staff member's teaching was...
- Overall, how would you rate the staff member's teaching in this unit?
- Overall, the lecturer was highly effective in facilitating my learning
- Overall I was satisfied with the quality of this unit of study.
- Overall, I was satisfied with this unit.
- Overall I was satisfied with the quality of this unit.
- Overall, I am satisfied with the quality of this course.
- All things considered, my overall rating of this unit is: (poor/satisfactory/excellent)

Such variability means that if institutions want to benchmark in the traditional manner by comparing SET results to identify areas to explore further,, they will first need to develop new shared items and scales. These might be based on agreeing to use items derived from existing SET data bases; however, these items are often unvalidated. We argue that if this is the approach to benchmarking that is desired, a better approach is to agree to use already validated items from other large scale surveys used in Australia or other countries, or where these are not available or not appropriate for use at a particular SET level, to develop and validate new scales and items. In Stage Two of the project, we therefore examined the SET research literature to identify validated scales and items which could be used for benchmarking, working within the broader agenda of the TQI Framework for Teaching and Learning Quality Indicators (Chalmers, 2007).

An alternative approach to benchmarking could use the shared focus on a particular dimension of teaching, reflected in differently worded SET items, as a basis for a conversation about that dimension. For example – five different universities using five different SET items to probe “feedback” (see below) could compare what they know about students’ experiences of feedback in their settings, with a view to learning how to further enhance their own students’ experiences. As can be seen in the specific feedback items noted previously – each probes a slightly different aspect of feedback, so this precludes a simple numerical comparison. However while a detailed statistical comparison of the results of five different SET feedback questions might not be possible, simple comparisons between institutional results on such items could indicate areas of significant difference in student experiences which might be explored further in a benchmarking dialogue.

- I found that the comments and feedback from assessment tasks were helpful
- Feedback on my work in this unit helps me to achieve the learning outcomes.
- The teaching staff gave me helpful feedback.
- The teaching staff normally give me helpful feedback on how I am going
- My understanding of this course has improved as a result of feedback.
- Adequate and appropriate feedback was provided to assist learning and keep students informed on their progress.
- The feedback that I receive in this subject is helpful.
- I was given helpful feedback on how I was going in the course.

This is a perspective on benchmarking that explicitly goes beyond ideas of ranking and numerical comparison. However such a dialogue as to the reasons for variations in student ratings is a vital if institutions are to learn from numerical ranking and comparison exercises. As such it is perhaps better characterised as an approach which requires an earlier commitment to dialogue and collaborative analysis of data.

It was also apparent that there was a considerable variation between institutional SET practices, not only in terms of item use but the internal administration and reporting processes. While this variation is not of concern for internal institutional use, it raises challenges for between-institutional use. Two sources of policy variability which will impact upon cross-institutional attempts to benchmark results are:

Confidentiality. Seven institutions reported that results were confidential to the individual teacher; eight institutions permitted access to individual teachers’ results by line managers; and seven institutions permitted either limited or unlimited public access to teacher-level results.

Frequency of surveys. Differences in policies around the frequency of surveys may affect the representativeness of results. For instance, if a University’s policy gives complete discretion about the frequency of a survey of a staff member’s teaching, or the quality of a unit of study, the results for that institution may be biased due to self-selection compared to institutions that have a clear sampling plan, or who adopt a census approach (i.e. conduct surveys of all staff members and/or all units).

Section Three

Exploring institutional SET practice in relation to the proposed Teaching Quality Teaching Quality Framework

6. MOVING FORWARDS: OUTLINE OF STAGE TWO

The second stage of the study explored the preliminary analyses and potential for participating institutions' SET data to contribute information relevant to the Teaching Quality Indicators framework. In doing so it considers some ways the issues identified in Stage One of this study might be addressed.

In relation to the use of SET data for the proposed TQI dimensions, the Abrami et al. (2007) 'Dimensions of Teaching' framework was mapped against the dimensions proposed for the 'Teaching Quality Framework' dimensions. This not only explores the coverage of the proposed dimensions in relation to the empirically derived Abrami et al. dimensions but it provides a way for institutional SET contacts to relate their existing SET items to the proposed TQI dimensions.

Following this the data collected in Stage One of the study was used to identify validated or currently used SET scales and items with the potential to provide reliable data in relation to the proposed Teaching Quality Indicators dimensions.

The study then considered some of the key issues to be addressed by those seeking to share SET data between institutions. These included the challenges of (i) Developing shared data sources based on common SET items, (ii) Establishing acceptable aggregation methods, (iii) Establishing a shared benchmarking philosophy and agreed data sharing procedures, (iv) Identifying appropriate alternative and complementary data sources, and (vi) establishing collaborative networks of SET experts to enable the potential of SET benchmarking to be realised.

The remainder of this report considers the following Stage Two outputs:

1. Mapping of the Abrami et al. dimensions to the proposed Teaching Quality Dimensions
2. A proposal for suitable SET survey scales identified as being relevant to the Teaching Quality Dimensions. These existing scales and items have potential for development as shared (cross-institutional) SET data sources.
3. A suggested SET item development and validation process. Where possible the scales and items proposed in relation to the TQI Dimensions were drawn from SET surveys with published reliability and validity data. Where no validated SET scales were identified in the literature, suitable items currently in use in universities' surveys are identified as a starting point for further consideration and validation. A process for undertaking such validation is outlined.
4. Preliminary guidelines are provided for acceptable aggregation methods for data collected at different institutional levels.

5. Preliminary considerations for establishing shared benchmarking philosophies and agreed data sharing procedures are discussed.
6. Suggestions for appropriate alternative and complementary data sources are considered.
7. Suggestions are provided for ongoing institutional collaborations to develop and validate new SET scales and items.

Context: A new agenda for teaching and learning in Australian universities

As part of the overarching Teaching Quality Indicators Project, Chalmers (2007) developed a framework identifying indicators and outcomes of teaching quality operating at multiple levels in tertiary institutions. This framework was developed based on an international review of performance indicators and is articulated in terms of four broad dimensions and 24 sub-dimensions. The dimensions are not independent of each other, but are intended to provide an institution with different lenses to examine its mission and practices. Such self-examination is intended to enable institutions to better use evidence gathered through this process to improve the student experience in general, and student learning in particular.

The framework takes as a starting point Swail, Redd, and Perna's (2003) geometric model of student persistence and achievement. This model distinguishes between *cognitive factors* related to the student (e.g. quality of learning; aptitude; academic-related extracurricular activities), *social factors* impacting on the student (e.g. financial issues; attitude towards learning; family influence), and *institutional factors* affecting student persistence and achievement (e.g. financial aid; student services; curriculum and instruction).

This model has been used extensively in the USA to develop programmes to improve retention (see <http://www.isra-online.com/>), but Chalmers (2007) notes that these programmes have focussed strongly on institutional factors, with much less emphasis on social and cognitive factors, or on the engagement, recognition and rewarding of staff. As a result,

"The Framework endeavours to encompass not only a number of important institutional factors but some of the social and cognitive factors related to students. In addition, it accounts for the importance of staff, their engagement and career development in the context of the institutional mission and aspirations."
(Chalmers, 2007; p.5).

It is important to note that the Framework is not overly prescriptive in its values, design or goals. It recognises the considerable diversity in the Australian higher education sector, and aims to support individual institutions in making their own judgments; defining their

own missions, goals and priorities; and planning their own course of action (Chalmers, 2007). As such the framework encourages institutions to adopt a more internal and contextual approach to data in relation to teaching quality. The considerable diversity in current internal SET data identified in the first part of this report poses a challenge for institutions seeking to use their internal SET data in support of meaningful cross-institutional dialogues under the new Teaching Quality Framework.

Table 5 summarises the proposed Teaching Quality Framework dimensions (Chalmers 2007).

Table 5: Proposed ' Teaching Quality Framework dimensions (Chalmers, 2007).

1. Institutional climate and systems	Institution-wide	Faculty	Department / program	Subject	Teacher / Individual
1.1 Adoption of a student-centred learning perspective					
1.2 Recognising and rewarding quality teaching and contribution to student learning					
1.3 Possession of desirable teacher characteristics					
1.4 Relevant and appropriate teaching experience, qualifications and development					
1.5 Use of current research findings in informing teaching and curriculum / course content					
1.6 Community engagement / partnership					
1.7 Funding model in support of teaching and learning					
2. Diversity & Inclusivity					
2.1 Valuing and accommodating student and staff diversity					
2.2 Provision of adequate support services					
2.3 Active recruitment and admissions					
2.4 Provision of transition and academic support					
2.5 Active staff recruitment					
2.6 Multiple pathways for reward and recognition of staff					
3. Assessment					
3.1 Assessment policies address issues of pedagogy					
3.2 Adopting an evidence-based approach to assessment policies					
3.3 Alignment between institutional policy for best practice and faculty/ departmental activities					
3.4 Commitment to formative assessment					
3.5 Provision of specific, continuous and timely feedback					
3.6 Explicit learning outcomes					
3.7 Monitoring and review of standards and assessment tasks					
4. Engagement and learning community					
4.1 Student engagement					
4.2 Fostering and facilitating (academic) learning communities					
4.3 Engaging and identifying with a learning community					
4.4 Staff engagement					



Mapping the Teaching Quality dimensions

The two sets of ‘teaching’ dimensions discussed in this report differ in several ways. The proposed Dimensions of Teaching Quality (Chalmers, 2007) were developed on the basis of a review of the literature and current practice in relation to teaching quality assurance. The Abrami et al. dimensions used in the first stage of this study were empirically derived from a different literature – the relevant research studies published on student evaluation of teaching. The Abrami et al. and TQI dimensions, while sharing an interest in the quality of teaching and learning, do differ in several ways:

Table 6: Differences between the Teaching Quality Framework dimensions and the Abrami et al. dimensions

	Teaching Quality Dimensions (Chalmers)	SET Dimensions (Abrami et al.)
Derivation method	Literature and practice review (inc empirical studies)	Empirical analysis of previously reported studies
Orientation	Future oriented (tomorrow’s practices)	Historical (previous SET studies)
Data Source	Information from different sources about T&L	Information from one source (student surveys) about T&L
Focus	A systemic/holistic evaluation of teaching and learning	Student evaluation of teaching and learning

In some ways the Abrami et al. dimensions might be thought of as forming a subset or part of the TQI dimensions. That is, they provide a particular perspective (that of the students) gathered in a particular way (surveys) about teaching. As such we were interested to map the Abrami dimensions to the proposed TQI dimensions to see if there were particular dimensions where the student perspective was more relevant, and conversely to see if there were any aspects of the student perspective which were not represented in the TQI dimensions.

In the first part of the study, we used the empirically derived framework based on Abrami et al. (2007) to understand the current foci of SET in Australian universities. The framework based on Abrami et al.’s work captures the historical variation in SET research foci, particularly in the USA’s higher education sector. In contrast, the framework developed by Chalmers (2007) was based on contemporary theory and research in higher education, to provide guidance to the sector in improving the student experience. In keeping with the Stage One strategy, the Abrami et al. dimensions were mapped against the proposed TQI dimensions to see if there were any dimensions that were



considered important in SET systems that were absent in the TQI framework. We validated the components of the proposed four Teaching Quality dimensions by mapping the Abrami et al. categories against them. The mapping was conducted by examining the literature review provided by Chalmers (2007) relating to each dimension and sub-dimension, and determining where there was a correspondence. These correspondences are presented in Table 7.

The degree of commonality between the TQI and Abrami et al. frameworks varied according to the TQI dimension in question. For “Institutional Climate and Systems”, there were many Abrami et al. dimensions which were encompassed by the broader dimension of “Possession of desirable teacher characteristics” (1.3). However, the other proposed components (1.1 Adoption of a student-centred learning perspective; 1.2 Recognising and rewarding quality teaching and contribution to student learning 1.4 Relevant and appropriate teaching experience, qualifications and development; 1.5 Use of current research findings in informing teaching and curriculum / course content; 1.6 Community engagement / partnership; and 1.7 Funding model in support of teaching and learning) were not reflected in the traditional SET literature.

Likewise, for “Diversity & Inclusivity” there was relatively little overlap, with only one TQI dimension (2.1 Valuing and accommodating student and staff diversity) having corresponding traditional SET measures. “Provision of adequate support services” (2.2), “Active recruitment and admissions” (2.3), “Provision of transition and academic support” (2.4), “Active staff recruitment” (2.5), and “Multiple pathways for reward and recognition of staff” (2.6) did not appear in the Abrami et al. framework.

For “Assessment”, there was considerably more overlap, with several of the proposed TQI sub-dimensions having correspondences in the traditional SET research literature. Exceptions included “Adopting an evidence-based approach to assessment policies” (3.2), “Alignment between institutional policy for best practice and faculty/ departmental activities” (3.3), and “Monitoring and review of standards and assessment tasks” (3.7).

Lastly, for “Engagement and learning community”, all of the proposed components of this dimension had correspondences in the traditional SET research literature, except for “Staff engagement” (4.4).

Table 7: Correspondences between the Teaching Quality Framework dimensions and the Abrami et al. dimensions

Teaching Quality Dimensions (Chalmers)	Mapped Abrami et al. (2007) dimensions
1. Institutional climate and systems	
1.3 Possession of desirable teacher characteristics	1.3 Knowledge of Domain: The students are assessing the instructor's knowledge of the specific course subject matter and its applications. 1.6 Preparation and Organization: The students are evaluating the extent to which the instructor prepared himself/herself for instruction. 3.1 Answering Questions: The students are evaluating the extent to which the instructor encouraged students to ask questions and responded to students' questions appropriately. 3.2 Vocal Delivery: The extent to which the instructor demonstrated skill in vocal delivery. 3.3 Clarity of Instruction: The students are evaluating the extent to which the instructor delivers clear, concise, understandable and accurate instruction (e.g., lectures, laboratories, etc.). 3.4 Dramatic Delivery: The students are evaluating the extent to which the instructor delivered instruction in an expressive, dynamic, dramatic or exaggerated manner. 3.5 Interaction and Discussion: The students are evaluating the extent to which the instructor modelled, encouraged and achieved interactive classes in which both students and instructor contributed to the class. 3.7 Monitoring Learning: The students are evaluating the extent to which the instructor monitored students' reactions and taught at the appropriate individual and class level. 3.8 Stimulation of Interest in the Course: The students are evaluating the extent to which the instructor stimulated their interest in the course by using a variety of activities, manifested by the extent to which good attendance, increased interest, outside reading, and liking/enjoyment for the subject matter were exhibited. 5.5 Enthusiasm for Teaching: The students are evaluating the extent to which the instructor communicates his/her enthusiasm, interest or liking for teaching. 5.9 Motivating Students to Greater Effort: The students are evaluating the extent to which the instructor motivated students to more effort, intellectual curiosity, love of learning, high academic aspirations, etc. 5.10 Respect for Others: The students are evaluating the extent to which the instructor modeled, encouraged and showed trust, respect, and consideration for others (e.g., listened without interruption, did not belittle or criticize others' criticism, treated others as equals, was punctual, etc.).
2. Diversity & Inclusivity	
2.1 Valuing and accommodating student and staff diversity	5.6 Friendly Classroom Climate: The students are evaluating the extent to which the instructor modelled, encouraged and achieved a friendly and safe classroom. 5.11 Tolerance of Diversity: The students are evaluating the extent to



	which the instructor modelled, encouraged and achieved tolerance for a diversity of opinions, ideas and viewpoints and an absence of prejudice in the classroom.
3. Assessment	
3.1 Assessment policies address issues of pedagogy	6.1 Assessment Tasks (labelled simply as 'Assessment' by Abrami): The students are evaluating the extent to which the instructor's tests were appropriate in terms of content, frequency, time allocation, weight, difficulty, validity and learning opportunity. They are also evaluating the instructor's fairness and consistency in grading.
3.4 Commitment to formative assessment	6.1 Assessment Tasks (labelled simply as 'Assessment' by Abrami): The students are evaluating the extent to which the instructor's tests were appropriate in terms of content, frequency, time allocation, weight, difficulty, validity and learning opportunity. They are also evaluating the instructor's fairness and consistency in grading. 6.2 Feedback: The students are evaluating the instructor's use of review and feedback (frequency, positive/negative) and its effect on students.
3.5 Provision of specific, continuous and timely feedback	6.2 Feedback: The students are evaluating the instructor's use of review and feedback (frequency, positive/negative) and its effect on students.
3.6 Explicit learning outcomes	4.1 High level Cognitive Outcomes: The students are evaluating the extent to which the instructor is promoting high level cognitive outcomes such as writing skills, reasoning, meta cognition, problem solving, etc. 4.2 Low level Cognitive Outcomes: The students are evaluating the extent to which the instructor is promoting low level cognitive outcomes (e.g., recall, recognition, knowledge, etc.). 4.4 Overall Learning: The students are evaluating the overall quality and relevance of the perceived learning that took place including the achievement of short and long term objectives. 6.3 Clear Criteria and Standards (Labelled as 'Objectives' by Abrami): The students are evaluating the extent to which the instructor communicated performance criteria and deadlines for assignments and tests. 6.4 Overall Assessment: The students are evaluating the overall extent to which the assessment supported their learning.
4. Engagement and learning community	
4.1 Student engagement	3.5 Interaction and Discussion: The students are evaluating the extent to which the instructor modelled, encouraged and achieved interactive classes in which both students and instructor contributed to the class.
4.2 Fostering and facilitating (academic) learning communities	5.12 Overall: Learning Climate: The students are evaluating the overall extent to which the learning environment fostered their active engagement as valued members of a learning community.
4.3 Engaging and identifying with a learning community	5.12 Overall: Learning Climate: The students are evaluating the overall extent to which the learning environment fostered their active engagement as valued members of a learning community.



The above mapping also identifies how items in a particular traditional SET categories might generate useful data in relation to the TQI dimension, and may assist SET experts in understand their existing surveys in terms of such dimensions and locating relevant items for the less familiar TQI dimensions.

However, there were additional TQI dimensions which could not be mapped against the Abrami et al. dimensions, given below in Table 8, suggesting that historical SET practices have not kept pace with developments in theory, nor has SET always gathered information on all important teaching quality dimensions. Indeed, Abrami, Rosenfield, and Dedic (2007) have argued that traditional SET foci – particularly a focus on teacher-centred rather than student-centred learning – are impeding these new developments in higher education pedagogy in that they support the status quo of teaching practices and institutional systems.

Table 8: Unmapped Teaching Quality Framework dimensions

Unmapped TQI Dimensions	Sub-Dimensions
1. Institutional Climate and Systems	1. Student-centred learning approach 2. Valuing teachers and teaching – surveys which identify the quality of staff experience 3. Provision of support services 4. Use of current research findings to inform teaching and curriculum/course content 5. Community engagement/partnership 6. Funding model that supports learning and teaching
2. Diversity and inclusivity	In particular, aspects of this dimension operating at the program, department and institutional levels (e.g. provision of relevant work-based/research-based experiences.)
3. Student engagement and learning community	1. Additional Student Engagement dimensions such as Active Learning, Academic Challenge, Enriching Educational Experiences, and Supportive Learning Environment (from AUSSE). 2. Staff engagement and learning community.

The mapping exercise served to identify the degree of commonality between the Abrami et al. SET dimensions, and the Teaching Quality Dimensions. While there was some overlap – raising the possibility of using some existing institutional scales and items – there were a number of TQI sub-dimensions which were not part of the Abrami et al. framework. In the next section of this report, we describe a process for identification and development of scales and items to address this shortfall.



Engaging institutional SET experts: National workshop

As noted in the Stage One report, SET data is a central component of teaching quality assurance evidence in Australian and internationally. It will provide an important source of data for the proposed Teaching Quality framework. As such the institutional contacts nominated for this study represent a key institutional resource in relation to implementing the proposed framework.

A workshop was held for nominated institutional contacts to facilitate engagement with the second stage of the project. The workshop was held in conjunction with the annual Australasian Evaluation Forum as this event is regularly attended by the SET representatives from many Australian universities.

In the lead-up to this workshop all participating Stage One institutions were invited to review the data submitted in Stage One, as well as the coding and analysis of this data for their particular institution. The first part of this workshop also provided those attending the event with the opportunity to review and make any additions to the data institutional contacts provided in Stage One. This data included an overview of SET practices at different institutional levels as well as the allocation of the core SET items from current institutional student feedback surveys, against the categories identified in a recent review of the major multi-section validity studies. This data was provided using a web based interface and in some cases involved respondents contributing many items from a large and complex collection of surveys or providing us with an item bank from which we made allocations. The meeting was an opportunity to review the descriptions and allocations and to add to the data for each participant's institution.

The second part of the workshop provided an opportunity for institutional contacts to review the results of the preliminary analysis and also to consider how they might like to share the results of the data collection to date.

The third part of the workshop considered how the results of the preliminary analysis might contribute to the larger Project on 'Rewarding and recognising quality teaching and learning in higher education'. The group considered the challenges of establishing meaningful conversations between institutions in relation to the SET data collected at different organisational levels. This consideration involved participants as collaborators in the identification of potential sets of items, scales and teaching evaluation processes which might facilitate sharing of such quality enhancement data in the future.

The workshop concluded with the invitation to participants to consider establishing small groups of potential benchmarking partner institutions to continue to work on developing and sharing SET data. The workshop presentation is included as Appendix 3 to this report.

7. DEVELOPING BENCHMARKING SCALES AND ITEMS FOR THE TEACHING QUALITY FRAMEWORK

As noted above in Chapters 5 and 6, the variability in SET scales and items used across the Australian higher education sector makes meaningful benchmarking difficult if not impossible at present, particularly at the Teacher and Subject levels of evaluation.

In order to support this capacity, a suite of research-based, validated scales and items is needed which institutions can use with confidence that results will be comparable. In this section, we make suggestions for such scales and items, as well as processes by which currently unavailable measures might be developed. We also discuss issues around the validation and analysis of such measures, and issues related to the aggregation of responses gathered at one level (e.g. Subject) to measure quality at higher levels (e.g. overall quality of teaching at the school or department level).

The process of scale and item selection or development was as follows.

1. Where possible, the scales and items suggested below were included on the basis of having gone through a process of empirical development reported in the SET research literature. References from peer-reviewed journals or Commonwealth Government documents, which describe each scale's theoretical underpinning and psychometric testing, are provided below. Having each gone through a process of scale development and testing, it is expected that these scales will be suitable for use in a variety of contexts, and for benchmarking specifically. (Ongoing testing of the psychometric properties of these instruments should nonetheless be carried out to ensure their fitness for purpose; for example, Graduate Careers Australia examines the functioning of the CEQ each year to ensure its purported factor structure is robust). In some cases, a suggested scale may have one particular item that addresses a desired dimension well, but other items which are less clearly related; these scales have been included on the basis that constituent items have undergone psychometric testing and have been judged as a suitable part of a broader dimension.
2. Where existing psychometrically validated scales and items were not available for specific dimensions, items provided by participating universities in the Study 1 dataset, or from Appendix A of items given in Feldman (1989), were examined for suitability, based on the quality of wording and face validity of the item. Where no pre-existing items were available, a proposed exemplar item is given, drawing on the essence of the dimension in question. It should be noted that while these items were selected and proposed on the basis of sound survey item design principles, none of the necessary validation or construct validity testing has been carried out. They are provided as a starting point only for such collaborative work.

In the case of items which are suggested as starting points for measures of dimensions where no previously developed scales or items are available, the proposed items are relatively broadly worded rather than focusing on very specific characteristics of the dimension. Examples of dimensions and associated items phrased in this way include:

- *Teacher's Knowledge Base*: "The teaching staff seemed to know the subject matter well."
- *Effective Communication with Students*: "Communication between my teachers and I helped me learn in this subject."
- *Staff engagement*: "Staff in this subject seemed really engaged in their teaching."

In this way, the proposed items resemble 'global SET ratings'. Abrami et al. (2007) argued on the basis of their research into the dimensionality of SET ratings that global ratings (e.g. of an instructor, a subject, or of perceived learning) had the highest validity coefficients, and were less likely to be affected by different instructional settings (e.g. disciplinary differences) than specific SET dimensions. There is no suggestion that a single item necessarily encapsulates a whole dimension. Moreover given the multidimensionality of SET and good teaching identified in other research it is indeed likely that each of the Teaching Quality dimensions will require several SET items. Nonetheless, we argue that beginning with more broadly framed items will provide a more productive starting point than items which specify tightly focussed facets of proposed dimensions.

Readers will note that several rows do not include suggested scales and/or items. The reasons for these exclusions are as follows:

Table 9: Reasons suggested scales / items not proposed

Dimension	Reason suggestion/s not proposed
1.5 Use of current research findings to inform teaching and curriculum/course content	A literature search did not find any sufficiently well-validated items or scales for this construct at any of the 5 levels. Extended scale development work will be needed if this construct is to be measured with SETs.
2.5 Active staff recruitment	Students cannot be expected to comment reliably on this aspect of University function.
2.6 Multiple pathways for reward and recognition of staff.	Students cannot be expected to comment reliably on this aspect of University function.
3.1 Assessment policies address issues of pedagogy	Students cannot be expected to comment reliably on this aspect of University function.
3.2 Adopting an evidence-based approach to assessment policies	Students cannot be expected to comment reliably on this aspect of University function.
3.3 Alignment between institutional policy for best practice and faculty/ departmental activities	Students cannot be expected to comment reliably on this aspect of University function.

The following table (Table 10) is designed to provide readers with a starting point in identifying SET items and scales that might be used for benchmarking specific dimensions in the TQI framework between universities. The tables present suggested



scales and/or items for benchmarking dimensions of the student experience across Australian universities.

The following table is divided in to 4 sections. Each section represents one of the four broad dimensions of the TQI framework. The columns of the table represent the 5 levels at which SET data might be collected. Readers who are interested in identifying potential SET items for a particular component of the TQI framework should first identify which organisational level data for which data is required, then look down the dimensions.

The four broad dimensions of the TQI framework are:

1. Institutional climate and systems
2. Diversity
3. Assessment
4. Student engagement and learning community

Where meaningful, alternative scales and/or items are given for each dimension at 5 levels:

1. Whole university (institution-wide)
2. Faculty/School/Department
3. Degree/Programme
4. Unit of Study/Subject
5. Individual Teacher

As noted above, we argue that not all of the dimensions can be meaningfully measured using SET scales or items. For these dimensions, alternative sources of institutional performance will be needed; for example, institutional performance on the “Active Staff Recruitment” and “Adopting an evidence-based approach to assessment policies” dimensions could be examined using document analysis of Human Resources and School/Department/Faculty records respectively.

For several cells in the table, there are multiple suggestions for scales and items measuring a given construct; institutions will therefore often have choices as to the scales or items which seem most appropriate in their context, or alternative forms could be piloted to determine which had the best psychometric properties in particular contexts.

Readers will also note that many cells at the “Degree/Programme”, “Faculty/School/Department” and “Whole university” levels suggest that SET scales used at lower levels be aggregated up to higher levels. Such aggregations are often used in SET data – for example, CEQ data collected at the degree/programme level is often aggregated up to broader fields of study – but evidence for the suitability of aggregation is less frequently provided. We discuss issues surrounding the use of aggregation in SET data further below.

Table 10: Suggested benchmarking scales & items for the TQI SET Framework

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.1 Adoption of a student-centred learning approach	Aggregated Ramsden (1991) CEQ <i>Good Teaching Scale</i>	Aggregated Ramsden (1991) CEQ <i>Good Teaching Scale</i>	Ramsden (1991) CEQ <i>Good Teaching Scale</i> 1. The teaching staff of this course motivated me to do my best work 2. The staff put a lot of time into commenting on my work 3. The staff made a real effort to understand difficulties I might be having with my work	Kember & Leung (in press-a) <i>Active Learning Scale</i> 1. Students were given the chance to participate in class. 2. There was discussion between students in class. 3. The teaching staff promoted discussion in class.	Kember & Leung (in press-a) <i>Active Learning Scale</i> 1. Students were given the chance to participate in class. 2. There was discussion between students in class. 3. My teacher promoted discussion in class. Kember & Leung (in press-a) <i>Flexibility Scale</i> 1. I found my teacher helpful when I had difficulty understanding concepts. 2. My teacher was sensitive to student feedback. 3. My teacher was helpful when asked questions.
	Aggregated Kember & Leung (in press-b) <i>Active Learning Scale</i>	Aggregated Kember & Leung (in press-b) <i>Active Learning Scale</i>	4. The teaching staff normally gave me helpful feedback on how I was going 5. My lecturers were extremely good at explaining things 6. The teaching staff worked hard to make their subjects interesting	Kember & Leung (in press-a) <i>Flexibility Scale</i> 1. I found teaching staff helpful when I had difficulty understanding concepts. 2. The teaching staff were sensitive to student feedback. 3. The teacher(s) were helpful when asked questions.	Kember & Leung (in press-a) <i>Teacher-Student Relationships Scale</i> 1. There was a friendly relationship between the teacher and students. 2. The communication between my teacher and students is good. 3. Our teacher knew the individuals in the class.
	Aggregated Kember & Leung (in press-b) <i>Relationship between Teachers and Students Scale</i>	Aggregated Kember & Leung (in press-b) <i>Relationship between Teachers and Students Scale</i>			Marsh & Roche (1994) <i>SEEQ Group Interaction Scale</i>
	Aggregated Kember & Leung (in press-b) <i>Teaching for Understanding Scale</i>	Aggregated Kember & Leung (in press-b) <i>Teaching for Understanding Scale</i>			
	Aggregated Richardson et al. (2007) NSS (UK) <i>Teaching Scale</i>	Aggregated Richardson et al. (2007) NSS (UK) <i>Teaching Scale</i>	Kember & Leung (in press-b) <i>Active Learning Scale</i> 1. Our teaching staff use a variety of teaching methods. 2. Students are given the chance to participate in class.	Kember & Leung (in press-a) <i>Teacher-Student Relationships Scale</i> 1. There was a friendly relationship between teaching staff and students. 2. The communication between teaching staff and students is good. 3. Our teacher(s) knew the individuals in the class.	
	Aggregated Coates (2006) <i>Constructive Teaching Scale</i>	Aggregated Coates (2006) <i>Constructive Teaching Scale</i>			
	Aggregated Coates (2006) <i>AUSSE Teacher Approachability Scale</i>	Aggregated Coates (2006) <i>AUSSE Teacher Approachability Scale</i>	Kember & Leung (in press-b) <i>Relationship between Teachers and Students Scale</i> 1. The communication between teaching staff and students is good.		

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.1 Adoption of a student-centred learning approach (Continued)			2. I find teaching staff helpful when asked questions. Kember & Leung (in press-b) <i>Teaching for Understanding Scale</i> 1. The teaching staff try hard to help us understand the course material. 2. The course design helps students understand the course content. Richardson et al. (2007) NSS (UK) <i>Teaching Scale</i> 1) Staff were good at explaining things. 2) Staff made the subject interesting. 3) Staff were enthusiastic about what they were teaching. Coates (2006) <i>Constructive Teaching Scale</i> 1. Materials were presented in a way that I could understand. 2. Staff valued students' ideas and questions. 3. Staff encouraged students to question what was being taught. 4. Academics used teaching approaches that suited students' needs. 5. I felt encouraged to	<i>Possible exemplar item</i> The teaching in this subject took account of my needs as a learner.	1. Students were encouraged to participate in class discussions. 2. Students were invited to share their ideas and knowledge. 3. Students were encouraged to ask questions and give meaningful answers. 4. Students were encouraged to express their own ideas and/or question the lecturer.

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.1 Adoption of a student-centred learning approach (Continued)			creatively explore ideas. 6 Staff talked about their research in ways that inspired me to learn. <i>Coates (2006) AUSSE Teacher Approachability Scale</i> 1. Staff seemed interested in helping students. 2. Staff were generally approachable. 3. Teaching staff were accessible. 4. Teaching staff made a real effort to interact with students.		
1.2 Recognising and rewarding quality teaching and contribution to student learning	This University recognises teachers for their contribution to student learning.	This school/department/faculty recognises teachers for their contribution to student learning.	Single Item - Outstanding subjects in my course were recognised by the university.		
1.3 Possession of desirable teacher characteristics – Overall Ratings	Aggregated Ramsden (1991) CEQ – <i>Good Teaching Scale</i> Aggregated Coates (2006) <i>Constructive Teaching Scale</i>	Aggregated Ramsden (1991) CEQ – <i>Good Teaching Scale</i> Aggregated Coates (2006) <i>Constructive Teaching Scale</i>	Ramsden (1991) CEQ – <i>Good Teaching Scale</i> <i>Coates (2006) Constructive Teaching Scale</i>	<i>Possible exemplar item</i> The teaching in this unit of study helped me to learn effectively (Ginns & Barrie, 2004)	Overall my lecturer effectively supported my learning
1.3 Possession of Desirable Teacher Characteristics – Specific Dimensions	<i>1.3.1 Teacher Clarity</i> Ramsden (1991) CEQ – Aggregated <i>Clear Goals and Standards Scale</i>	<i>1.3.1 Teacher Clarity</i> Ramsden (1991) CEQ – Aggregated <i>Clear Goals and Standards Scale</i>	<i>1.3.1 Teacher Clarity</i> Ramsden (1991) CEQ – <i>Clear Goals and Standards Scale</i> 1. It was always easy to know the standard of work expected	<i>1.3.1 Teacher Clarity</i> Kember & Leung (in press-a) <i>Understanding Fundamental Concepts</i> Scale 1. This unit concentrated on fundamental	<i>1.3.1 Teacher Clarity</i> <i>Students' Evaluation of Educational Quality</i> (SEEQ; Marsh & Roche, 1994) <i>Organisation/Clarity Scale</i>

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.3 Possession of Desirable Teacher Characteristics – Specific Dimensions (Continued)	<i>1.3.2 Teacher Organisation</i> Aggregated Kember & Leung (in press-b) <i>Coherence of Curriculum Scale</i> Aggregated Griffin et al. (2003) Extended CEQ Course Organisation Scale	<i>1.3.2 Teacher Organisation</i> Aggregated Kember & Leung (in press-b) <i>Coherence of Curriculum Scale</i> Aggregated Griffin et al. (2003) Extended CEQ Course Organisation Scale	<i>1.3.2 Teacher Organisation</i> Kember & Leung (in press-b) <i>Coherence of Curriculum Scale</i> 1. I can see how units fitted together to make a coherent programme for my major. 2. The programme of study for my major was well integrated.	concepts. 2. In each class the key points were made clear. 3. In this unit I learnt the key principles. <i>Possible exemplar item</i> Ideas in this subject were clearly communicated. <i>1.3.2 Teacher Organisation</i> Kember & Leung (in press-a) <i>Organisation Scale</i> 1. This unit was well organized. 2. This unit was well planned. 3. Each class was well planned.	1. Lecturer's explanations were clear. 2. Class materials were well-prepared and carefully explained. 3. Proposed objectives agreed with those actually taught so you knew where the class was going. 4. Lecturer gave presentations that facilitated taking notes. <i>Items from Phase 1 Survey</i> 1. The lecturer explained important concepts and ideas in ways that I could understand. 2. During each lecture, the lecturer described what students were expected to learn from that lecture. <i>1.3.2 Teacher Organisation</i> <i>Students' Evaluation of Educational Quality (SEEQ; Marsh & Roche, 1994)</i> <i>Organisation/Clarity Scale</i> 1. Lecturer's explanations were clear. 2. Class materials were well-prepared and carefully explained. 3. Proposed objectives agreed with those actually taught so you

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.3 Possession of Desirable Teacher Characteristics – Specific Dimensions (Continued)			Griffin et al. (2003) Extended CEQ <i>Course Organisation Scale</i> 1. The course was well organised. 2. I was given helpful advice when planning my academic programme. 3. The course content was organised in a systematic way. 4. There was sufficient flexibility in my course to suit my needs. 5. I had enough choices of the topics I wanted to study.	<i>Possible exemplar item</i> The way this subject was organised helped me to learn.	knew where the class was going. 4. Lecturer gave presentations that facilitated taking notes. <i>Items from Phase 1 Survey</i> 1. My teacher came to class well-prepared to help me learn. 2. The lecturer produced classes that were well organised.
1.3 Possession of Desirable Teacher Characteristics – Specific Dimensions (Continued)	<i>1.3.3 Motivation & Stimulation of Interest</i> Aggregated Griffin et al. (2003) Extended CEQ Intellectual Motivation Scale Aggregated Ramsden (1991) CEQ Good Teaching Scale item 1. The teaching staff worked hard to make their subjects interesting.	<i>1.3.3 Motivation & Stimulation of Interest</i> Aggregated Griffin et al. (2003) Extended CEQ Intellectual Motivation Scale Aggregated Ramsden (1991) CEQ Good Teaching Scale item 1. The teaching staff worked hard to make their subjects interesting.	<i>1.3.3 Motivation & Stimulation of Interest</i> Griffin et al. (2003) Extended CEQ Intellectual Motivation Scale 1. I found my studies intellectually stimulating. 2. I found the course motivating. 3. The course has stimulated my interest in the field of study. 4. Overall, my university experience was worthwhile. Ramsden (1991) CEQ Good Teaching Scale item 1. The teaching staff worked hard to make their subjects interesting.	<i>1.3.3 Motivation & Stimulation of Interest</i> Kember & Leung (in press-a) Motivation Scale 1. The teacher(s) were enthusiastic. 2. I found the classes enjoyable. 3. This was an interesting unit. Kember & Leung (in press-a) Relevance Scale 1. Local examples were used to show the relevance of material. 2. I could see the relevance of materials because real life examples were given. 3. Current issues were used to make the course interesting.	<i>1.3.3 Motivation & Stimulation of Interest</i> Students' Evaluation of Educational Quality (SEEQ; Marsh & Roche, 1994) Learning/Academic Value Scale 1. You found the class intellectually challenging and stimulating. 2. You have learned something you consider valuable. 3. Your interest in the subject has increased as a consequence of this class. 4. You have learned and understood the subject materials in this class. Items from Phase 1 Survey 1. The lecturer stimulated

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.3 Possession of Desirable Teacher Characteristics – Specific Dimensions (Continued)				<i>Possible exemplar item</i> I was motivated to engage with the learning activities in this subject.	my interest in the topic area. 2. The lecturer stimulated me to think about issues in this subject area.
1.3 Possession of Desirable Teacher Characteristics – Specific Dimensions (Continued)	<i>1.3.4 Teacher Enthusiasm Towards Teaching</i> Aggregated item from Kember & Leung (in press-a) <i>Motivation Scale</i> 1. The teacher(s) were enthusiastic.	<i>1.3.4 Teacher Enthusiasm Towards Teaching</i> Aggregated item from Kember & Leung (in press-a) <i>Motivation Scale</i> 1. The teacher(s) were enthusiastic.	<i>1.3.4 Teacher Enthusiasm Towards Teaching</i> Aggregated item from Kember & Leung (in press-a) <i>Motivation Scale</i> 1. The teacher(s) were enthusiastic.	<i>1.3.4 Teacher Enthusiasm Towards Teaching</i> Kember & Leung (in press-a) <i>Motivation Scale</i> 1. The teacher(s) were enthusiastic. 2. I found the classes enjoyable. 3. This was an interesting unit. <i>Possible exemplar item</i> My teachers enjoyed teaching this subject.	<i>1.3.4 Teacher Enthusiasm Towards Teaching Students' Evaluation of Educational Quality</i> (SEEQ; Marsh & Roche, 1994) <i>Instructor Enthusiasm Scale</i> 1. Lecturer was enthusiastic about teaching the class. 2. Lecturer was dynamic and energetic in conducting the class. 3. Lecturer enhanced presentations with the use of humour. 4. Lecturer's style of presentation held your interest during class.
1.3 Possession of Desirable Teacher Characteristics – Specific Dimensions (Continued)	<i>1.3.5 Teacher's Knowledge Base</i> Aggregate up from Degree/Programme level	<i>1.3.5 Teacher's Knowledge Base</i> Aggregate up from Degree/Programme level	<i>1.3.5 Teacher's Knowledge Base Adapted from Items from Phase 1 Survey</i> 1. The teaching staff exhibit sound knowledge of the degree. 2. The teaching staff in this degree seemed to know the subject matter well.	<i>1.3.5 Teacher's Knowledge Base Items from Phase 1 Survey</i> 1. The teaching staff exhibit sound knowledge of the course. 2. The teaching staff seemed to know the subject matter well.	<i>1.3.5 Teacher's Knowledge Base Items from Phase 1 Survey</i> 1. This teacher really understands the content of this subject. 2. The staff member seemed to know the subject matter very well.

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.3 Possession of Desirable Teacher Characteristics – Specific Dimensions (Continued)	<i>1.3.6 Effective Communication with Students</i> Aggregate Kember & Leung (in press-b) <i>Teachers Encouraging Students</i> Scale	<i>1.3.6 Effective Communication with Students</i> Aggregate Kember & Leung (in press-b) <i>Teachers Encouraging Students</i> Scale	<i>1.3.6 Effective Communication with Students</i> Kember & Leung (in press-b) <i>Teachers Encouraging Students</i> Scale 1. There is a close relationship between teaching staff and students. 2. The communication between teaching staff and students is good.	<i>1.3.6 Effective Communication with Students</i> Kember & Leung (in press-a) <i>Teacher-Student Relationships</i> Scale 1. There was a friendly relationship between teaching staff and students. 2. The communication between teaching staff and students is good. 3. Our teacher(s) knew the individuals in the class. <i>Possible exemplar item</i> Communication between my teachers and I helped me learn in this subject.	<i>1.3.6 Effective Communication with Students Phase 1 items</i> From Feldman (1989) 1. I could comprehend the language and vocabulary used by the lecturer. From Feldman (1989) 1. The instructor has a good vocal delivery.

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.3 Possession of Desirable Teacher Characteristics – Specific Dimensions (Continued)	<i>1.3.7 Respect for Students Items from Phase 1 study</i> 1. The University is free from discrimination.	<i>1.3.7 Respect for Students Items from Phase 1 study</i> 1. The learning environment is free from discrimination.	<i>1.3.7 Respect for Students Items from Phase 1 study</i> 1. The learning environment is free from discrimination.	<i>1.3.7 Respect for Students Items from Phase 1 study</i> 1. The learning environment is free from discrimination. Students were treated with respect in this subject	<i>1.3.7 Respect for Students - Students' Evaluation of Educational Quality (SEEQ; Marsh & Roche, 1994) Individual Rapport Scale</i> 1. Lecturer was friendly toward individual students. 2. Lecturer had a genuine interest in individual students. 3. Lecturer made students feel welcome in seeking help/advice in or outside of class. 4. Lecturer was adequately accessible to students during office hours or after class. <i>Items from Phase 1 study</i> 1. The lecturer treated all students fairly and equally. 2. The lecturer treated students with respect.

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.3 Possession of Desirable Teacher Characteristics – Specific Dimensions (Continued)	<i>1.3.8 Expectations of high quality learning</i> Aggregated Coates (2006) <i>AUSSE Academic Challenge Scale</i>	<i>1.3.8 Expectations of high quality learning</i> Aggregated Coates (2006) <i>AUSSE Academic Challenge Scale</i>	<i>1.3.8 Expectations of high quality learning</i> Coates (2006) <i>AUSSE Academic Challenge Scale</i> 1. Assessment tasks challenged me to learn. 2. I was given enough material to keep up my interest. 3. I was encouraged by teachers to go beyond set materials. 4. Academic staff gave me comments on my work that helped me learn. 5. I received feedback quickly enough to improve subsequent work.	<i>1.3.8 Expectations of high quality learning</i> <i>Possible exemplar items</i> I felt encouraged to <i>really</i> learn in this subject. I felt encouraged to achieve at a high standard in this subject	<i>1.3.8 Expectations of high quality learning</i> Adapted NSSE item – “About how often this semester have you worked harder than you thought you could to meet this teacher's standards or expectations?”
1.4 Relevant and appropriate teaching experience, qualifications and development	<i>Possible exemplar item</i> The experience of teaching staff in teaching this degree helped me to learn.	<i>Possible exemplar item</i> The experience of teaching staff in teaching this degree helped me to learn.	<i>Possible exemplar item</i> The experience of teaching staff in teaching this degree helped me to learn.	<i>Possible exemplar items</i> This subject was taught by skilled teachers. My learning in this subject was supported by skilled and experienced teachers.	<i>Possible exemplar item</i> The experience of my teacher in teaching this topic helped me to learn.

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.5 Use of current research findings to inform teaching and curriculum/cou rse content					

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
1.6 Community engagement/ partnership	Coates (2006) AUSSE <i>Complementary Activities Scale</i> 1. I met people with different perspectives to mine. 2. University resources helped me to develop my non-academic interests. 3. I participated in interesting events and activities around campus. 4. I participated in social clubs and societies with other students.	Disaggregated Coates (2006) AUSSE <i>Complementary Activities Scale</i>	Disaggregated Coates (2006) AUSSE <i>Complementary Activities Scale</i> Kabanoff et al. (2003) <i>Workplace Skills Scale</i> 1. The course developed an understanding of workplace skills. 2. The course provided the opportunity for putting theory into practice. 3. The course developed current professional skills. 4. The course was a good combination of theory and practice. 5. I got practical experience dealing with actual work situations. 6. What I learnt benefited my future work. 7. The course helped prepare me for the workforce. 8. Subjects had no direct relation to the world of work (reverse scored)	<i>Possible exemplar item</i> Teaching and learning in this subject was enhanced by links with external communities.	<i>Possible exemplar item</i> My teacher extended this topic to show its relevance to broader communities than our class.
1.7 Funding model that supports learning and teaching.	<i>Aggregated Possible exemplar item</i> Teaching and learning in this degree was well resourced.	<i>Aggregated Possible exemplar item</i> Teaching and learning in this degree was well resourced.	<i>Possible exemplar item</i> Teaching and learning in this degree was well resourced.	<i>Possible exemplar item</i> Teaching and learning in this subject was well resourced.	<i>Possible exemplar item</i> My teacher had sufficient resources to support teaching and learning.

	Institution-wide	Faculty/School/Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
2.1 Valuing and accommodating student and staff diversity .	Aggregated Ramsden (1991) CEQ <i>Emphasis on Independence Scale</i> Coates (2006) AUSSE <i>Supportive Learning Environment</i> 1. Staff respected students' backgrounds, perspectives and needs. 2. The university campus felt like a supportive place to learn. 3. Staff seemed responsive to feedback from students. 4. I felt part of an academic community at university.	Aggregated Ramsden (1991) CEQ <i>Emphasis on Independence Scale</i>	Ramsden (1991) CEQ <i>Emphasis on Independence Scale</i> 1. There are few opportunities to choose the particular areas you want to study. (reversed) 2. The course has encouraged me to develop my own academic interests as far as possible. 3. Students have a great deal of choice over how they are going to learn in this course. 4. Students here are given a lot of choice in the work they have to do. 5. We often discuss with our lecturers or tutors how we are going to learn in this course. 6. There's very little choice in this course in the ways you are assessed. (reversed)	<i>Possible exemplar items</i> Teaching and learning in this subject made the most of student diversity. Teaching and learning in this subject benefited from considering student diversity. Teaching in this subject took student diversity into account.	<i>Phase 1 study items</i> 1. My teacher was open to new ideas and viewpoints. 2. The lecturer accommodates cultural differences amongst students in the class. 3. Students are encouraged to think about different viewpoints and cultural perspectives.
2.2 Provision of adequate support services	Griffin et al. (2003) Extended CEQ <i>Student Support Scale</i> 1) I was able to access information technology resources when I needed them. 2) Relevant learning resources were	Aggregated Richardson et al. (2007) NSS (UK) <i>Support Scale</i>	Richardson et al. (2007) NSS (UK) <i>Support Scale</i> 1) There was good advice available to ensure that I made the choices in my studies that were right for me. 2) During my first year, I received sufficient	<i>Possible exemplar item</i> I was provided with helpful academic support during my study in this subject.	

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
2.2 Provision of adequate support services (Continued)	accessible when I needed them. 3) Health, welfare and counselling services met my requirements. 4) The library services were readily accessible. 5) I was satisfied with the course and careers advice provided. Griffin et al. (2003) <i>Extended CEQ Learning Resources Scale</i> 1) The library resources were appropriate for my needs. 2) The study materials were clear and concise. 3) It was made clear what resources were available to help me learn. 4) Course materials were relevant and up to date. 5) Where it was used, the information technology in teaching and learning was effective. Richardson et al. (2007) <i>NSS (UK) Resources Scale</i> 1) The library resources were good enough for my needs. 2) I was always able to access general IT resources when required.		support and advice with my studies. 3) Over the course as a whole, I received sufficient support and advice with my studies.		

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
2.2 Provision of adequate support services (Continued)	3) When it was needed, there was sufficient access to specialised equipment, computing facilities, or rooms.				
2.3 Active recruitment and admissions	<i>Aggregated Possible exemplar item</i> I was able to access good information about this degree to inform my choices.	<i>Aggregated Possible exemplar item</i> I was able to access good information about this degree to inform my choices.	<i>Possible exemplar item</i> I was able to access good information about this degree to inform my choices.		
2.4 Provision of transition and academic support	Griffin et al. (2003) Extended CEQ <i>Student Support Scale</i> 1) I was able to access information technology resources when I needed them. 2) Relevant learning resources were accessible when I needed them. 3) Health, welfare and counselling services met my requirements. 4) The library services were readily accessible. 5) I was satisfied with the course and careers advice provided. Griffin et al. (2003) Extended CEQ <i>Learning Resources Scale</i> 1) The library resources were appropriate for my needs. 2) The study materials were clear and concise. 3) It was made clear	Aggregated Richardson et al. (2007) NSS (UK) <i>Support Scale</i>	Richardson et al. (2007) NSS (UK) <i>Support Scale</i> 1) There was good advice available to ensure that I made the choices in my studies that were right for me. 2) During my first year, I received sufficient support and advice with my studies. 3) Over the course as a whole, I received sufficient support and advice with my studies.	<i>Possible exemplar item</i> I was provided with helpful academic support during my study in this subject.	<i>Possible exemplar item</i> My teacher supported me in starting to study this unit.

	Institution-wide	Faculty/School/Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
2.4 Provision of transition and academic support (Continued)	what resources were available to help me learn. 4) Course materials were relevant and up to date. 5) Where it was used, the information technology in teaching and learning was effective. Richardson et al. (2007) NSS (UK) <i>Resources Scale 1</i> The library resources were good enough for my needs. 2) I was always able to access general IT resources when required. 3) When it was needed, there was sufficient access to specialised equipment, computing facilities, or rooms.				
2.5 Active staff recruitment					
2.6 Multiple pathways for reward and recognition of staff					

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
3. Assessment - Overall	Aggregated Ramsden (1991) CEQ <i>Appropriate Assessment Scale</i> (cf. Ginns, Prosser & Barrie, 2007) Aggregated Kember & Leung (in press-b) <i>Assessment Scale</i> Aggregated Richardson et al. (2007) NSS (UK) <i>Assessment Scale</i>	Aggregated Ramsden (1991) CEQ <i>Appropriate Assessment Scale</i> (cf. Ginns, Prosser & Barrie, 2007) Aggregated Kember & Leung (in press-b) <i>Assessment Scale</i> Aggregated Richardson et al. (2007) NSS (UK) <i>Assessment Scale</i>	Ramsden (1991) CEQ <i>Appropriate Assessment Scale</i> 1. To do well in this course all you really needed was a good memory. 2. The staff seemed more interested in testing what I had memorised than what I had understood. 3. Too many staff asked me questions just about facts. Kember & Leung (in press-b) <i>Assessment Scale</i> 1. The programme uses a variety of assessment methods. 2. To do well in assessment in this programme you need to have good analytic skills. 3. The assessment tested our understanding of key concepts in this programme. Richardson et al. (2007) NSS (UK) <i>Assessment Scale</i> 1) The criteria to be used in marking were clear in advance. 2) Assessment arrangements and marking were fair. 3) It was clear what standard was required in assessed work	Kember & Leung (in press-a) <i>Assessment Scale</i> 1. The type of assessment related closely to the expected learning outcomes. 2. The assessment tested our understanding of key concepts. 3. A variety of assessment methods were used. Richardson et al. (2007) NSS (UK) <i>Assessment Scale (not currently validated at this level)</i> 1) The criteria to be used in marking were clear in advance. 2) Assessment arrangements and marking were fair. 3) It was clear what standard was required in assessed work. <i>Possible exemplar item</i> Assessment allowed me to demonstrate important things I had learned in this subject work	<i>Students' Evaluation of Educational Quality</i> (SEEQ; Marsh & Roche, 1994) <i>Examinations/Grading Scale</i> 1. Feedback on examinations/graded material was valuable. 2. Methods of evaluating student work were fair and appropriate. 3. Examinations/graded materials tested class content as emphasised by the lecturer. <i>Assessment Phase 1 study items</i> 1. This teacher marks assigned work fairly. 2. The staff member clearly explained what I was required to do in assessment tasks.

	Institution-wide	Faculty/School/Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
3.1 Assessment policies address issues of pedagogy					
3.2 Adopting an evidence-based approach to assessment policies					
3.3 Alignment between institutional policy for best practice and faculty/departmental activities					

	Institution-wide	Faculty/School/Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
3.4 Commitment to formative assessment	<i>Aggregated Possible exemplar item</i> Formative assessment tasks were used in this unit to assess ongoing learning.	<i>Aggregated Possible exemplar item</i> Formative assessment tasks were used in this unit to assess ongoing learning.	<i>Possible exemplar item</i> Formative assessment tasks were used in this degree to assess ongoing learning.	<i>Possible exemplar item</i> I had the opportunity to learn from early assessment tasks in this subject.	<i>Possible exemplar item</i> My teacher used formative assessment tasks to assess ongoing learning.
3.5 Provision of specific, continuous and timely feedback	Aggregated Ramsden (1991) CEQ <i>Good Teaching Scale</i> feedback items Aggregated Kember & Leung (in press-b) <i>Feedback to Assist Learning Scale</i> Richardson et al. (2007) Aggregated NSS (UK) <i>Feedback Scale</i>	Aggregated Ramsden (1991) CEQ <i>Good Teaching Scale</i> feedback items Aggregated Kember & Leung (in press-b) <i>Feedback to Assist Learning Scale</i> Richardson et al. (2007) Aggregated NSS (UK) <i>Feedback Scale</i>	Ramsden (1991) CEQ <i>Good Teaching Scale</i> feedback items 1. The staff put a lot of time into commenting on my work. 2. The teaching staff normally gave me helpful feedback on how I was going. Kember & Leung (in press-b) <i>Feedback to Assist Learning Scale</i> 1. When I had difficulty with learning materials, I found the feedback provided by the teaching staff useful. 2. There was sufficient feedback on activities and assignments to ensure that we learnt from the work we did. 3. When I was unsure about an assignment, the teaching staff helped me to reach an understanding about how to finish it.	Richardson et al. (2007) NSS (UK) <i>Feedback Scale</i> (not currently validated at this level) 1) I received prompt feedback on my work. 2) I received detailed comments on my work. <i>Possible exemplar item</i> I received helpful feedback on my work in this subject.	<i>Feedback Phase 1 study items</i> 1. Assignments marked by this teacher have been returned within a reasonable time frame. 2. This teacher gave constructive feedback (in class, on assessment tasks, etc) that helped my learning.

	Institution-wide	Faculty/School/Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
3.5 Provision of specific, continuous and timely feedback (Continued)			Richardson et al. (2007) NSS (UK) <i>Feedback Scale</i> 1) I received prompt feedback on my work. 2) I received detailed comments on my work.		
3.6 Explicit Learning Outcomes	Aggregated Ramsden (1991) CEQ <i>Clear Goals and Standards Scale</i> Aggregated Richardson et al. (2007) NSS (UK) <i>Assessment Scale</i>	Aggregated Ramsden (1991) CEQ <i>Clear Goals and Standards Scale</i> Aggregated Richardson et al. (2007) NSS (UK) <i>Assessment Scale</i>	Ramsden (1991) CEQ <i>Clear Goals and Standards Scale</i> 1) It was always easy to know the standard of work expected. 2) I usually had a clear idea of where I was going and what was expected of me in this course. 3) It was often hard to discover what was expected of me in this course. (reversed) 4) The staff made it clear right from the start what they expected from students. Richardson et al. (2007) NSS (UK) <i>Assessment Scale</i> 1) The criteria to be used in marking were clear in advance. 2) Assessment arrangements and marking were fair. 3) It was clear what standard was required in assessed work.	Richardson et al. (2007) NSS (UK) <i>Assessment Scale (not currently validated at this level)</i> 1) The criteria to be used in marking were clear in advance. 2) Assessment arrangements and marking were fair. 3) It was clear what standard was required in assessed work <i>Possible exemplar items</i> Assessment clearly focussed on what we were learning. This subject's intended learning outcomes were clear to me. The expected standards for assessed work were clear to me.	<i>Assessment Phase 1 study items</i> The lecturer made clear the standard of the work expected. The staff member clearly explained what I was required to do in assessment tasks.

	Institution-wide	Faculty/School/Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
3.7 Monitoring and review of standards and assessment tasks	<i>Aggregated Possible exemplar item</i> Across my degree, learning outcomes and standards of assessment were made clear.	<i>Aggregated Possible exemplar item</i> Across my degree, learning outcomes and standards of assessment were made clear.	<i>Possible exemplar item</i> Across my degree, learning outcomes and standards of assessment were made clear.	<i>Possible exemplar item</i> Assessment standards were set appropriately.	<i>Possible exemplar item</i> The lecturer clearly explained the learning outcomes and expected standards of this unit of study.

	Institution-wide	Faculty/School/Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
4.1 Student Engagement	Aggregated Coates (2006) <i>AUSSE Active Learning Scale</i> Aggregated Coates (2006) <i>AUSSE Academic Challenge Scale</i> Aggregated Coates (2006) <i>AUSSE Student and Staff Interactions Scale</i>	Aggregated Coates (2006) <i>AUSSE Active Learning Scale</i> Aggregated Coates (2006) <i>AUSSE Academic Challenge Scale</i> Aggregated Coates (2006) <i>AUSSE Student and Staff Interactions Scale</i>	Coates (2006) <i>AUSSE Active Learning Scale</i> 1. I set high performance standards for myself. 2. I tried to make connections between things I was learning. 3. I pushed myself to understand things I found puzzling. 4. I sought out my own resources to help me understand topics. 5. I thought about the practical applications of materials that I studied. 6. I thought about the ethical issues related to the material that I studied. Coates (2006) <i>AUSSE Academic Challenge Scale</i> 1. Assessment tasks challenged me to learn. 2. I was given enough material to keep up my interest. 3. I was encouraged by teachers to go beyond set materials. 4. Academic staff gave me comments on my work that helped me learn. 5. I received feedback quickly enough to improve subsequent work.	<i>Students' Evaluation of Educational Quality</i> (SEEQ; Marsh & Roche, 1994) <i>Group Interaction Scale</i> 1. Students were encouraged to participate in class discussions. 2. Students were invited to share their ideas and knowledge. 3. Students were encouraged to ask questions and were given meaningful answers. 4. Students were encouraged to express their own ideas and/or question the lecturer. Coates (2006) <i>AUSSE Scales</i> (not validated at this level) - <i>Active Learning Scale</i> , <i>Academic Challenge Scale</i> , <i>Student and Staff Interactions Scale</i> , <i>Enriching Educational Experiences Scale</i> <i>Possible exemplar item</i> I make the best use of the learning opportunities in this subject.	<i>Students' Evaluation of Educational Quality</i> (SEEQ; Marsh & Roche, 1994) <i>Group Interaction Scale</i> 1. Students were encouraged to participate in class discussions. 2. Students were invited to share their ideas and knowledge. 3. Students were encouraged to ask questions and were given meaningful answers. 4. Students were encouraged to express their own ideas and/or question the lecturer.

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
4.1 Student Engagement (Continued)			Coates (2006) AUSSE <i>Student and Staff Interactions Scale</i> 1. I had one-to-one conversations with academic staff. 2. I initiated individual contact with a member of academic staff. 3. I sought advice from staff on how to improve my performance. 4. I developed a valuable rapport with a member of academic staff. 5. I met with academic staff outside class.		
4.2 Fostering and facilitating (academic) learning communities				<i>Possible exemplar item</i> Teaching in this subject helped establish a learning community	

	Institution-wide	Faculty/School/ Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
4.3 Engaging and identifying with a learning community	Aggregated Griffin et al. (2003) Extended CEQ <i>Learning Community Scale</i> Coates (2006) <i>AUSSE Supportive Learning Environment</i> 1. Staff respected students' backgrounds, perspectives and needs. 2. The university campus felt like a supportive place to learn. 3. Staff seemed responsive to feedback from students. 4. I felt part of an academic community at university.	Aggregated Griffin et al. (2003) Extended CEQ <i>Learning Community Scale</i> Disaggregated Coates (2006) <i>AUSSE Supportive Learning Environment</i>	Griffin et al. (2003) Extended CEQ <i>Learning Community Scale</i> 1. I felt part of a group of students and staff committed to learning. 2. I was able to explore academic interests with staff and students. 3. I learned to explore ideas confidently with other people. 4. Students' ideas and suggestions were used during the course. 5. I felt I belonged to the university community.	Griffin et al. (2003) Extended CEQ <i>Learning Community Scale</i> (not validated at this level) 1. I felt part of a group of students and staff committed to learning. 2. I was able to explore academic interests with staff and students. 3. I learned to explore ideas confidently with other people. 4. Students' ideas and suggestions were used during the course. 5. I felt I belonged to the university community. Summers et al. (2005) <i>Classroom Community Scale</i> 1. I feel connected to people in this class. 2. I've made friends in this class. 3. I feel I fit into this class. 4. I know other people well in this class. <i>Possible exemplar item</i> I felt part of a group of students and staff committed to learning	

	Institution-wide	Faculty/School/Department	Degree/Program me	Unit of Study/Subject	Teacher/Individual
4.3 Engaging and identifying with a learning community			Disaggregated Coates (2006) AUSSE <i>Supportive Learning Environment</i>		
4.4 Staff engagement	Aggregated SOHQ, Prosser & Trigwell, or HERI scales	<i>Student Orientation Scale</i> (SOHQ; Hart et al., 2000 – survey of teachers); scales on Curriculum Coordination; Excessive Work Demands; Goal Congruence; Participative Decision-Making; Professional Growth; Professional Interaction; Student Orientation; and Supportive Leadership. (The SOHQ is a proprietary instrument.) <i>Perceptions of the Teaching Environment</i> (Prosser & Trigwell, 1997) Transformational Leadership 1. The Head inspires respect for his/her own ability as a teacher. 2. The Head of this Department enables you to think about old problems in new ways. 3. The Head motivates you to do more in your teaching than you ever thought you could. 4. The Head gives me confidence as a teacher. Teacher Involvement 1. Academic staff spend a good deal of time talking to each other about		<i>Possible exemplar item</i> Staff in this subject seemed really engaged in their teaching	

	Institution-wide	Faculty/School/Department	Degree/ Programme	Unit of Study/Subject	Teacher/Individual
4.4 Staff engagement (Continued)		their teaching. 2. People here don't communicate much. (reversed). 3. There isn't much discussion about educational issues among the staff here. (reversed). 4. Academic staff here are keen to learn from each other. 5. Staff here discuss their teaching problems with each other. Clear Goals & Contingent Reward 1. My Head of Department gives a lot of praise for good teaching. 2. You usually have a good idea of what's expected of you as a teacher here. 3. The Head readily acknowledges your contributions to teaching. 4. The Head makes it clear what is required from teaching staff. Collaborative Management 1. In this department staff are often consulted on matters of policy. 2. You feel free to express your own point of view in this department. 3. The Head of Department listens to what you have to say. 4. The Head delegates responsibility fairly and consistently.			

	Institution-wide	Faculty/School/Department	Degree/ Programme	Unit of Study/Subject	Teacher/Individual
4.4 Staff engagement (Continued)		Teacher Workload 1. There is ample time to give students proper comments on the quality of their work. 2. Large classes here mean students don't get the attention they deserve. (reversed). 3. Academic staff here just don't have enough time to help students properly. (reversed). 4. It is difficult to really assist students in this subject in the time we have available for teaching. (reversed). Class Size 1. In large classes I give students less encouragement to see me. (reversed). 2. In large classes I try to avoid questions from students. (reversed). 3. Large classes discourage contact between me and students. (reversed). Teaching Control 1. I have very little say in the way this subject is run. (reversed). 2. The Department allows considerable flexibility in the way I teach this subject. 3. I feel a lack of control over what I teach in this subject. (reversed). 4. I am under increasing pressure from students to stick to the syllabus. (reversed)			

	Institution-wide	Faculty/School/Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
4.4 Staff engagement (Continued)		Student Characteristics 1. Having a range of student talent makes it difficult to appropriately direct my teaching. (reversed). 2. Students act as though I am a school teacher, not someone who assists adult learning. (reversed). 3. Students have such variable skills that I find it hard to predict what they know. (reversed). Commitment to student learning 1. You get the feeling in this department that the staff really care for students. 2. Teaching students can be a bit of a chore for most people. (reversed). 3. Staff here make a real effort to understand difficulties students may have with work. 4. Staff here take account of students' views when planning changes to subjects. HERI (Higher Education Research Institute) Faculty Survey (see http://www.gseis.ucla.edu/heri/facoverview.php)			

	Institution-wide	Faculty/School/Department	Degree/Programme	Unit of Study/Subject	Teacher/Individual
4.4 Staff engagement (Continued)		Includes: Time Diary for faculty work and life; Faculty productivity; Perceptions of institutional priorities; Perceptions of institutional climate; Teaching and learning outcomes for undergraduate education; Evaluation methods; Instructional techniques; Goals and outcomes for general education; Engagement with undergraduates in Research, Teaching, Service, Diversity issues, Integration of diversity in the curriculum, Integration of diversity in research, Institutional climate for women, minorities and gay/lesbian faculty, Institutional diversity priorities. Faculty job satisfaction; Sources of faculty stress, plans to leave, or potential for faculty job turnover; Faculty goals and attitudes that match items on CIRP student surveys; Faculty salary and satisfaction.			

8. KEY ISSUES FOR PILOT UNIVERSITIES SEEKING TO USE SET DATA IN RELATION TO THE TEACHING QUALITY INDICATOR DIMENSIONS.

Using existing SET data

This report concluded that existing *internal* university SET data does not provide particularly useful, sector-wide, quantitative benchmarking data at all the organisational levels covered by the TQI framework. This is largely because there is little overlap in the SET items used by different universities, particularly at the level of individual teachers and subjects.

Where overlap in SET items did exist it was primarily at the level of degrees/courses, largely through the historical use of the CEQ and the development of internal versions of this survey for use with current students. If, in the future, new SET surveys are adopted nationally by government (eg the AUSSE), then there might be additional commonality at this level as institutions build items from such surveys into their internal surveys.

At other organisational levels there was more overlap in 'overall' or comprehensive SET items (e.g. 'The assessment supported my learning in this subject') than there was in SET items that reflected specific sub-elements of dimensions of teaching (e.g. sub-elements of assessment; 'the assessment was marked fairly', 'the assessment was relevant to the learning outcomes').

However while there was considerable variability in the actual SET items there was some commonality in terms of the dimensions these various SET items reflected. That is, many universities ask about the same dimensions of teaching using different questions. This overlap offers the potential to compare data in different ways. The comparison cannot be based on a simple comparison of numerical data, but rather the numerical data can be used as the starting point for a conversation. The comparison may then be made in terms of a qualitative comparison of the interpreted SET data, i.e. the nature of the issues identified in the SET data and the responses to these is the basis for comparison – not the raw numbers themselves. However it should be noted that this is quite different to the Australian higher education sector's current approach to benchmarking which focuses on statistical comparisons of numerical scores on SET surveys.

Developing a new generation of SET data

As current internal SET data is limited in its ability to support meaningful between-institution dialogue based on comparisons of numerical scores, this report suggests either; (1) the adoption of a different approach to benchmarking; one that focuses on comparisons arising from conversations to interpret SET data than a simple comparison of SET data itself, or (2) the development of new SET items and scales for shared use



between benchmarking partners. In the context of current uses of performance indicators a combination of these approaches might be most appropriate. This could involve the development of some shared items or scales for the TQI dimensions, which would also support dialogue and conversations based on idiosyncratic SET items related to common dimensions. This could entail the collaborative revision and piloting of some existing items by participating universities, and agreement to incorporate new scales and items into existing SET surveys. Either approach should also involve a consideration of how to strategically extend beyond a reliance on SET data in relation to the full range of TQI dimensions, as well as a consideration of how to facilitate the sharing of data and the establishment of productive benchmarking dialogues between partner institutions.

To support such collaborations this report has provided some initial suggestions for items and scales drawn from existing SET surveys in relation to the new Teaching Quality Indicator dimensions. Where possible these scales and items have been drawn from SET instruments with published validity and reliability data. Where this is available a reference has been provided in the table listing these suggestions. However these items and scales have typically been validated at one level of use only (e.g. at the course level) and re-validation may be needed if these are to be adapted for use in a different context.

It is possible that data gathered at lower levels can usefully be aggregated to provide a picture of teaching quality at higher levels, for instance, aggregating subject level data to provide a proxy for department/course level. However there are some caveats suggested in terms of the statistical processes used and it should also be remembered that course quality, is usually 'more than the sum of the parts' as different issues (e.g. integration between subjects) become more or less relevant at different organisational levels.

Where no validated scale or item was identified, a potential item has been suggested from the data gathered in this study. These items may provide a basis for development of suitable shared items by the participating universities. While such items may have a degree of face validity in relation to the dimension, published data on their reliability or validity is not available. Reliability and validity work should therefore be undertaken by universities seeking to use such items to generate TQI data. Some suggestions are offered below for how such items or scales might be validated.

A framework for scale & item development and testing

Barrie and Ginns (2006; Ginns & Barrie, under review) provide a general framework for the development of student-focused SET questionnaires. We outline this framework as a potential starting point for collaboration between benchmarking institutions in the development and testing of the scales and items suggested below.

1. Identify relevant SET research literature

The scales and items suggested above draw on a wide range of theoretical frameworks and scale/item developmental research, including Student Learning Theory (e.g. Ramsden, 1991); Student Engagement Theory (Coates, 2006); and broadly based investigations of the multiple dimensions underlying effective teaching (e.g. Abrami et al., 2007; Marsh & Roche, 1994).

2. Develop items with a focus on the student's experience of the teacher

Many teaching evaluation instruments are worded in ways that focus on judgments of the teacher, rather than the student's experience of the teaching, and how that experience related to student learning. A core assumption of the Student Approaches to Learning paradigm (e.g. Prosser & Trigwell, 1999) is that it is students' perceptions of the teaching environment which affect their approaches to learning, rather than an objective appraisal of that environment. Where possible, then, items are worded in the first person, and emphasise the students' personal experience of teaching. The use of Likert scale values of 'strongly disagree' to 'strongly agree', rather than alternative rating schemes like 'poor' to 'excellent', is similarly driven by the desire to emphasise a student perception focus rather than an appraisal focus. Where possible, institutions may consider designing surveys with space for comment on each item, rather than simply a space for overall comments. Using this method allows staff to interpret numerical responses and "triangulate" quantitative results for each item with qualitative feedback. This avoids the common problem with numerical ratings faced by teachers, of why students give high, middling or low ratings with respect to a given facet of teaching. For an extended discussion of teaching evaluation from a student-focussed perspective, see Ramsden and Dodds (1989).

3. Iterative peer review and consultation with stakeholders

Following initial item development, discussions are held with key individuals, university committees and working groups with expertise in quality assurance and/or specific teaching contexts. These discussions cover the breadth of the item coverage, and the "face validity" of the items; that is, the extent to which an item is clearly recognisable as a meaningful indicator of teaching quality for a particular context. In a similar way, students are asked for their comments on the survey.

Hinkin and Tracey (1999) developed a method for assessing the adequacy of the content of multi-item scales which is also recommended. Using this method, key individuals, university committees and working groups with expertise in quality assurance (roughly 30 to 50) rate the consistency of proposed items against their expected dimensions using a rating scale of 1 (not at all) to 5 (completely). The ratings can then be analysed using analysis of variance techniques, reducing potential biases which might be associated with the exclusive use of subjective judgments to make item inclusion decisions. For a recent example of the use of this technique in organisational research, see Tracey and Tews (2005).

An additional qualitative research method for testing the suitability of SET scales and items with students is the cognitive interview (Willis, 2005). Cognitive interviews allow problems in the thought processes required to make a response to an item to be identified using either concurrent or retrospective think-aloud protocols. For an example of the application of this method to refine SET items in an Australian university, see McCormack (2007).

4. Pilot testing of surveys and statistical validation

In this stage, a draft version of scales and items is piloted across benchmarking institutions. The overall goal of this process is to ensure that the selected scales and items provide useful and meaningful information to teaching staff. Following the above process, the surveys are made available to all teaching staff. Psychometric analyses of the surveys (e.g. determining the factor structure of instruments through exploratory and confirmatory factor analyses; determining the consistency of responses among items and between students using internal consistency and inter-rater reliability analyses) are conducted. For excellent introductions to the process of developing self-report scales, see deVellis (1991) and Netemeyer, Bearden, & Sharma (2003).

The reliability of self-report scales is typically measured by the extent of agreement between related items (e.g. internal consistency measures like coefficient alpha; Cronbach, 1951). However, as Marsh (2007; p.333) notes,

“...this approach, while potentially useful, does not provide an adequate basis for assessing the reliability of SET responses. The main source of variability is lack of agreement among different students’ ratings of the same teacher rather than lack of agreement among different items. Hence, the reliability of SETs is most appropriately determined from studies of interrater agreement that assess the lack of agreement among students within the same course”.

Inter-rater reliability analyses can be conducted using analysis of variance, where the subject of analysis (e.g. teacher, unit of study) is used as the “level” (i.e. independent variable). An estimate of the inter-rater reliability (IRR) can then be derived using the F value ($IRR = [F-1]/F$; Gillmore, Kane, & Naccarato, 1978; Winer, Brown, & Michels, 1991). The inter-rater reliability of a teacher- or class-average depends on the number of raters; Marsh (2007) gives estimates of typical IRRs with different numbers of raters as follows:

- .95 for 50 raters
- .90 for 25 raters
- .74 for 10 raters
- .60 for 5 raters.

More accurate estimates of IRR with different numbers of raters can be derived from the collected data using an on-line calculator developed by Solomon (2004); see <http://www.med-ed-online.org/rating/reliability.html>

Because of the “nested” nature of such teaching evaluation data, in which students are taught by individual teaching staff, care must be taken to account for this nestedness in subsequent analyses - using standard General Linear Model statistics on individual-level data will mean that parameter estimates may be biased (Rowe, 2003). Alternative analytic methods for testing psychometric properties of teaching evaluation instruments include conducting analyses on class-

average responses (e.g. Marsh, 1991), or conducting multi-level confirmatory factor analysis, which allows investigation of the factor structure of an teaching evaluation instrument at the individual- and class-level simultaneously (for a recent example of this approach, see Toland & de Ayala, 2005). **Guidelines for aggregating SET data**

Some validated or suggested scales and items appear to be appropriate for use at different levels of analysis. This requires results for a given scale or item collected at one level (e.g. degree/programme) to be aggregated up to a higher level (e.g. multiple degrees administered by the same school, department or faculty).

This is a common practice in higher education quality assurance of teaching (e.g. the aggregation of CEQ results into broad fields of study); however, it is important to provide evidence that suitable levels of *inter-rater agreement* exist across multiple lower-level units of measurement before aggregating results to a higher level (Chan, 1998). An example of this type of analysis is given in Ginns, Prosser, & Barrie (2007), who investigated inter-rater agreement for Student Course Experience Questionnaire degree-level ratings at the faculty level, using the *Average Deviation* (AD) metric (Dunlap, Burke, and Smith-Crowe, 2003). Similarly, Barrie and Ginns (2007) provided evidence of suitable levels of inter-rater agreement for both SCEQ faculty-level aggregated results, and faculty-level aggregated Unit of Study Evaluation (USE) results, before investigating correlations between faculty scores on these two instruments.

Currently, there is limited available software for conducting inter-rater agreement analyses. A free programme developed by Dunlap et al. (2003) is suitable for calculating AD on single Likert-type items (see <http://www.tulane.edu/~dunlap/psylib.html>, programme agree.exe), and SAS and R macros are available for multiple-item inter-rater agreement analyses (Cohen, Doveh, & Nahum-Shani, in press). It is also expected that Professor Jim Jaccard will develop SPSS- and Excel-compatible software (www.zumastat.com) for multiple Likert-type item SET scale inter-rater agreement in the near future (personal communication, February 2, 2008).

Developing new SET processes

The report identified that there was considerable variability in the SET processes institutions have in place to gather, analyse, share and respond to SET data internally. This report describes some of the variations in relation to how data is gathered, used and aggregated, how standards are established and comparisons made and what sort of access different members of the (university) community have to different levels of SET data.

For universities seeking to use SET data for cross institutional benchmarking, the report has suggested that common procedures should be established for all these aspects of the process. Once developed such procedures should be noted or referred to in the memorandum of understanding which describes the benchmarking relationship. Some preliminary suggestions are provided in terms of what such a memorandum of understanding might cover in relation to SET.

As a precursor to sharing institutional SET data for collaborative benchmarking activities the following preliminary considerations were identified in discussions with workshop participants:

- 1 Institutional agreement and commitment to share data should be discussed and formalised in a memorandum of understanding between senior university officers. This agreement should include;
 - o A clear statement of the intent of the activity. Is it for institutional learning, quality assurance or marketing purposes etc.
 - o An indication of how the shared data and any subsequent analysis of the data will be used. For instance will there be any publication of comparative analyses? What approval or consultation process will be undertaken prior to any publication? Will partners agree to joint public reporting or internal reporting only?
 - o Clarification of the level of information to be associated with the data and the format in which it will be shared. For instance will SET data be provided as raw data suitable for statistical analyses or as aggregated reported data? Will it be identified by field of study; degree course; subject name; year of study; background of respondents (i.e. international/local student etc).
 - o How long the agreement will be in place.
2. A communication process should be established to inform members of the participating university communities (staff and students) as to what SET data will be used and how it will be used. This was seen to be particularly important for sharing SET data below that of the 'course-level' as concerns might exist that such data could be associated with particular individuals (staff or students) unless suitably de-identified. Institutions should also consider and compare their existing ethics clearances in relation to the use of SET data.
3. Processes should be established and responsibilities allocated for acting upon such data once shared. These processes should include reporting back to the university community on what has occurred in response to any issues identified. There was a strong suggestion from some participants that such processes should be based at the local (i.e. discipline/department level rather than with a central quality assurance unit.

Suggestions for appropriate alternative and complementary data sources to be considered

In mapping the potential available SET scales to the Teaching Quality Dimensions it was noted that traditional quantitative SET data did not provide a full picture for any of the dimensions. The rich variety of available data in relation to the proposed dimensions is described in the main report (Chalmers 2007). In some cases participants at the SET experts workshop noted that no form of SET data was relevant to particular dimensions (see annotations in summary table of potential SET scales). In other cases they noted the usefulness of collecting complementary data to support the interpretation of quantitative SET data. In particular participants noted the importance of utilising qualitative student evaluation of teaching data gathered in open response comments and student group interviews.

While noting the importance of such qualitative SET data it was also noted that processes to analyse and share such data were currently relatively unsophisticated in relation to those used for comparative analyses of numerical SET data. There were however some notable exceptions where analyses of qualitative data (CEQ data in particular) had provided valuable insights into the range of students' experiences of teaching and learning. The report "Accessing the student voice" (Scott 2006) is a case in point. Several institutional contacts reported internal university processes for using the results of analyses of such qualitative data to support quality enhancement activities. There were no institutional reports of such data being used in internal teaching quality indicator schemes. The use by participating institutions of shared consistent qualitative data analysis procedures was seen by some participants to be a precursor to using such data to compare institutions.

It was also apparent that there were some notable absences in the survey data currently available. While some workshop participants reported including student effort and motivation items in SET surveys to generate performance indicator data related to student engagement, there were no reports of comparable indicators of staff effort, motivation or engagement. Surveys of staff perceptions of teaching and learning experiences, and their conceptions of teaching were notably absent from the work of those running institutional teaching survey systems. However, such perceptions were perceived by workshop participants to be highly relevant to a consideration of the quality of the learning environment, especially in relation to some of the Teaching Quality Indicators. There is also strong support for the impact of such factors on the quality of student learning in the literature (see Trigwell et al 1999; Prosser et al 2003).

Workshop participants also suggested the use of student learning outcome data derived from assessment as a useful complement to current SET data. Meta-analytic studies of the multisection validity SET research literature (eg. Feldman 1989) have found moderate to strong correlations between overall teacher or course ratings and student achievement, and moderate correlations between some specific dimensions (eg teacher preparation) and student outcomes. While this is true statistically, in a particular class the relationship between a student's perception of a high quality learning experience and the quality of learning outcomes indicated by assessment results is of course mediated



by the nature of the assessment task, criteria and standards and the assumption that better performance relates to 'better learning' rather than, (for instance) 'generous marking'.

Suggestions for fostering a new approach to benchmarking using SET data

Universities' current approaches to using SET data for benchmarking tend to focus on a comparison of SET scores and rarely extend to exploring the reasons underlying any observed differences. However, contemporary perspectives on benchmarking emphasise the centrality of organisational learning from quality assurance in order to inform quality enhancement. This study suggested two ways forwards in supporting Australian universities in making better use of internal SET data for benchmarking. The first involved developing SET shared data sets to enable comparisons and the second involved using different SET data sets but with a shared focus. This second approach, while more challenging for institutions has much to recommend it. The key benefit is that because the numerical data generated by different SET items is *not* amenable to simple comparison it requires interpretation before any comparison can be made. This interpretation requires *real* engagement by staff with the SET data. By necessity it involves members of the university in understanding *why* their students have rated aspects of their experiences in a particular way. In this approach, benchmarking would then involve the members of one institution sharing their *understanding* of why students had experienced an aspect of teaching and learning in a particular way, with colleagues at another institution interested in the same dimension of teaching. This exchange would focus on participants learning more about why students might experience teaching and learning in particular ways and sharing strategies for how these experiences might be enhanced.

The adoption of such an approach to benchmarking would require:

- o Support for participating institutional representatives in identifying appropriate SET data (and useful complementary data) relevant to the selected TQI dimensions.
- o Support for participating universities' in developing strategies and mechanisms to enable the interpretation of SET data in relation to the TQI dimensions.
- o Development of mechanisms to enable conversations between institutions in relation their respective interpretations of SET data on a shared TQI dimensions.
- o Facilitation of learning by participants from these conversations in relation to planning improvements in their students' experiences of teaching and learning.
- o Development of strategies to document and record the outcomes of the process as evidence of successful benchmarking.



9. CONCLUDING COMMENTS & SUGGESTIONS: SUSTAINING CHANGE IN SET

This study was not aimed at considering best practice in universities' internal SET systems. However, in the course of undertaking this study it was apparent that many universities possess innovative, systematic and theoretically rigorous SET systems which are effectively supporting internal processes of change. In some cases these internal institutional change processes and SET indicators articulated in a coherent way with higher level external SET based measures of change. However this coherence rarely extended throughout all levels of an institution (eg to department, subject, and individual level) and where it did, the manner in which different institutional levels were related varied. The result was that the SET systems at different universities (especially at the level of subject and teacher) don't always ask about the same things (in terms of the empirically derived dimensions of teaching used in this analysis) and certainly not in the same way, in terms of how the SET items are worded or scored. This makes cross institutional data comparison difficult.

		External SET indicators	
	University A		University B
Institution wide	SET data		SET data
Faculty	SET data		SET data
Department	SET data		SET data
Subject	SET data		SET data
Individual	SET data		SET data

As a consequence, current internal university SET systems, while they may support individual institutions in changing practice relative to external measures, seem poorly suited to the task of supporting outwards-looking institutions seeking to cooperate with each other in *improving* teaching and learning. Such improvement is a focus of the proposed Teaching Quality Indicators framework.

This study was focused on the extent to which SET systems might support a relatively new activity; that of cross-institutional benchmarking using internal SET data.

In considering the potential for current SET systems to support cross institutional benchmarking at the organisation levels where significant teaching and learning change typically occurs (department, subject or individual), there was little or no commonality between internal SET survey data. As such, the current SET systems do little to encourage comparisons and collaborations between discipline groups or similar organisational groups or between 'like' subjects taught in different universities.

While some universities and discipline networks (eg Deans of Economics & Business) do already exchange SET data for benchmarking purposes, the level of this exchange is



limited and typically involves national graduate data collected using the CEQ or internal SET data in relation to whole-degrees / courses collected using variants of established national surveys. In the context of the proposed new TQI dimensions such data is limited both in terms of the organisational level it relates to (degree course) and the dimensions of the teaching and learning experience it relates to.

Opportunities for institutions and networks to engage in more productive benchmarking activities will be provided during the piloting of the Teaching Quality Indicators framework in 2008. This report has suggested some ways in which current SET systems might be developed to better support the sorts of quality enhancement systems envisaged in the proposed Teaching Quality Framework. However such development will require a different approach to the development of Student Evaluation of Teaching systems than has typically occurred in the past. These systems now need to consider the wider context of teaching and learning change in terms of:

- What is known from the theoretical and empirical research about university teaching and learning
- What is known about how university communities manage and engage in change related to teaching and learning
- What is known about the changing societal and knowledge contexts which universities operate in

Any undertaking to develop SET systems for the new purpose of cross-institutional benchmarking for learning will need to be a collaborative undertaking. It will need to be a collaboration *within* universities that involves people with different expertise, perspectives, agendas and powers in the institution. Perhaps most importantly it will involve collaborations by these people *across* universities.

In laying the groundwork for such collaborations many of the SET experts at the national 'Evaluation Forum' workshop in November 2007, who range of roles encompassed, Pro Vice-Chancellor, academics and administrative officers, declared their interest in working together to develop their existing systems to better support the proposed new TQI Framework. Such ongoing collaboration could usefully support and inform the pilot of the proposed TQI Framework.

10. REFERENCES

Abrami, P.C., d'Apollonia, S., & Rosenfield, S. (2007). The dimensionality of student ratings of instruction: What we know and what we do not. In R.P. Perry and J.C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence based perspective*. Dordecht: Springer.

Abrami, P.C., Rosenfield, S., & Dedic, H. (2007). The dimensionality of student ratings of instruction: An update on what we know, do not know, and need to do. In R.P. Perry and J.C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence based perspective*. Dordecht: Springer.

Australian Council for Educational Research (2007). The idea of student engagement. Downloaded 16 October 2007 from <http://www.acer.edu.au/ausse/engagement.html>

Barrie S.C., & Ginns P. (2006). A research based approach to developing student feedback surveys. Paper presented at HERDSA Perth, July 10–12, 2006.

Barrie, S., & Ginns, P. (2007). The linking of institutional performance indicators to improvements in teaching and learning in classrooms. *Quality in Higher Education*, 13, 275 – 286.

Chalmers, D. (2007). *An agenda for teaching and learning in universities (Version 3)*. Sydney: The Carrick Institute for Learning and Teaching in Higher Education.

Chan, D. (1998) Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, 83, 234-246.

Coates, H. (2006). *Student engagement in campus-based and online education: University connections*. London: Routledge.

Cohen, A., Doveh, E., & Nahum-Shani, I. (in press). Testing agreement for multi-item scales with the indices $r_{WG(J)}$ and $AD_{M(J)}$. *Organizational Research Methods*.

Cooperative Institutional Research Program (CIRP) (2007). Downloaded 16 October 2007, from <http://www.gseis.ucla.edu/heri/cirpoverview.php>

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

Davies, M., Hirschberg, J., Johnstone, C., & Lye, J. (2007). An Exploration of the Relationship between the Quality of Teaching Survey and the CEQ. Paper presented at



the 2007 Quantitative Analysis of Teaching and Learning in Business, Economics and Commerce Forum, The University of Melbourne.

deVellis, R.F. (1991). *Scale development: Theory and applications*. Newbury Park: Sage.

Dunlap, W. P., Burke, M. J., & Smith-Crowe, K. (2003). Accurate tests of statistical significance for r_{wg} and AD interrater agreement indices. *Journal of Applied Psychology*, 88, 356-362.

Feldman, K.A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, 30, 583-645.

Gillmore, G. M., Kane, M. T., & Naccarato, R. W. (1978). The generalizability of student ratings of instruction: estimation of the teacher and course components. *Journal of Educational Measurement*, 14, 1-21.

Ginns, P., & Barrie, S. (under review). Developing and testing a student-focussed teaching evaluation instrument for university instructors.

Ginns, P., & Barrie, S.C. (2004). Reliability of single-item ratings of quality in higher education: A replication. *Psychological Reports*, 95, 1023-1030.

Ginns, P., Prosser, M., & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. *Studies in Higher Education*, 32, 603-615.

Graduate Careers Australia (2006). Enhancing the GCA national surveys: An examination of critical factors leading to enhancements in the instrument, methodology, and process. DEST: Commonwealth of Australia. Downloaded 29 March 2007, from http://www.dest.gov.au/sectors/higher_education/publications_resources/profiles/enhancing_gca_national_surveys.htm

Griffin P., Coates H., McInnis, C., & James R. (2003). The development of an extended Course Experience Questionnaire. *Quality in Higher Education*, 9, 259-266.

Hart, P.M., Wearing, A.J., Conn, M., Carter, N.L., & Dingle, R.K. (2000). Development of the School Organisational Health Questionnaire: A measure for assessing teacher morale and school organisational climate. *British Journal of Educational Psychology*, 70, 211-228.

Hinkin, T.R., & Tracey, J.B. (1999). An analysis of variance approach to content validation. *Organizational Research Methods*, 2, 175-186.

Huang, S.L. (2006). Validation of an instrument for assessing psychological environments at colleges and universities in Taiwan. In D.L. Fisher and M.S. Khine (Eds.), *Contemporary Approaches to Research on Learning Environments: World Views*. Singapore: World Scientific.



Huang, S.L. (in review). The university environment and its relation to student academic aspirations and satisfaction.

Kabanoff, B., Richardson, A., & Brown, S. (2003). Business graduates' perceptions of the quality of their course: A view from their workplace. *Journal of Institutional Research*, 12, 1-12.

Kember, D., & Leung, D.Y.P. (in press-a). Establishing the validity and reliability of course evaluation questionnaires. *Assessment and Evaluation in Higher Education*.

Kember, D., & Leung, D.Y.P. (in press-b). Development of a questionnaire for assessing students' perceptions of the teaching and learning environment and its use in quality assurance. *Learning Environments Research*.

Krause, K-L., Hartley, R., James, R., & McInnis, C. (2005). The First Year Experience in Australian Universities: Findings from a Decade of National Studies. CSHE: University of Melbourne. Downloaded 29 March 2007, from <http://www.cshe.unimelb.edu.au/pdfs/FYEReport05KLK.pdf>

Kuh, G. D. (2001). The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties. Indiana University Center for Postsecondary Research. Bloomington, IN. Downloaded 30 March 2007, from http://nsse.iub.edu/html/psychometric_framework_2002.cfm

Laird, T.F.N., Shoup, R., and Kuh, G.D. (2005). Deep learning and college outcomes: Do fields of study matter? Paper presented at the Annual Meeting of the Association for Institutional Research, May 29 – June 1, 2005, San Diego, CA. Downloaded 16 October 2007, from nsse.iub.edu/pdf/conference_presentations/2005/AIR2005DeepLearning-PAPER.pdf

Marsh, H.W. (1991). Multidimensional students' evaluations of teaching effectiveness: A test of alternative higher-order structures. *Journal of Educational Psychology*, 83, 285-296.

Marsh, H.W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R.P. Perry and J.C. Smart (eds.), *The scholarship of teaching and learning in higher education: An evidence based perspective*. Dordrecht: Springer.

Marsh, H., & Roche, L. (1994). *The use of students' evaluations of university teaching to improve teaching effectiveness*. Canberra: Australian Government Publishing Service.

McCormack, C. (2007). What not to ask: The use of cognitive interviews to refine evaluation questions. Paper presented at the 5th Evaluation Forum, University of Queensland, 27-28 November 2007.



National Survey of Student Engagement (2007a). About NSSE. Downloaded 30 March, 2007, from http://nsse.iub.edu/html/quick_facts.cfm

National Survey of Student Engagement (2007b). What is BCSSE?. Downloaded 30 March, 2007, from <http://bcsse.iub.edu/about.cfm>

National Survey of Student Engagement (2007c). Publications and Presentations. Downloaded 30 March, 2007, from <http://nsse.iub.edu/html/pubs.cfm?viewwhat=Research%20Paper>

Netemeyer, R.G., Bearden, W.O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage.

Prosser, M., & Trigwell, K. (1997). Relations between perceptions of the teaching environment and approaches to teaching. *British Journal of Educational Psychology*, 67, 25-35.

Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education*. Buckingham: SRHE & Open University Press.

Prosser, M., Ramsden, P., Trigwell, K. & Martin, E. (2003). Dissonance in experience of teaching and its relation to the quality of student learning. *Studies in Higher Education*, 28:1, 37-48.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education*, 16, 129-50.

Ramsden, P., & Dodds, A. (1989). *Improving teaching and courses: A guide to evaluation*. Parkville: University of Melbourne.

Richardson, J.T.E., Slater, J.B., & Wilson, J. (2007). The National Student Survey: Development, findings, and implications. *Studies in Higher Education*, 32, 557-580.

Rowe, K. (2003). *Practical multilevel analysis with MLwiN & LISREL: An integrated course* (Revised edition). 19th ACSPRI Summer Program in Social Research Methods and Research Technology, Australian National University. Camberwell, VIC, Australia: Australian Council for Educational Research.

Scott, G. (2006). *Accessing the Student Voice: Using CEQuery to identify what retains students and promotes engagement in productive learning in Australian higher education*. Canberra: DEST. Downloaded 12 March, 2008, from http://www.dest.gov.au/sectors/higher_education/publications_resources/profiles/access_student_voice.htm

Sharpe, A. (2007). *Comparative review of British, American and Australian national surveys of undergraduate students*. York: Higher Education Academy. Downloaded 30 March 2007, from



http://www.heacademy.ac.uk/documents/National_Survey_Comparative_Review_Feb_2007.doc

Solomon, D. (2004). The rating reliability calculator. *BMC Medical Research Methodology*, 4, 11. See <http://www.med-ed-online.org/rating/reliability.html>

Swail, W.S., Redd, K., & Perna, L. (2003). *Retaining minority students in higher education*. San Francisco, CA: Jossey-Bass.

Summers, J.J., Beretvas, S.N., Svinicki, M.D., & Gorin, J.A. (2005). Evaluating collaborative learning and community. *Journal of Experimental Education*, 73, 165-188.

Toland, M., & de Ayala, R.J. (2005). A multilevel factor analysis of students' evaluations of teaching. *Educational and Psychological Measurement*, 65, 272-296.

Tracey, J.B., & Tews, M.J. (2005). Construct validity of a general training climate scale. *Organizational Research Methods*, 8, 353-374.

Trigwell, K., Prosser, M. & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, 37, 57-70.

Willis, G.B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.

Wilson, K.L., Lizzio, A., & Ramsden, P. (1997). The development, validation and application of the Course Experience Questionnaire. *Studies in Higher Education*, 22, 33-53.

Winer, B., Brown, D., & Michels, K. (1991). *Statistical principles in experimental design*. (3rd ed.) New York: McGraw-Hill.

Your First College Year (2007). Downloaded 16 October 2007, from <http://www.gseis.ucla.edu/heri/yfcy/>

Yorke, M., with Bell, R., Dove, A., Haslam, L., Hughes Jones, H., Longden, B., O'Connell, C., Typuszak, R. & Ward, J. (1997). *Undergraduate non-completion in England*. Bristol: HEFCE.

Yorke, M., & Longden, B. (2007). The first year experience in higher education in the UK: Report on phase 1 of a project funded by the Higher Education Academy. York: Higher Education Academy. Downloaded 30 March 2007, from <http://www.heacademy.ac.uk/research/FirstYearExperience.pdf>

11. APPENDICES

Appendix 1: Sectoral & Regional Groups in Australian Higher Education

The Group of Eight (Go8)

Australian National University
Monash University
University of Adelaide
University of Melbourne
University of New South Wales
University of Queensland
University of Sydney
University of Western Australia

Australian Technology Network (ATN)

Curtin University of Technology
Queensland University of Technology
RMIT University
University of South Australia
University of Technology, Sydney

Innovative Research Universities Australia (IRU)

Flinders University
Griffith University
LaTrobe University
Macquarie University
Murdoch University
University of Newcastle

New Generation Universities (NGU)

Australian Catholic University
Central Queensland University
Edith Cowan University
Southern Cross University
University of Ballarat
University of Canberra
University of the Sunshine Coast
University of Western Sydney
Victoria University

Non-aligned / No grouping

Bond University
Charles Darwin University
Charles Sturt University
Deakin University
James Cook University
Swinburne University of Technology
University of New England
University of Notre Dame Australia
University of Southern Queensland
University of Tasmania
University of Wollongong



Appendix 2: Examples of items in most frequently used dimensions

In this section we give examples of items from dimensions used by a medium to high number of universities.

Teacher Level Items

Clarity of Instruction

- The lecturer explained important concepts and ideas in ways that I could understand
- The lecturer regularly summarised main points of the presentation to the class.
- The quality of this teacher's explanations was...
- During each lecture, the lecturer described what students were expected to learn from that lecture
- (name of teacher) gives clear explanations.

Overall Instructor ratings

- Overall my lecturer effectively supported my learning.
- Overall, I was satisfied with the performance of this lecturer.
- Overall, the quality of this staff member's teaching was...
- Overall, how would you rate the staff member's teaching in this unit?
- Overall, the lecturer was highly effective in facilitating my learning

Preparation and Organisation

- My tutor came to class well-prepared to help me learn.
- This teacher's organisation was...
- This teacher is well prepared for class.
- (name of teacher) is well organised.
- The lecturer produced classes that were well organised.

Feedback

- During lectures, the lecturer provided useful feedback on student work.
- The tutor returned assessed work within a reasonable time.
- This teacher gave constructive feedback (in class, on assessment tasks, etc) that helped my learning.
- The lecturer provided timely and constructive feedback on my assessment tasks.
- The staff member put a lot of time into commenting on my work.

Stimulation of Interest in the Course

- The lecturer stimulated my interest in the topic area.
- The lecturer stimulated me to think about issues in this subject area.
- The lecturer helped make the content interesting and engaging.
- My attention or interest was sustained during each lecture.
- (name of teacher) stimulates my interest in learning in this course.

Availability

- The lecturer was readily available for consultation.
- The lecturer was available to answer student inquiries.
- This teacher's availability to students was...
- It was possible for me to consult the lecturer out of class about each lecture.
- The lecturer was available for consultation and individual help where needed.

Objectives

- The lecturer explained the assessment processes in ways I could understand.
- The lecturer made clear the standard of the work expected.
- How this teacher clarified the subject's expectations of students was...
- This teacher communicated clearly what was expected of me to be successful in this unit.
- This teacher explains the purpose of each class as it relates to the unit.

Concern for Students

- The lecturer was responsive to student learning needs.
- The lecturer helped me to understand problems with which I had difficulty
- The lecturer appeared willing to assist students (either during or outside lectures) with difficulties that they had with lecture content.
- I found this teacher helpful if I encountered difficulties in this unit.
- The staff member shows a genuine concern for the quality of my learning.

Interaction and Discussion

- The lecturer provided sufficient opportunities for active participation of students in this unit.
- The tutor encouraged my participation in the unit.
- This teacher encourages student participation in classes and/or groups.
- (name of teacher) encourages student participation.
- The tutor encouraged students to participate actively.

Subject Level Items

Overall Course ratings

- Overall I was satisfied with the quality of this unit of study.
- Overall, I was satisfied with this unit.
- Overall I was satisfied with the quality of this unit.
- Overall, I am satisfied with the quality of this course.
- All things considered, my overall rating of this unit is: (poor/satisfactory/excellent)

Objectives

- The learning outcomes and expected standards of this unit of study were clear to me.
- Requirements for completing the assessment tasks in this unit were clear.
- I had a clear idea of what had to be completed and the level of work that was expected.
- The criteria for success in this unit were made clear.
- It was clear what I was expected to learn in this unit.

Feedback

- I found that the comments and feedback from assessment tasks were helpful.
- The teaching staff gave me helpful feedback.
- I received timely feedback that assisted my learning.
- The feedback on my marked work was useful for my learning in this unit.
- The teaching staff normally give me helpful feedback on how I am going.

Overall Learning

- The content of the unit contributed constructively to my learning in this subject.
- The unit advanced my understanding of the subject.
- The level to which we accomplished the aims of the subject was...
- The unit allowed me to achieve the learning objectives.
- Overall, I have learned a lot in this course.

Workload

- The workload in this unit of study was too high.
- The workload in this unit was manageable.
- The amount of work required of me in this unit was reasonable.
- The overall amount of work required of me for this unit was appropriate.
- The amount of work required in this course is about right.

Assessment

- The range of assessment tasks allowed me to demonstrate what I had learned.
- The assessments were strongly linked to the unit aims and objectives.
- Assessment tasks were set an appropriate level.
- The assessment requirements of this unit are appropriate to its learning objectives.
- The assessment tasks were appropriate to the aims of this course.

Appropriate Use of Methods/Materials

- The content presented in this unit reflected the declared outcomes/ objectives.
- The activities in the unit supported my learning.
- The learning activities (assessment tasks, in-class exercises, homework, etc) were useful for building up my understanding of this unit.
- This subject was delivered in a way which was consistent with its stated objectives.



- the substance of this course was appropriate for the level.

High-level Cognitive Outcomes

- This unit of study helped me develop valuable generic attributes.
- My experience in this unit has enhanced my ability to solve problems.
- Intellectual inquiry was encouraged in this unit.
- The assessment tasks tested my understanding of the subject area, rather than just memory.
- This course helps develop my thinking skills (problem solving, analysis etc.)

Choice of Required Materials

- Overall I found the selection of print, software and hardware resources for this unit was suitable to satisfy unit requirements.
- The course materials in this unit were of high quality.
- The unit materials were current and up to date.
- The teaching materials for this unit (handouts, unit outlines, websites, etc) were helpful to my learning.
- The learning resources (e.g. handouts, web resources) are valuable for my understanding of the course.

Course/Degree Level Items

Overall Course

- Overall I was satisfied with the quality of this course.
- Overall I was satisfied with the quality of this degree course.

Clarity of Instruction

- My lecturers were extremely good at explaining things.

Monitoring Learning

- The staff made a real effort to understand difficulties I might be having with my work.

Overall Instructors

- Good Teaching Scale total score

High-level Cognitive Outcomes

- Generic Skills Scale total score

Overall Learning

- The course provided me with a broad overview of my field of knowledge.
- I consider what I learned valuable for my future.
- I consider what I learned valuable for my future.

Enthusiasm for Subject

- The teaching staff worked hard to make their subjects interesting.

Motivating Students to Greater Effort

- The teaching staff of this course motivated me to do my best work.

Feedback

- The teaching staff normally gave me helpful feedback on how I was going.
- The staff put a lot of time into commenting on my work.

University Level Items

As noted above, at the University level, there were few dimensions which are measured, and of those that were, only a few universities measure them. The most frequently used dimensions were as follows:

Overall University Experience

- I enjoyed being a student at X.
- I am satisfied with my academic experience at X.
- Teaching approaches used assist my learning.
- Overall, I am satisfied with my study experience at Y.
- I think university life really suits me.

Tolerance of Diversity

- I have learned to explore ideas confidently with other people.
- University X acknowledges the deep relationship between Indigenous Australians and the land.
- My personal beliefs were respected on campus.
- At this university I was encouraged to engage and communicate with students from diverse cultures.

Overall Learning Climate

- I feel I belong to the university community.
- How satisfied are you with Academic Support?
- I find the academic environment supportive to the learning process.
- At this university I was encouraged to seek academic support when I need it.

Additional Dimensions and exemplar items

In addition to the 42 dimensions discussed above, Survey 2 also invited universities to provide examples of additional dimensions reflecting items which could not be categorised according to the Abrami et al's framework.

No additional dimensions were volunteered at the Teacher level.

The following additional dimensions were suggested:

Learning Environment:

Facilities (classrooms, laboratories etc), which constituted the learning environment

- The facilities (such as classrooms, lecture theatres, studios, labs, workshops) are adequate for this course.
- Teaching resources and facilities (laboratories, studios, equipment) are appropriate for my needs.

Information technology enabled learning environments

- The technologies used to deliver the online content in this unit performed satisfactorily.
- Where it was used, information technology helped me to learn; The online learning experiences of my degree course were well-integrated with my face-to-face learning; My online experiences helped me engage actively in my learning.
- How satisfied are you with Information Technology Services?

Library and other learning resources present in the environment

- The Library support for my whole course this semester was...
- How satisfied are you with the services of the Library;
- The library resources are appropriate for my needs
- Student & learning support services which create an effective learning environment
- I believe appropriate support was available to help me to achieve learning outcomes in this course (technical support, on-line support, scholarly information resources, distance education support
- I was able to access quality support (e.g. from staff, other students, the university) when appropriate.

Administrative services which create an effective learning environment

- The degree course administration is effective in supporting my learning.

Student Motivation and Effort:

Student Motivation

- I am motivated to achieve the learning outcomes in this unit
- I make best use of the learning experiences in this unit
- I think about how I can learn more effectively in this unit
- I do some of my best work in this subject.

Student Effort

- On average, how many hours a week did you work on this unit? (0-5/6-10/11-13/14-19/20+).

Active Student Learning

- I have collaborated with other students in learning

Appendix 3: Forum Workshop Powerpoint Presentation

THE CARRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

carrick
INSTITUTE

Rewarding and recognising quality teaching and learning in higher education: Study 6: Student surveys on teaching and learning

A/Prof Simon Barrie, Dr Paul Ginns & Ms Rachel Symons
The University of Sydney

An initiative of the Australian Government Department of Education, Science and Training

Today's Workshop

- Overview of study 6
- Outline of Phase 1: Survey 1 & 2
- Review of preliminary findings
- Data sharing arrangements
- Coffee break
- Outline of Phase 2: supporting the Carrick pilot
- Opportunity for review and update to Survey 1 and 2 data
- Collaborative exploration of shared scales and items
- Planning for ongoing collaboration

THE CARRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Study 6 Overview

- Larger project on recognising and rewarding quality teaching & learning
- SET as core element of QA
- Current practice across the sector
- Suggestions in relation to the use of SET in the 2008 pilot phase
- Foster ongoing collaboration & development

THE CARRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Review of Survey 1

Four levels

1. Individual teacher
2. Department/school/faculty
3. Whole institution
4. Across sector.

Focus questions

- **Purpose & use:** Please describe the sorts of surveys and survey practices as well as the ways student feedback data are used at this level?
- **How is the data interpreted?** Please describe how the data is aggregated and how comparisons are made? Are standards applied? What are these? How are benchmarks or averages decided? Who has access to the results?

THE CARRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Key Findings Survey 1

Considerable variation in SET policy and practice

Limited internal benchmarking of SET data – and significant variation

Little or no external benchmarking of SET (except CEQ & some discipline networks, ATN)

Increasing use of SET internally – strategy & research

THE CARRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Review of Survey 2

Core teaching evaluation scales/items at four levels

1. Individual teacher
2. Unit of study
3. Degree
4. Whole university

42 dimensions clustered under 6 areas

1. Teacher predispositions / personality
2. Course preparation and organisation
3. Approaches to teaching and teaching strategies
4. Quality of learning outcomes
5. Learning climate
6. Assessment

THE CARRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Key Findings Survey 2

- Inconsistent coverage – variation between unis
- Some notable gaps – e.g. institutional climate
- Most overlap at degree level (CEQ)
- Most overlap in relation to “overall” type items
- Focus shifts between levels
- Little internal alignment

THE CURRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Moving forwards

Other issues?

Opportunity to review & update?

Sharing data – useful? (Survey 1? Survey 2?)

Process?

Coffee

THE CURRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Phase 2

- Collaborative development of alternative approach to teaching quality indicators by Carrick intended to better support enhancement activities in Australian universities
- 8 pilot universities in 2008
 - Deakin
 - Griffith
 - Macquarie
 - RMIT
 - University of Queensland
 - University of Western Australia
 - University of South Australia
 - University of Tasmania
- Develop all institutions' potential for SET benchmarking and collaboration

THE CURRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Outline of Study 6: phase 2

SET in relation to pilot of new QA framework

Goal – explore potential for shared SET data relating to dimensions:

1. Institutional climate and systems
2. Diversity
3. Assessment
4. Student engagement and learning community

At the following levels:

1. Whole university
2. Faculty/School/Department
3. Degree/Programme
4. Subject
5. Teacher

THE CURRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Phase 2: Getting started

How we might build from existing SET practice?

Parameters that would allow SET to be used in this way?

Where we are looking?

- Phase 1 data
- Other countries' surveys
- Surveys in development
- Potential surveys in development (ASV)

Where we have got to

- Some initial ideas
- Some limitations

Invitation to work with us to shape next generation of SET practices

THE CURRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Collaborative development: Task 1

• Four groups – one for each dimension

- Read through and see if you think the potential scales/items meet the parameters
- Suggest additional scales items – or additional parameters
- Help fill in any blanks you think appropriate
- Reflect if your group could agree to share any of the common SET scale / items for any dimension

- Laptop/group – annotate using track changes
- Report back: Key observations & any potential commonalities

(If you would rather work on reviewing your survey 1 & 2 data please do)

THE CURRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Collaborative development: Task 2

- Pairs of 'like' institutions – potential or current benchmarking partners
- What would have to happen in terms of your SET policy and practice to enable you to share useful data from any potentially common scales / items identified

Some possible considerations

- What would be useful data
- Who would lead this decision making
- Who would need to agree to any sharing
- How would you use the data to start conversations to support interpretation
- What system do you need to use the data to inform enhancement
- What other data would be useful to supplement your SET data
- What alternative data might be better

Report Back: Action plan

THE CARRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD

Where to next?

Finalising our report December 2007

- Share phase 1 data if/as agreed?
- Opportunities for further conversations & collaborations?
- Potential to contribute to SET aspects of pilot

THE CARRICK INSTITUTE FOR LEARNING AND TEACHING IN HIGHER EDUCATION LTD