

Chapter XIII

E-Assessment: The Demise of Exams and the Rise of Generic Attribute Assessment for Improved Student Learning

Darrall Thompson
University of Technology, Sydney, Australia

Abstract

This chapter explores five reasons for a reduced focus on exams by questioning their value and sustainability in the assessment of student learning. It suggests that exam grades cannot provide accruing developmental information about the students' attributes and qualities vital for a changing world and workplace. It then argues for the integrated assessment of generic attributes (including those developed through exams) and describes two e-assessment tools developed by the author to facilitate this approach. These tools are based on the concept that assessment criteria should encompass the complete range of attributes and qualities that institutions proclaim their students will acquire. Given that assessment drives learning, explicit alignment between assessment tasks and criteria is essential. It is proposed by this chapter that the development of formative criteria (numerically valued) together with expert-derived criteria groups can facilitate students' development of important qualities, or generic attributes at both school and tertiary levels of education.

Introduction

The term *generic attributes* used in this chapter (sometimes referred to as *graduate attributes*) is intended to incorporate a broad range of qualities that are often claimed by educational institutions describing those who complete their courses of study. It is broader than the terms *key skills*, *generic skills*, and *key competencies* often interchangeably used in this research area.

The reader may well ask why a senior lecturer teaching visual communication is writing about e-assessment and educational research. It may be wise to slip into first person for a paragraph at the beginning of this introduction to contextualize and validate the contribution that this chapter is attempting to make.

On entering university teaching, my knowledge of educational research was limited to a 1-year postgraduate teaching certificate. A 6-month secondment to the University of Technology, Sydney (UTS), Centre for Learning and Teaching initiated my focus on research in this area. The realization that my background in information design and visual communication had something to bring to the design of learning environments led eventually to a research master's in design education and the design and development of online assessment systems described in this chapter. I do not teach on an exam-based course but have worked for 15 years with colleagues that do so. It is not my intention here to present an in-depth study of exams as an assessment strategy but rather to provide powerful reasons and supporting references that may encourage greater questioning of the value and sustainability of exams in educational contexts. My reasons for encouraging the assessment of graduate attributes are based on a long association with criteria-based assessment. I believe that, in a rapidly changing world and workplace, students, staff, and employers need much more feedback about the development of graduate attributes. These are hidden or simply not assessed by exam-based summative approaches.

The first part of this chapter explores five reasons to question the value and sustainability of exams in formal educational contexts. The references used include educational research, a recent United States patent granted to Microsoft[®], and studies on youth suicide.

The second part explores five reasons for the explicit integration of graduate attributes in curricula and assessment processes. The Australian government's concern with the fact that graduate attributes publicized by universities were often not explicit in curricula or assessed in practice led them to initiate an independent Graduate Skills Assessment (GSA) test. This out-of-context approach is diametrically opposed to the integrated systems proposed in this chapter. A brief analysis of the GSA in this text concludes that whilst a great deal

of work has been put into it, there are flaws in both the approach and the questions themselves.

The third part describes an e-assessment system developed by the author (Re:View) and how knowledge and skills usually assessed by exams can be integrated with graduate attribute descriptive assessment that exams fail to address. This is done using assessment criteria for marking and formative feedback for learning tasks. These criteria are categorized in groups of attributes that accrue summative percentage marks in a secure online database. Students can visually monitor their development during a course of study in each category over a range of tasks (including exams assessed with criteria grouped under the same categories). Through research and experience in using this system, five optimal groups of attributes have emerged: creativity and innovation, communication skills, attitudes and values, professional skills, critical thinking, and research.

The second e-assessment system developed by the author in collaboration with other academic colleagues at UTS is called the Self and Peer Assessment Resource Kit (SPARK). It is used for the assessment of group projects using online self and peer ratings of group-performance criteria by students. These ratings produce a “factor,” which when multiplied by the group mark produces individual marks for each group member.

The last part of the chapter illustrates the use of these two systems in an undergraduate program. The examples from case studies outlined are from a 4-year university undergraduate honors degree course and are included to provide evidence of graduate attribute development facilitated by these online systems.

The Demise of Exams: Five Reasons to Question Their Value and Sustainability

- Exams encourage “rote learned” responses and reliance on factual memory.
- Exams encourage surface approaches to learning.
- Exams viability will soon be challenged by the availability of memory-enhancing drugs and emerging technologies.
- Exams facilitate the “fee for degree” commercialization of education.
- Exams cause physical illness, depression, and youth suicide.

The first two reasons are different but related reasons and will be considered together.

Considering the first two reasons as they operate in a highly competitive environment the extensive use of practice exams and the study of answers to past exam papers can dominate the learning environment to the exclusion of more innovative strategies. The rote learning of facts and the one-off extrinsic motivation to succeed in exams is also known to encourage surface approaches, minimal retention of facts and poor knowledge transferability to other contexts. Thorough research on the effects and consequences of surface/atomistic approaches and deep/holistic approaches to learning are readily available in many published works by Marton, Hounsell, and Entwistle (1984), Ramsden (1992), Ramsden and Entwistle (1981), and others:

Surface approach and the motive of fulfilling demands raised by others (i.e., [sic] extrinsic motivation) seem to go together. (Marton & Saljo in Marton, et al., 1984, p. 51).

The following quote from a student in a study by Ramsden in the same book *The Experience of Learning* is just one from hundreds of similar studies:

I hate to say it, but what you've got to do is have a list of the facts, you write down ten important points and memorize those, then you'll do alright in the test.... If you can give a bit of factual information - so and so did that, and concluded that, for two sides of writing, then you'll get a good mark (Ramsden in Marton et al., 1984, p. 144)

This quote from Ramsden's research is a student who received a first-class honors degree (ironically) in Psychology, and reveals a surface approach to learning even though they "hate to say it." It implies they are uncomfortable with achieving success through a memory-based formulaic approach that gives them good marks every time.

The highly complex issue of what is actually being assessed in the psychology student's essay is not addressed here, but new approaches to the assessment of essays online continue to be onerous and problematic (Shermis, Mzumara, Olson, & Harrington, 2001). It could be argued for example, that the psychology student is actually showing a certain degree of synthesis. However the point being made here is that the assessment method itself is adversely affecting the students' approach to their study, and this impact is supported in many specific researches, for example Biggs (1995), Marton and Saljo (1976), Ramsden and Entwistle (1981), and Steinberg (1997).

Apart from essay exam questions, there has been a large increase in the popularity of multiple-choice questions (MCQs), easily ported to online Web

sites and with the added advantage of automatic marking. These have been criticized for their bias toward memory testing and the design of these exams has been identified as a major area for concern by educational researchers. It is apparently very difficult to provide MCQs that do not cause the selection of the correct answer to be too easy, too difficult or just plain ambiguous (Pritchett, 1999).

The main issue centres upon the institutions' and lecturers' responsibility to design courses that encourage deep student engagement and their development of qualities and attributes essential for lifelong learning and continuing employment.

Exam's viability will soon be challenged by the availability of memory-enhancing drugs and emerging technologies. As early as February 2000 an article appeared in the Times Educational Supplement, titled "Spectre of Exam Drug Test Looms", that showed successful results with mice and reporting that "according to scientists brain-boosting drugs may soon hit students" (Bunting, 2000, p.3). By 2004, the economic potential for pharmaceutical companies had expanded the drive for further research and development:

At least 40 potential cognitive enhancers are currently in clinical development, says Harry Tracy, publisher of NeuroInvestment, an industry newsletter based in Rye, New Hampshire... The interest in such drugs will not stop there, predicts James McGaugh, who directs the Centre for the Neurobiology of Learning and Memory at the University of California at Irvine. Next in line could be executives who want to keep the names of customers at the tips of their tongues, or students cramming for exams. (Economist, 2004, p. 27)

Apart from the problems associated with these advances in drug development, there are also the rapid advances in communication and computing technology. The integrity of exam submissions will be unverifiable due to the miniaturization of digital technology and wireless communication systems. To highlight the extent of this problem with regard to cheating in exams it might be worth considering the following extract from US patent no. 6,754,472 applied for on April 27, 2000, and granted to *Microsoft*[®] on June 22, 2004.

The human body is used as a conductive medium, e.g., a bus, over which power and/or data is distributed. Power is distributed by coupling a power source to the human body via a first set of electrodes. One or more device

to be powered, e.g., peripheral devices, are also coupled to the human body via additional sets of electrodes. (Williams, Vablais, & Bathiche, 2004)

Whilst there may be concern about health issues arising from electronic data transfer using the human body, the recent history of technological development has shown that the rate of technology adoption has been exponential and inevitable. Devices that transmit and receive data using human skin conductivity and human eardrums instead of headphones may soon become the technology that makes all information invisibly and undetectably available.

If the only educational challenge for young people at school or university is to pass an exam, Microsoft[®] and the drug companies are about to sell students the means to overcome that challenge, unless of course examinees are all drug tested, strip-searched and conducted to a specially designed radiation-proof room. Perhaps it would be better, and much less expensive, to encourage them to have a deep approach to their own learning and personal development using assessment processes that encourage and facilitate that approach.

Exams facilitate the “fee for degree” commercialization of education. It is conceded that the idea of the demise of exams is counter to current education and e-learning trends. Massachusetts Institute of Technology (MIT) and other North American universities have put their course content free online but charge large fees to take the exam. Given the focus on short-term memory and problems of invigilation already mentioned, is not this rather a dangerous direction? Does it fulfil the developmental needs of young people for survival in this new millennium, or just rely on the fact that those who can afford the fee are likely to have backgrounds that guarantee survival anyway?

The Australian government is making strong statements about the professionalism and quality of teaching in higher education, but whilst pressing for an increase in quality teaching, the government is reducing funding and encouraging commercialization:

Most importantly, teaching needs to be accorded a much higher status in universities. It is necessary to take a broader conception of academic work and the validation of alternative career paths to improve the status of teaching. The quality of teaching is absolutely central to the learning experience. There needs to be a renewed focus on scholarship in teaching and a professionalisation of teaching practice. (Department of Education, Science and Training, 2002, p. 10)

The demand for financial viability means that a focus on quality teaching is unlikely to happen with this economic pressure. Any attempt to improve the

quality of teaching through teacher development is expensive and accrues no direct financial benefit. Gaining research grants and collecting exam fees are far more profitable strategies. Equivalent financial support for teaching and learning is unlikely to be forthcoming unless governments are prepared to invest in providing resources.

Requiring fees for exams may also be a strategy that only the high profile “brand name” universities can apply. If all universities try to follow this approach, will education become like the car industry, where 50 smaller manufacturers have now been replaced by 4 massive ones? The early signs of this are already occurring in Australia, and with the advent of the US free trade agreement, this is likely to gather momentum. In 1999, evidence was emerging in the media:

Large for-profit corporations like Jones International University and the University of Phoenix have entered the huge and growing “virtual university” market to claim their share... indications are that virtual higher education will surely become a large enterprise. According to John Chambers, CEO of Cisco Systems, the company that makes routers that direct traffic on the Internet, education is the next big “killer application.” Chambers believes that “Education over the Internet is going to be so big it is going to make e-mail usage look like a rounding error! Chambers warns, “Schools and countries that ignore this will suffer the same fate as big department stores that thought e-commerce was overrated.” (New York Times, 1999)

In October 2004, the South Australian government announced that Carnegie Mellon will be opening a new university in Adelaide in 2006, offering American undergraduate and postgraduate degrees with private students receiving government loans on the same basis as local students.

Whilst the media makes no mention of free content and fee for exam models, the potential for this development is clearly evident. This further commodification of education in Australia may reduce the emphasis on, and encouragement of, good teaching and effective assessment of high quality learning outcomes.

Exams cause physical illness, depression, and youth suicide. This is perhaps the most tragic reason for reducing the focus on exam-based systems.

New South Wales (NSW) in Australia has a Higher School Certificate (HSC) exam, which is used to determine a University Admission Index (UAI). A report commissioned by the NSW Commission for Children and Young People (Sankey & Lawrence, 2003) studied the population of all deaths of children and young people in NSW by suicide or risk-taking, over a 5-year period (January

1996 to December 2000). The upper age limit for this study was 17 years 11 months, which suggests that its findings could be an underestimate of the number of HSC-related deaths, as many taking the exam were over 18. From 187 young people committing suicide in this period, 38 students were reported to have committed suicide as a result of school-related problems. Ten of these were directly related to HSC stress, and a further eight from related learning difficulties.

Previous research had not documented a link between HSC stress and suicide in NSW, however Smith and Sinclair (2000) found that more than 40% of Year 12 (HSC) students in their study reported symptoms of depression, anxiety, and stress that fell outside normal ranges.

In considering the university age range, the alarming increase in youth suicide (particularly males aged 15-24) prompted the Australian government in 1997 to allocate \$31 million to the National Youth Suicide Prevention Strategy aimed at reducing youth suicides in this age range by June 1999 (Australian Government Publishing Service, 1997).

Although the following extracts are from a specific study, my brief research into reports from Japan and the United States shows that they are typical of extracts from studies in other countries with exam-based systems.

Report Extracts (Sankey & Lawrence, 2003):

The young people whose records indicated significant stress levels associated with their impending HSC exams all appear to have suicided in a state of acute stress and in close proximity to an event relating to their exams. (p. 67)

Of the 8 young people experiencing learning difficulties... When Chris told his father that he was at the bottom of his year, his father said that he didn't mind, he just wanted him to complete his HSC. A few weeks prior to his death, Chris told a friend that he felt "dumb" and was finding it very difficult to cope with this. He further said that he would rather be dead, in heaven where it was more peaceful. (p. 57)

Of the JO young people who experienced significant levels of HSC-related stress, all died by suicide. As a group, these were successful students, with records indicating that they set high standards for themselves and worked extremely hard. Documentation also showed that the period leading up to their deaths was typically characterised by feelings of overwhelming pressure to succeed, coupled with an intense fear of failure. (p. 55)

The finding of an association between HSC stress and suicide warrants urgent investigation of how to support young people during this stressful period and how to work with parents and the community to provide realistic guidance to students. The Child Death Review Team (CDRT) considers that there is a need for the Strategy to address this important issue. (p. 114)

Dr. Anthony Kidman, director of the UTS Health Psychology Unit, conducted a study in 2004 in which an average of two out of five teenagers believed that the HSC exams would affect the rest of their lives. In an article about the study he said, “There is significant anecdotal evidence to suggest burn-out in a large number of students as well as sleeplessness, suicidal ideas and anxiety” (*Sydney Morning Herald*, 2004).

Dr. Gary Galambos in the same article commented, “You have to take that really seriously in a student population because the risk of suicide in teenagers is very high in Australia” (*Sydney Morning Herald*, 2004).

Dr. Kidman and his team are taking a very positive approach to the situation and have designed a “psycho-educational” program for teenagers studying the HSC. The program is called “Taking Charge! A Guide for Teenagers, practical ways to overcome stress, hassles and other upsetting emotions.”

The pressure exerted on students by themselves, parents, relatives, peers, and the institutions involved, to achieve good exam results is clearly intense. It has become common practice for students to be advised to choose particular subjects to gain high university entrance scores. This extrinsic motivation overrides the intrinsic motivation to follow a natural inclination or interest.

Few would argue that stresses have not increased, even in the last 5 years; and burgeoning plagiarism and cheating in exams have become major concerns for many educators. *Test anxiety* and *exam stress* are now common terms and inevitably some students who cannot cope with these pressures can become reliant on prescribed antidepressant drugs. Selective serotonin reuptake inhibitors (SSRIs), are the new popular group of drugs that are supposed to have fewer side effects than previous tricyclic drugs.

Speaking at the 11th Annual Suicide Prevention Australia National Conference in October 2004 in Sydney, director of the Australian Institute for Suicide Research and Prevention, Professor Diego De Leo warned that SSRIs may not be a cure-all for depressed kids. He reported that prescriptions for the drugs were increasing, with some given to children presenting signs of suicidality as young as 10 or 11 years old. His concern was that no studies have been done with children in Australia, but recent investigations in the United States urged extreme caution in the prescribing of these inhibitors.

There has also been a link made between physical illness and exam stress:

The psychological stress of school exams can increase the severity of asthma by increasing airway inflammation response, said Dr. Lin Ying Liu and associates at the University of Wisconsin, Madison. In a study of 20 college students with mild allergic asthma, the percentage of sputum eosinophils was 10.5% at 6 hours and 11.3% at 24 hours following an antigen challenge during final exam week—significantly higher than the 7% level during a periods of low stress. (Ying Liu, 2002, p. 15)

It is clear from these studies that young people do not need an assessment strategy that adds aggravation and stress to the other sociocultural pressures during these important developmental periods. Is there a better and more developmental approach to the assessment of knowledge, skills, and attributes, instead of what has proved to be a stressful hurdle biased towards those who can cope and/or cleverly regurgitate?

For all the reasons mentioned in the first part of this chapter, exams have been reasoned to be an indefensible strategy for assessment. Their retention is not inevitable as there are viable alternatives that improve student learning and reduce aggravation and stress for both staff and students.

The following part of this chapter outlines the reasons for the assessment and development of graduate attributes, followed by a description of the effective application of e-assessment systems to enable this process.

The Rise of Generic Attributes: Five Reasons to Encourage Their Explicit Integration in Curricula and Assessment

- Australian Universities Quality Agency (AUQA) audits identify graduate attributes as a problem area.
- Universities need to validate their attribute statements through assessment.
- Schools and universities are responsible for student employability not just accreditation.
- Students actually need these attributes to cope with a changing world.
- Well-designed learning tasks can develop a very broad range of generic attributes.

AUQA audits identify graduate attributes as a problem area. Graduate attributes are referred to in the United Kingdom as key skills and other terms include generic attributes (Wright, 1995), *key competences* (Mayer, 1992) or *transferable skills* (Assiter, 1995). In addition to discipline knowledge, and skills, graduate attributes include, for example: problem solving, interpersonal understanding, critical thinking, written communication, and teamwork skills. Statements of graduate attributes can be found in one form or another on the Web sites and in the documentation of most universities in Australia. However, the statements appear to have minimal implementation in curricula. In a report on the 2002 AUQA audits, this was identified as a major problem:

The audit sampling process picks out the teaching of generic skills or graduate attributes for particular attention. Reflecting the policies of an earlier incarnation of the Department of Education, Science and Training, most of the institutions have some formal statement about the skills they aim to instill but are labouring to devise a means of ensuring that this occurs. One institution was commended for its explicit attention to graduate attributes; at least three institutions were seen to be fumbling the implementation of their policies. (Martin, 2003, p. 15)

It is interesting to note that perhaps thirty years ago “a university education” was assumed to instil implicit attributes in its graduates through traditions, social events, field activities, sports, and so on. This was clearly the case with Oxford and Cambridge whose graduates were considered to have the “moral fibre” and high standards suitable for service in the government or the church. The reduction of traditional socialization could be considered a backward step, but the explicit statements are more appropriate in a deregulated context where a university education is now ubiquitous. However, universities should perhaps be less forthright in the boldness of their graduate-attribute claims in an increasingly litigious society. AUQA’s audit and open-publishing process may well provide public assurance of the improvements needed in this area. Their mission reads: “By means of quality audits of universities and accrediting agencies, and otherwise, AUQA will provide public assurance of the quality of Australia’s universities and other institutions of HE, and will assist in improving the academic quality of these institutions” (Australian Universities Quality Agency, 2004, p. 2).

Universities need to validate their attribute statements through assessment. In order for universities to ensure their graduates’ achievement in the qualities in attribute statements, they must initiate methods of assessing and monitoring attribute development in students.

In 1999, the Australian government decided to test the efficacy of universities' graduate attribute development by commissioning the Australian Centre for Educational Research (ACER) to develop the GSA test. This is completed voluntarily by students at entry, and again at graduation from the university sector, in order to compare results. Whilst the government's concerns are justified by recent AUQA audits, the test does not assess attributes in context and the questions are unrelated to the students' course of study. The test consists of a 2-hour multiple-choice exam and 1-hour essay, but to date has not been received favorably by universities or students. However, a brief analysis is included here, because the thrust of this chapter is diametrically opposed to this external examination method by which the development of graduate attributes or qualities can be measured.

The Australian GSA Test

Graduate Skills Assessment: GSA Introduction

ACER was commissioned by DETYA, in 1999, under the Higher Education Innovation Program to develop a new Graduate Skills Assessment test. The test has been designed to assess generic skills of students when they begin at university and just before they graduate. The four areas currently included in the test are: Critical Thinking, Problem Solving, Interpersonal Understandings, and Written Communication. Many universities have identified these skills as important and are included in their graduate attributes or capabilities statements. The test consists of a multiple-choice test and two writing tasks. The multiple-choice test is currently two hours long and the writing test is sixty minutes long. (Australian Council for Education Research, 2001, p. 1)

This is a very long test by normal standards, and if implemented by all universities will need a large resource in adjudication and benchmarking processes to achieve consistency. The practical implications are enormous in regard to the storage, translation, and analysis of information, and the marking of a massive amount of essay material. The data would also have to be stored for a minimum of four years for comparison to be made with exit scores, and no mention is made of the institutions that will have access to it, the privacy issues involved, equity and language issues, and so on.

The GSA claims to test a range of graduate attributes such as problem-solving and interpersonal skills. For example, Unit 1, Question 1 from the sample questions on Problem Solving, asks the student to work out the scheduling of

netball teams to meet, whilst avoiding their regular training sessions that are listed. It then shows drawings of four sample schedules from which to select, in a multiple-choice list. This task involved simply matching gaps in times listed on pairs of charts. The answer to this question does not require problem-solving skills but the ability to match diagrams with lists, and is clearly below university level.

The following question is a sample from the Interpersonal Understandings section of the GSA: UNIT 11 Question 23:

23 A job interviewer asks an applicant the following question:

“How would you persuade a person working with you in a team to follow your suggestions when that person is reluctant to do so?”

Which one of the following responses by the applicant most strongly suggests an ability to work well in a team?

- A: I would do the work myself. At least that way I know it would be done properly.*
- B: I would make it very clear that I was more experienced in these matters than him/her.*
- C: I would find out more about the person's concerns and then discuss these with him/her.*
- D: I would follow his/her suggestions rather than mine to show the person they are wrong.*
- E: I would point out that in a team there has to be some give-and-take, and that he/she should listen to me this time.*

(Australian Council for Education Research, 2002, p. 14)

Firstly, the question is contextualized, as though it is being asked in a job interview rather than a real group situation. It is possible that this is to trick the examinee (and the applicant) into a preset persuasive position, which they would need to avoid to answer the question correctly. This convoluted context is asking a young person to make a judgement from the position of an employer asking a trick question of an illusory applicant. How this is related to the examinees natural response when immersed in a group context is hard to fathom. Secondly, there are perhaps more appropriate responses, such as: F: I would apologize for trying to impose my suggestions through persuasion!; G: I would suggest we have a look

at the criteria against which the project will be assessed; or H: I would suggest we have a brief discussion about why we are doing the project and what we want to get out of it.

MCQs are unlikely (even with clever design) to test higher-level learning or attributes because of forced choice options (Pritchett, 1999). Given all the reasons discussed in the first part of this chapter, it would clearly be preferable not to add yet another onerous exam (for staff and students alike).

This chapter proposes that the solution is to ensure that the educational experiences in which students engage encourage the development of graduate attributes and that universities monitor the development of them throughout a student's course of study. The facilitation of this process is addressed by the database-driven Web applications described in this chapter.

Schools and universities are responsible for student employability not just accreditation. Business organizations rely on universities to stay ahead of changing requirements and to provide graduates with appropriate attributes and skills. However, because of their own increasingly pressured environment, employers are becoming more vocal in what they expect the university environment to deliver.

Graduate attributes have been identified as vital qualities for successful employment and lifelong learning. A number of employer studies have identified serious flaws in graduates' attributes. For example a study of design engineering graduates and their employers in the United Kingdom (Garner & Duckworth, 2000, p. 208) revealed a deep dissatisfaction with current graduate profiles. In their study, the employers' criticisms of graduates included the following points:

- They need greater ability to take other people's ideas on board.
- They have a lack of resilience to criticism.
- They have a weak ability to muster a reasoned defence of their contribution.
- They need to improve listening skills.
- They need higher-quality written, graphic, and verbal communication.
- They need to be able to be critical of their own work and contributions.

The following quote from the summary of their study resonates with a need for versatility and diversity in employees and lays a huge expectation at the doors of the university learning environment:

A breadth of skills and knowledge seems vital—as one manager put it, “we can’t afford specialists. The desired profile seems a broad one: creative and analytical; practical and academic; numerate and literate; able to exploit both divergent and convergent thought processes; sensitive and strong!” (Garner & Duckworth, 2000, p. 211)

Salchow (1994) in his critical writing on “employer-driven” responses in the educational provision for professional graphic designers suggests a more balanced view by suggesting that: “We should not attempt to give the provincial employer everything he expects of an applicant if it contradicts the needs of our students, society, and the profession” (p. 221).

However, a graduate fulfilling the attributes described in Garner and Duckworth’s (2000) summary, could hardly fail to satisfy Salchow, the students, society, or the professions.

Students actually need these attributes to cope with a changing world. The exponential changes outlined in the first part of this chapter require graduates to have a broad range of attributes, including the versatility with which to apply them. Toffler (1980) goes further in his prophetic work: “The illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn, and relearn” (p. 12).

It is ironic, in my view, that the educational focus on the examination of reading and writing in children, often ignores, and perhaps even stifles, their natural curiosity and desire for learning. The ability to cope with, and adapt to change, is also hinted at in Toffler’s (1980) definition of literacy for the 21st century. It is interesting to note that these abilities or qualities are supported in literature by those studying the natural worlds. Suzuki (1998) in his book *Earth Time* suggests that the two most important qualities of a sustainable life form are versatility and diversity. He also implies that these qualities may equally be necessary for the sustainability of business organizations and individuals in society. The knowledge and skills fundamental to an undergraduate degree are likely to be less important than the acquisition of these generic attributes for survival in a changing world and workplace.

Well-designed learning tasks can develop a very broad range of attributes. It is clear that the development of the broad range of generic attributes promoted in university statements can never be exclusively due to their educational experiences and may develop as a result of engagement external to the university. The lack of teaching and assessment of these attributes has been a cause for government concern in recent university audits by AUQA mentioned

earlier. However, well-designed learning environments offer a unique opportunity for the development of this broad range of attributes. The reasons for a mismatch in attribute statements and actual outcomes identified in AUQA audits are difficult to define. If graduate attributes are to be achieved through learning experiences they ought to be fundamental to the design of learning tasks and the basis for assessment criteria.

The following part of this chapter outlines two e-assessment systems developed by the author at UTS for the assessment of graduate attributes as part of an integrated approach, incorporating other knowledge and skills criteria.

E-Assessment for the Development of Generic Attributes Through Individual Tasks and Group Work

It ought never to be said that software of any kind is a solution, and later parts of this chapter propose that any assessment system needs to be carefully integrated at all stages of the learning design.

It is clear that exam grades do not provide the range of information about a graduate attributes necessary for a changing world and workplace. Written reporting does not accrue summatively to clearly indicate development in key areas. So how can software assist in this context?

In the first instance, the designers of exams need to develop criteria that actually make explicit what an exam is measuring or testing. For example, if an exam is testing the memorizing of facts, then this should contribute to a student's development of memory skills. If it is testing whether they can reference an academic document correctly then that should contribute to the students' development of research skills. If it is testing their ability to communicate rather than remember a concept, then the assessment should be adding to the students' ongoing development of communication skills. If they are asked to apply a concept innovatively, then assessment should be adding to their ongoing development of creativity and innovation; and if they are asked to analyze a text, the assessment should add to their ongoing development of critical thinking skills... and so on. If these criteria are made explicit and each one graded rather than all this information being hidden in a single grade, then the software can ascribe a value to each and store the numbers under different criteria categories.

If the exams are marked using criteria in this way, then other tasks throughout a course of study that also contribute to these criteria categories (including group

projects) can be added giving ongoing feedback about the students' development within these categories of attributes.

There is an obvious advantage in using systems that can combine the assessment of graduate attributes that are impossible to assess with exams, with criteria-based measurement of discipline knowledge and skills, which is normally assessed summatively.

The two software systems described in the following sections have been used in a university context, where the development of assessment criteria in various attribute categories has gradually occurred over a number of years.

- **Re:View - Online Criteria-Based Assessment.** A database-driven Web system for integrating graduate-attribute assessment with other criteria, and enabling students to monitor their progress in graduate attributes across a range of subjects.
- **Self and Peer Assessment Resource Kit (SPARK).** An online assessment system to enable self and peer ratings against group-work criteria, to be used in calculating individual marks for a group project.

Both these systems have been designed, at a grass-roots level, by academics attempting to solve problems related to assessment and feedback. As such, they have both been very successful and continue to develop and refine through pilot schemes in various educational contexts.

It is generally accepted in educational literature that to try to teach graduate attributes as separate subjects is not a successful strategy, and that attributes need to be embedded in curricula but made explicit in assessments. The Re:View

Figure 1. Logotype designed by the author for the Re:View Web-service assessment system:



The logotype for Re:View features the word "Re:View" in a bold, sans-serif font. The colon is a simple vertical line. The "View" is in a lighter grey color. To the right of the "View" is a small graphic consisting of four horizontal bars of different colors: red, yellow, green, and blue.

Figure 2. Logotype designed by the author for the SPARK Web-based group-work assessment system:



Figure 3. Lecturer's view of assessment and feedback marking screen:

The screenshot shows the UTSAssess marking interface. At the top, there are navigation buttons: Logout, Back to Menu, and Unpublish Marks. Below this, there are tabs for Students, Teaching, and Learning. A list of criteria is displayed on the right, each with a vertical slider and a numerical score. The criteria include: visual exploration and experimentation (65), origination of imaginative visuals by hand and digital production methods (61), thoughtful composition and hierarchical structuring of text and handwritten type as relevant to the subject (66), demonstration of an understanding of typographic practice and detailing as evidenced in presentation and engagement (peer assessed) (53), an indication of sensitivity to typographic detailing, impact and clarity of final presentation (64), and presentation of carefully considered final work, presented with professional skill (42). A total score of 319 is shown at the bottom right. There are also buttons for Special Consideration, Late Penalty, and Save.

Criteria	Score
visual exploration and experimentation	65
origination of imaginative visuals by hand and digital production methods	61
thoughtful composition and hierarchical structuring of text and handwritten type as relevant to the subject	66
demonstration of an understanding of typographic practice and detailing as evidenced in presentation and engagement (peer assessed)	53
an indication of sensitivity to typographic detailing, impact and clarity of final presentation	64
presentation of carefully considered final work, presented with professional skill	42
Total	319

Graduate Attribute Development Categories

Colour-Coded Criteria
Appropriate Categories are allocated to criteria

Data-Sliders
Vertical bar slides whilst marks are automatically calculated for each criteria

Total Data-Slider
This slider can be adjusted whilst keeping all other criteria sliders proportionally accurate

system is designed to facilitate this assessment within normal learning tasks and alongside other more content-focused discipline knowledge and skills criteria. SPARK is designed to bring group work into mainstream summative assessment using criteria to feed back to students about their development of group skills through self and peer assessment ratings that modify group marks into individual marks for each student.

A Description of the Re:View - Online Criteria-Based Assessment System

One of the major problems in criteria based assessment is that percentage marks allocated to particular criteria are rarely presented to the students as an additive progression contributing to important strands of learning. By categorizing criteria under graduate attribute development categories, the Re:View system can gather and accrue ongoing “profiles” showing progress in each category over time and across many different subject assessments.

Another problem is the difficulty of delivering private individual assessment and feedback to students other than by laboriously copying and pasting to online grade books or to e-mail addresses that are often out of date. With this system the student can log on at anytime from anywhere via the Web and see assessment and feedback the moment the assessing lecturer clicks the “publish marks” button.

Figure 4. Student view of assessment-feedback screen:

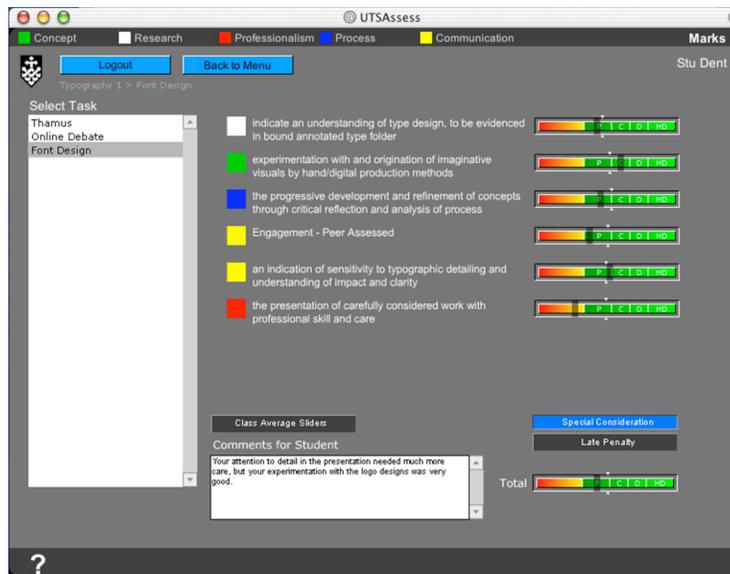
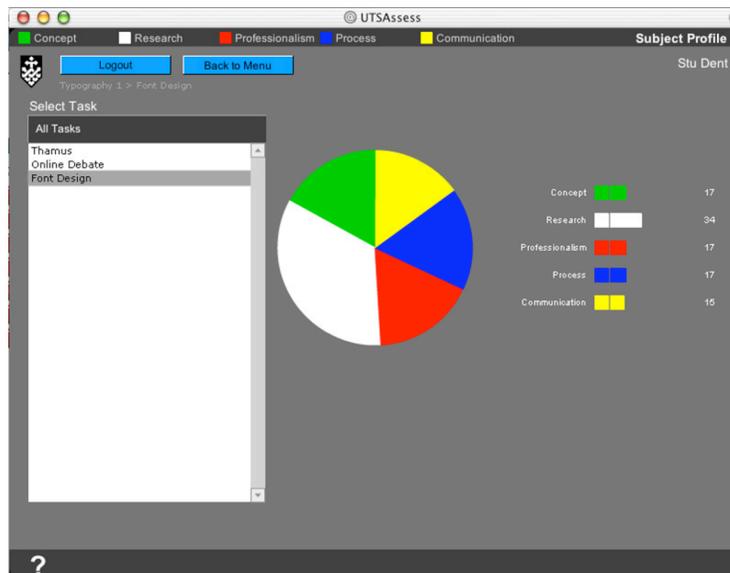


Figure 5. Feedback against an early set of graduate attribute categories based on an interpretation of the visual design process:



In the development of this e-assessment system, the criteria categories and individual criteria had to be written and refined. One side effect of entering these into the database system was that curriculum developers and staff were able to see a progressive overview of learning objectives reflected in assessment criteria across a broad range of subjects. There can be a great deal of hidden repetition in a course where individual lecturers are left to design their own

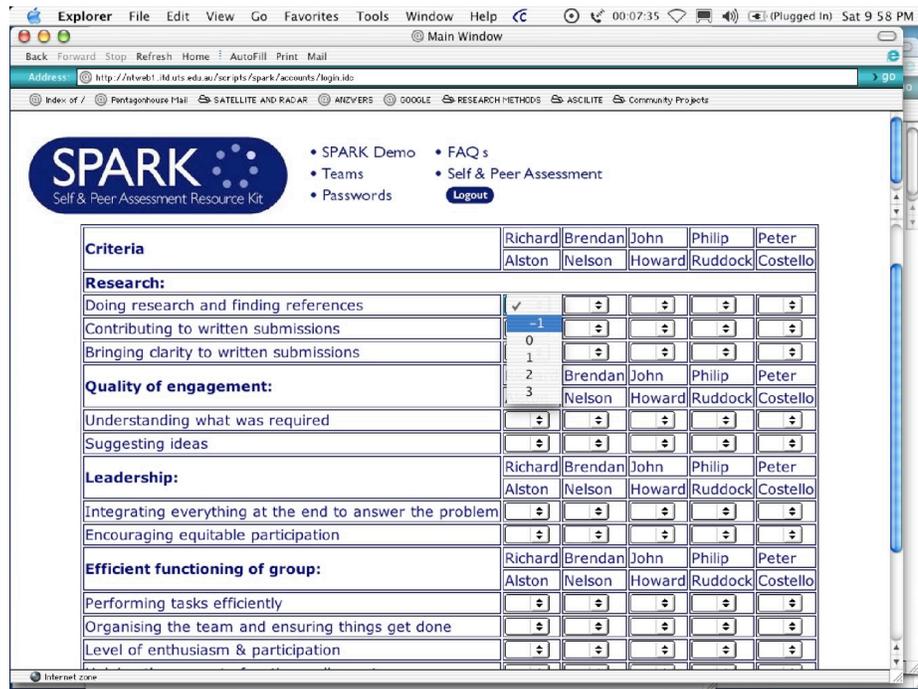
assignments independently. The Re:View system reduces the effect of this on assessment by displaying visual pie and bar charts of criteria categories, to illustrate the attribute development intended through different subject areas and during the progressive years of a course. It also displays students' achievement in those by individual, group, subject, year, and so on.

A Description of SPARK

This has also been a successful e-assessment tool aimed at the development of graduate attributes related to working in groups. It enables students' engagement in criteria-based reflection on their own and their peers' contributions (anonymously), through a Web interface. One of the factors produced by the system is then used to calculate an individual mark based on these self and peer ratings.

It is normally important to ensure that group work does not constitute the majority of a subjects' learning tasks, however, with the SPARK system, problems associated with all group members receiving the same mark are removed because individual marks are generated through the self and peer ratings process. I have successfully used SPARK for up to 70% of a subject's assessment.

Figure 6. Student's view of group rating screen against sample criteria:



Case Study Outline Using SPARK and Re:View

The figures reproduced for this case study were from a design history assignment traditionally assessed using exams and allocated 20% of the total assessment for the subject. The learning task was designed as a group-based online debate project ending in a live debate in the lecture theatre.

First semester, 2nd-year students were randomly formed into 18 learning groups of 5 students each for the task at the beginning of semester, as a method for managing tutorial and studio sessions throughout the 13-week period.

The *Online Debate* ran whilst the other two tasks were being undertaken and spanned a period of 6 weeks in the middle of the semester. There were two online submission deadlines on Sundays at midnight to stimulate online activity and also to avoid other interim deadlines within the subject.

An extract from the briefing document shows the briefing explanation students received:

“Scenario:

Imagine we have just been joined on our UTSONline website by 18 famous typographers / artists / designers. They represent strong views about both the expressive and functional approaches to typography and design.

Some of them are in disagreement and want to have a discussion / argument about their approaches.

However quite a few of them are dead and the others not here in Australia so we have arranged for this to happen using your learning groups as champions of their points of view and philosophy.”

The students were encouraged to “become” the typographer or designer using three devices:

- (a) a photograph of each learning group posted online with their new persona’s name, at the time when they needed to challenge their opposing debate partner;
- (b) an instruction that all online written submissions, and the live debate at the end, were to be written or spoken in first-person;
- and (c) group research of a given persona through five holistic research questions, which I designed, to broaden the information which forms the basis for the debate submissions:

1. **Propositions:** What did they believe and what were they trying to do?
2. **Influences and connections:** Where did they draw their influences from and who were they connected to, for example, artistic movements, etc?
3. **Principles and ways of working:** How did they go about their work, and what did they consider important in the way it was done?
4. **Character/Personality:** What were they like as a person and why do you think they did what they did?
5. **Facts and Figures:** What interesting facts can you find as well as the usual birth/death/education that is available?

Students were encouraged, through video clips of a previous iteration of the project, to be creative in costume and drama, in making their points in the live debate at the end of the task.

The debate, both online and live was structured to reveal the positions of the individuals from history (or living) in reference to a spectrum—from functional, problem-solving design at one end, to expressive, artistic approaches at the other. Students were instructed that they would have to position their persona on a spectrum line as part of the live debate session. They were also asked to reflect on their own position with regard to this spectrum both prior to and after the project.

Functional ←-----→ Expressive

As this learning task was only a 20% part of the subject assessment it was specifically targeted to develop two basic attributes in the context of an online group engagement. The two assessment criteria used to cover all three parts of the task (*Opening Statements*, *Challenging Statements* and *Live Debate*) related to subject-outline learning objectives in the attribute groups of research skills and process skills:

- **Research:** Depth of research in substantiating the points of view, and
- **Process:** The cogency with which arguments and rebuttals were developed.

The development of ability to work in teams was also part of the learning objectives of this task but the second online system, SPARK, was used for this purpose.

The Re:View software was used by the lecturer to mark and give feedback on this task (and the other tasks in this subject). Students logged on to see their group's assessment and feedback online. The staff members involved in the subject were keen to use the software, albeit at pilot stage, due to the timesavings they had experienced in both marking and then publishing grades and feedback online.

The live debate was videoed and both stills and clips from the video were put online as a reminder of the process, whilst group members had 10 days within which to rate each other and themselves using SPARK.

The factors produced from SPARK ratings were then used to modify the group marks to individual marks, based on the self and peer ratings against nine criteria (these were explained and agreed to by the students, with opportunity for comment and amendment, at a session introducing the learning task):

Category	Criteria
Online Submissions	Contributing to the cogency of written submissions
Online Submissions	Doing research and finding references
Efficient functioning	Helping the group to function well as a team
Efficient functioning	Level of participation in the online debate project
Efficient functioning	Performing tasks efficiently
Quality of engagement	Suggesting ideas
Quality of engagement	Understanding what was required
Leadership	Helping decide who does what and when
Leadership	Bringing things together for a good team result

Self assessment (SA) and peer assessment (PA) was new for these first-semester second-year students but categorized, criteria-based assessment had been experienced in other subjects.

There were two different factors produced by SPARK from these ratings, revealing some interesting aspects of this self and peer reflective assessment process. The first is a factor produced from the average of SA as a ratio to the PA. If this factor is less than 1.0, then the student underrated their own performance in comparison with their group members' ratings of them. The second is the SPA factor, which is a combination of the average of self and peer ratings and is the factor used to calculate individual marks through multiplication with the group mark given by the lecturer. The method for SPARK and the formulas for the calculation of factors were based on educational research (Thompson, 2002).

The analysis of overall ratings shown in this table reveals a very responsible approach to rating, in that the -1 rating was only used 12 times and 0 was used only 45 out of 2986 ratings. These ratings are used by SPARK to calculate the factors mentioned and the effect on individuals' marks when multiplying their factor by the group mark can be significant as seen in Table 2.

This table shows an example of the two factors generated by SPARK in the ratings process. If the SPA (Self + Peer Assessment) factor was 1.0 then the group member received the same mark that the group was given by the lecturer (e.g., 62% multiplied by 1.0= 62%). The SPA factor is the one used to modify the group mark, so that if a student has a factor of 0.9 they would only receive 90% of the group mark given for the group task (e.g., 62% multiplied by 0.9 = 55.8%). Therefore, as previously shown in the case of Student 2, the group mark was 75.5 and the student received .86 of that, that is 64.9%.

The SA/PA factors show significant underrating by individuals of their own performance. It is interesting that both factors for Student 1 show that the group rated this team member's performance much more highly than they rated

Table 1. Total numbers of case study ratings for each rating level

Ratings Entered	Total
-1 (detrimental contribution):	12
0 (no contribution):	45
1 (below average contribution):	370
2 (average contribution):	838
3 (above average contribution):	1631
All Ratings	2986

Table 2. SPARK factors showing effects on students' individual mark compared with the group mark given by the lecturer:

SA/PA	SPA	Group Mark given by lecture	Individual Mark GrpMark x SPA
Student 1: 0.82	1.05	80.5%	84.5%
Student 2 0.83	0.86	75.5%	64.9%
0.84	1.04	73.5%	76.4%
0.84	0.84	76%	63.8%
0.86	1	62%	62.0%
0.86	0.97	86%	83.4%
0.87	0.86	71.5%	61.5%
0.88	0.99	75.5%	74.7%
0.88	0.99	75.5%	74.7%
0.89	1.01	75	75.8
0.89	0.99	86	85.1
0.9	0.96	62	59.5

themselves, and their individual mark shifted from a Distinction to a High Distinction (High Distinction is the grade given in the Australian New South Wales banding system when a mark is 85% or above). SPA factors for Student 2 show that the group agreed with this student's low ratings of their own contribution and their individual mark was reduced by 11% from a Distinction (75-84%) to a Borderline Credit (65-74%).

Overall from this group of 90 students, 24 underrated themselves (as in the example of Student 1) with varying degrees of agreement from their peers.

Students reflected that they felt the SPARK process was fair, and that they could relax into group work knowing that individual contributions would be taken into account in the final individual mark. The criteria for the group project were simple, but thought to be appropriate for a task worth 20% and anonymous feedback against teamwork criteria was added to their ongoing profile under the five strands of development: research, concept, communication, process, and professionalism.

Conclusion

The argument for reducing the focus on exams becomes clear in the light of research surrounding the two sets of reasons explored in the first two parts of this chapter:

- Consistent educational research identifies assessment as a powerful driver in directing student learning.
- Exams appear to drive learning down pathways that no educator would support, with deadly side effects and the prospect of “invigilation” costs rising exponentially.
- Graduate attributes are a range of qualities that students need to develop in order for them to survive in a rapidly changing world, whilst contributing positively to it.
- The qualities that are consistently part of graduate-attribute statements therefore must become the core focus of our mainstream assessment systems.

Educational literature poses an important question underpinning assessment:

What do we want students to learn and how does our assessment encourage that learning? Given that part of necessary student learning is that learning itself is

a developmental lifelong process, our assessment processes should themselves give developmental feedback over time.

The two assessment systems described in this chapter are designed to bring the development of generic attributes into mainstream assessment processes, integrated with other more content-focused criteria. However, the introduction of e-assessment systems is not viable without delivering time and flexibility benefits to teaching staff. Change relies on the genuine reflection by curriculum developers and teaching staff, about the real impacts of their assessments on young people. An important rider to the changes suggested in this chapter is that the thorough explanation and careful introduction of an e-assessment system (to both staff and students), is an essential ingredient in its success.

Through early pilot schemes introducing the two systems, one of the most interesting aspects was the opening of a dialogue about assessment processes between lecturers and students. The lecturers have had to explain the reasons for these processes and be far more explicit in the definition of criteria used in the marking of work. The students, on the other hand, have had to relinquish surface approaches through discussing criteria and exercising responsibility in rating their own and their peers' contributions. The benefit to students of developmental feedback across subject boundaries was not studied in this research although the benefits of formative assessment with accruing summative marks was positively noted by teachers.

It is hoped that the reasons and references relating to the use of exams will assist in their gradual demise as a dominant feature of the assessment landscape, and that the case studies and descriptions provided will encourage a serious attempt to implement less stressful and more useful assessment processes.

References

- Assiter, A. (Ed.). (1995). *Transferable skills in higher education*. London: Kogan Page.
- Australian Council for Education Research (2001). Higher Education Innovation program. *Graduate Skills Assessment test*. DETYA. Canberra, Australian Capital Territory: Author.
- Australian Council for Education Research (2002). GSA report. Canberra, Australian Capital Territory: Author.
- Australian Government Publishing Service (1997). *Youth suicide in Australia: A background monograph* (2nd ed.). Author.

- Australian Universities Quality Agency (2004). *Annual report 2003*. Retrieved January 1, 2005, from <http://www.auqa.edu.au>
- Biggs, J. B. (1995). Learning in the classroom. In J. Biggs & D. Watkins (Eds.), *Classroom learning: Educational psychology for the Asian teacher* (pp. 147-166). Singapore: Prentice-Hall.
- Bunting, C. (2000, February 18). Spectre of exam drug tests looms. *Times Educational Supplement*, 4364, 3.
- Department of Education, Science, and Training (2002). *Striving for quality: Teaching, learning and scholarship* (Vol. 6891, HERCO2A). Canberra, Australia Capital Territory: Author.
- Economist*. (2004, September 18). Supercharging the brain. *Economist*, 372(8393).
- Garner, S., & Duckworth, A. (2000). In C. Swann & E. Young (Eds.), *Reinventing design education in the university* (pp. 206-212). Perth, West Australia: Curtin University of Technology.
- Martin, A. (2003). *2002 institutional audit reports: Analysis and comment*. Retrieved January 13, 2005, from Australian Universities Quality Agency Web site: <http://www.auqa.edu.au/qualityenhancement/occasionalpublications/>
- Marton, F., Hounsell, D., & Entwistle, N. (Eds.). (1984). *The experience of learning*. Edinburgh, Scotland: Scottish Academic Press.
- Marton, F., & Saljo, R. (1976). On qualitative differences in learning, II—Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology*, 46, 115-127.
- Mayer, E. (1992). Canberra, Australian Capital Territory: Australian Government Publishing Service.
- New York Times* (1999, November 17). NYT Education Supplement.
- Pritchett, N. (1999). Effective question design. In S. Brown et al. (Eds.), *Computer-assisted assessment in higher education*. London: Kogan Page.
- Ramsden, P. (1992). *Learning to teach in higher education*. London: Routledge.
- Ramsden, P., & Entwistle, N. J. (1981). Effects of academic departments on students' approaches to studying. *British Journal of Educational Psychology*, 51, 368-383.
- Salchow, G. (1994). in M. Bierut (Ed.), *Looking closer 1: Critical writings on graphic design*. New York: Ailsworth Press.
- Sankey, M., & Lawrence, R. (2003). *Suicide and risk-taking deaths of children and young people*. Sydney: New South Wales Commission for Children and Young People, Child Death Review Team, & Centre for Mental Health.

- Shermis, M., Mzumara, H., Olson, J., & Harrington S. (2001). Online grading of student essays: PEG goes on the World Wide Web. *Assessment & Evaluation in Higher Education*, 26(3), 247-259.
- Smith, L., & Sinclair, K. E. (2000). Transforming the HSC: Affective implications. *Change: Transformations in Education*, 3(2), 67-79.
- Steinberg, R. J. (1997). *Thinking styles*. New York: Cambridge University Press.
- Suzuki, D. (1998). *Earth time*. Toronto, Canada: Stoddart Publishing Company Ltd.
- Sydney Morning Herald* (2004, October 28). Stress put to the test. Australia.
- Thompson, D. (2002). In A. Davies (Ed.), *Enhancing curricula: Exploring effective curriculum practices in art, design and communication in higher education* (pp. 360-392). London: Centre for Learning and Teaching in Art and Design.
- Toffler, A. (1980). *The third wave*. New York: William Morrow.
- Williams, L., Vablais, W., & Bathiche, S. (2004). *US Patent No. 559746*. Washington, DC: U.S. Patent and Trademark Office.
- Wright, P. (1995). Canberra, Australian Capital Territory: The Higher Education Quality Council, Quality Enhancement Group.
- Ying Liu, L. (2002, August 1). In *Family Practice News*, 32(15).