

Investigating Factors Affecting the Uptake of Automated Assessment Technology

Carl Dreher

Curtin Business School, Curtin University, Perth, Australia

c.dreher@curtin.edu.au

Torsten Reiners

***Curtin Business School, Curtin University, Perth, Australia, and
Institute of Information Systems, University of Hamburg,
Germany***

reiners@econ.uni-hamburg.de

Heinz Dreher

Curtin Business School, Curtin University, Perth, Australia

h.dreher@curtin.edu.au

Executive Summary

Automated assessment is an emerging innovation in educational praxis, however its pedagogical potential is not fully utilised in Australia, particularly regarding automated essay grading. The rationale for this research is that the usage of automated assessment currently lags behind the capacity that the technology provides, thus restricting the pedagogical benefits for students, job satisfaction for staff, and quality assurance and financial benefits for universities. The exploration of the different perspectives of stakeholders regarding their needs and expectations of automated assessment shows the high-stake game of surviving and thriving. Inter alia, students value education as a means to gain employment, educators' value education as their means of employment, and universities value education as their means of existence. The various facets of this interplay are described, including the value of knowledge and learning (pedagogical considerations); society's system for measuring worth and exchanging value (economic considerations); socio-technological evolution and emerging innovations in educational assessment, and the adoption and resistance thereof.

We contribute to the field a national survey that investigated and reveals potential reasons for this sub-optimal utilisation by exploring the candid perceptions of Australian university students and

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Publisher@InformingScience.org to request redistribution permission.

teachers regarding automated assessment. Students and staff from Australian universities were invited to participate in an anonymous online survey. Two hundred and sixty five people completed the survey. The analysis utilised a mixed-method approach, combining quantitative and qualitative research paradigms. We present a sub-set of the survey results; this sample's use of automated

assessment is mapped via quantitative data, and respondents' perceptions of automated assessment are examined via a constructivist grounded theory analysis of short answer responses. Specific aims of this research were to survey the human and automated assessment practices in Australian universities (specifically, the educational roles of users (e.g., students, educators, management, IT-support, HR-administration), assessment types used, frequency and mode of use (individual vs. group), mode of awarding marks (individual vs. group), and mode of marking (human vs. computer)), examine what 'automated assessment' means to respondents, explore respondents' impressions of automated essay grading, and survey respondents' use of automated assessment technologies.

This sample's concept of 'automated assessment' is multifaceted but can be summarised by four categories of themes that describe automated assessment as involving computers or technology, electronic media, 'marking' or 'assessing', and degrees of automation. Understanding stakeholders' perceptions of automated assessment is instrumental in promoting its use and benefits. The implications from the survey transfer the focal point of our research agenda to building a profound methodology for future automated assessment and, in particular, to delineate a context-specific process of integration for the relevant stakeholders (e.g., students and staff). Most important, the marketing portfolio calls for the following, among others, to be demonstrated across various educational domains: robustness regarding marking accuracy and time (compared to human marking), ease of use for inexperienced users (students as well as teachers), and economic savings. The automated assessment has to meet pedagogical benefits first, but also has to provide commercial benefits. Especially the latter one is related to the reputation of the university, which is judged in society by successful graduations and post-graduation job performance. High quality education is the harbinger of new student enrolments and the attractor of research and business projects, including endowments and sponsorship, in total increasing the financial scope. In order to increase adoption of these technologies we have to learn more about stakeholders' concerns, desires, and speak their language in order to have them move-in and feel at home using educational innovations such as automated essay grading. The contribution provides an inside view and understanding of where we are with respect to automated assessment and what is still part of the roadmap for a successful integration.

Keywords: Educational assessment; automated assessment; automated essay grading; online survey; constructivist grounded theory analysis.

Introduction

Assessment is integral to education. Assessment involves both determining students' status in the process of learning (summative assessment) and generating meaningful feedback to learning progress (formative assessment) (Black & Wiliam, 1998). Therefore, assessment guides the teaching and learning process by providing reciprocal feedback to both educators and students so that they may improve in their respective tasks. While the kind of assessment that is used may depend on the pedagogical model being implemented, all types of assessment place demanding requirements on educators. Such demands include the design, development, execution, and evaluation of assessments, and the provision of feedback. From a pedagogical perspective, assessment needs to be an accurate and informative indicator of students' learning. From a pragmatic perspective, the quality of assessment is often mitigated by economic pressures that result in educators being required to do more with less. The 'do more' often involves meeting strategic goals to improve teaching and learning outcomes and research targets (ranked nationally and internationally), in addition to multifaceted administrative tasks. The 'with less' often involves restricted budgets, and (due to increasing task demands) less time available in an educator's working week (often requiring staff to work unpaid overtime). These pressures can lead educators to use summative assessment methods in ways that limit the provision of quality feedback to benefit student learn-

ing (Butler & Winne, 1995; Nicol & Macfarlane-Dick, 2006). However, if we as educators focus too much on summative assessment, we may fall short of satisfying community expectations relating to education system outcomes.

In this paper, we discuss the role of academic assessment in the educational process at universities from various perspectives (i.e., regarding pedagogical, economic, and socio-technological influences). It is the confluence of these issues that motivated our research into automated assessment technologies and stakeholders' perceptions of these technologies. Our research is also informed by our experience of staff and students being preoccupied with the term *automated* and their underlying assumptions about what this implies for marking and quality of education (i.e., regarding essay grading).

In the following section, we explore the different perspectives of stakeholders regarding their needs and expectations of automated assessment. In subsequent sections we present the method of investigation and the results (focusing on both respondents' use and understanding of automated assessment). We conclude the paper by reviewing the relevance of these findings for university students, educators, and management who in their various roles are faced with the demands of adapting to socio-technological evolution and producing educational outcomes that are valued by a global society.

Different Perspectives of Educational Assessment

By virtue of measuring valued outcomes, educational assessment is situated at the centre of a high-stakes game of surviving and thriving. Inter alia, students value education as a means to gain employment, educators' value education as their means of employment, and universities value education as their means of existence. In the following section, we discuss various facets of this interplay, including the value of knowledge and learning (pedagogical considerations), society's system for measuring worth and exchanging value (economic considerations), socio-technological evolution and emerging innovations in educational assessment, the adoption and resistance thereof, and our rationale for studying the Australian university community's perceptions of automated assessment technologies.

Pedagogical Considerations of Assessment

Assessment depends on the educational learning theory employed (e.g., behaviourism, cognitivism, constructivism, connectivism, multi-medial, social-cognitive, post-modern), it is linked to the educational objectives, it reflects the knowledge/skill domain being assessed, and it is heavily dependent on the purpose – summative assessment versus formative assessment (Gipps, 1994). Here, we concentrate on the purpose of the assessment and the type of knowledge domain being assessed. Bloom's taxonomy of educational objectives (Bloom, 1956) can be used to guide the setting of objectives and consequently their assessment in both summative and formative modes.

Learning is a process comprising the three functional domains of cognitive (thinking), affective (feeling emotion), and psychomotor (doing), with assessment being primarily in the cognitive domain (Bloom, 1956). The cognitive domain is further divided by Bloom's (1956) well-accepted taxonomy into three categories (each with two hierarchical levels): recall, interpretation, and problem-solving. The recall category has the lowest learning process requirements and involves memorizing and remembering without necessarily understanding (i.e., the knowledge level) and being able to restate or summarize given exercises (i.e., the comprehension level). The second category, interpretation, demands transfer of theory to practical situations (i.e., the application level) and the identification of relevant components and logic in the learning material (i.e., the analysis level). The last category, problem-solving, is about combining information to produce new products (i.e., the synthesis level) and to make decisions with respect to creating an impact in

a given application (i.e., evaluation level). Bloom's taxonomy was recently revised (see Anderson & Krathwohl, 2001; Krathwohl, 2002).

For educators, assessment tasks from the first category (recall) require the least preparation because evaluating memorized knowledge requires simple assessments like multiple-choice or direct questions. In contrast, assessment of higher levels of knowledge (e.g., tasks that require students to demonstrate synthesis and evaluation) requires assessment formats that are more demanding for educators to develop and assess and to provide feedback on (e.g., essays and practical projects).

From an alternative approach, the purpose for which assessment is used in education (summative vs. formative) is integral to its implementation. For example, it can affect the value attached to assessment (e.g., students' perceptions of its relevance and their motivation to achieve). Assessment can be conceptualised as either an outcome of, or an element in, the educational process. When considered as an outcome (as with summative assessment), assessment affords various benefits endowed by society's value of the result (e.g., students pass a unit or course and qualify for a certificate, scholarship, further study, or employment). When considered as an element in the teaching and learning process, formative assessment can be regarded as the interaction between learner and teacher during which the learner's ideas are characterised and communicated to the teacher, and the teacher's reflective assessment thereon is offered to underpin further learning by the student. Educational assessment is ideally aligned with pedagogical theory and praxis. Inherently, it is a social process to which we attached value – value that is unavoidably measured in economic terms.

Economic Imperatives of Education and Assessment

As organisational entities, universities are multifaceted. They are simultaneously cultural-epistemological (research) institutions, pedagogical (teaching) institutions, and economic (business) entities. Our discussion focuses primarily on the interplay of the pedagogical and economic demands on universities. Common strategic goals for universities are to maintain or increase enrolments and to maintain or improve rates of course completion. However, today's potential students are often opting to work full-time rather than to study, especially where there are highly paid jobs on offer such as in the resources sector of Australia's Northwest. In attempts to maintain growth many universities are recruiting overseas students. This is putting pressure on class sizes, creating challenges resulting from culture-clash and English language deficit, and of course, increasing marking loads.

Already we have seen union action against increased workloads and incommensurate remuneration. However, in order to maintain course completion rates, with larger classes and lower entrance standards, teachers are hard pressed to provide the formative feedback that is required to enhance student outcomes. Lowering academic standards can be one unfortunate outcome of this.

In order to cope with increased marking loads, teachers use, *inter alia*, the following options: working more hours (working harder), working more efficiently (working smarter), or they can lower the quality of work (working quickly, marking less accurately and/or with less formative feedback). Given the trend of increasing workloads, working harder can lead to decreased employee satisfaction and wellbeing. However, working smarter is likely to benefit students via the maintenance of academic standards through summative assessment and the improvement of knowledge and skills through formative assessment. Working more efficiently will also benefit the institution via a good reputation resulting from student satisfaction and competent graduates performing well in the industry.

One method of working smarter is to employ sessional or contract staff to both teach and mark. This may impact on quality of teaching and marking. Another method of working smarter is to

use technology to augment our human capacities. In this regard, automated assessment can provide quick, reliable, cost-effective means of assessing large numbers of students and has the added benefit of freeing the educator to teach students and to provide them with formative feedback for both directive and facilitative purposes (see Black and Wiliam, 1998, for a discussion of the benefits and forms of feedback). Oftentimes this automation of assessment means educators turn to multiple-choice tests (e.g., computerised in formal testing labs or provided online for self-study). These have certain pedagogical benefits (e.g., allow students to self-test and gain instant feedback to facilitate their revision of misconceptions) but are also criticised for assessing surface learning (e.g., recognition and/or recall) and providing limited capacity to assess depth of knowledge or application of industry-relevant skills (Ramsden et al., 1993). Another approach to automated assessment is that of automated essay grading. Essays facilitate assessment of deeper learning processes that require students to synthesise and/or analyse domain-relevant information. However, essays are labour-intensive to mark manually. Fortunately technology has advanced sufficiently to allow essays to be automatically graded as accurately as humans under certain conditions (Williams, 2006) and provides interactive formative assessment (H. Dreher, 2006). In short, by working smart through using various state-of-the-art automated assessment technologies, we contend that both the pedagogical and economic goals of the university can be met.

Socio-Technological Change and Emerging Innovations in Educational Assessment

Education has traditionally been the gateway to gainful employment in industry. However the rate of technological and consequential societal change is increasing, which is resulting in a generation of ‘digital native’ students who are engaging with an educational system that was designed in a pre-digital age and is comparatively slow to adapt (Prensky, 2001). Consequently, students may find less relevance in their education and must teach themselves modern life-skills through participating in the networked society/economy and must learn industry relevant skills and knowledge on the job. For universities to remain the gateway to the future, they must keep pace with, and be at the leading edge of, technological innovation and socio-cultural evolution.

Regarding assessment of educational outcomes, technological advances now permit and support:

- moving beyond online multiple-choice testing to assessment of higher-order learning outcomes via use of automated essay grading;
- improving formative feedback by using automated assessment technologies to decrease submission-feedback latencies and freeing educators’ resources to interact with students to provide formative feedback and to assist training to mastery (as a cohort, in groups, and individually) (H. Dreher, 2006);
- improving self-assessment through online assessments that provide quick/immediate interactive feedback (self-assessment is integral to students learning how to learn independently, which is a valuable skill in life generally as well as academia and industry specifically);
- integrating assessment with the learning process in a comprehensive system that is intrinsically motivating to students (e.g., educational gaming innovations such as Alice or 3D virtual world constructivist learning applications such as Students@Work and the Online Automated Assessment Laboratory) (C. Dreher, Reiners, Dreher, & Dreher, 2009);
- involving students in the design of assessment tasks and assessment criteria (Hulsart & McCarthy, 2009) to enhance ownership of and motivation in the learning process and educational outcomes, and;
- involving students in self-assessment during group-work (see for example Thompson & McGregor, 2009) for formative reflection and to inform educators’ summative assessment.

Students, and the younger generations in general (called *inter alia*, digital natives, Generation Y, Generation Z, and the Net Generation), are products of this technology-integrating culture (Prensky, 2001). Therefore, keeping abreast (or ideally leading) this social-technological innovation is the Gateway to the Future for universities. It is therefore important to examine the adoption and resistance of educational innovations in universities.

Adoption and Resistance of Automated Assessment in Educational Contexts

As an IT-enabled activity, automated assessment technology is both a well-established technology (e.g., computer-scored objective tests) and an emergent technology (e.g., automated essay grading). Indeed such emergent technologies go beyond the automation of repetitive processes, to the realm of analysing meaning through the semantic processing of text. In its various forms automated assessment is both welcomed and resisted by various stakeholders in the educational setting. For example, Livingston and Condie (2006) describe their implementation of an online learning program in secondary schools in Scotland. They report that the students readily utilised the system, whereas teachers did not. In discussing their experiences, Livingston and Condie note that one reason teachers can resist the use of online learning programs is that it facilitates a more independent approach to learning for students, which can be challenging for some teachers who see their role as being experts who transmit knowledge as “the sole leaders of learning in the classroom” (p. 155). Consequently, such teachers can be “unwilling to relinquish control of the learning process” (p. 155). The main issues here appear to be job-roles and control. In relation to assessment, if educators see their capacity to mark assessments as being integral to their role, then they may fear being replaced by automated assessment software. Thus stakeholders’ perceptions of automated assessment are an important consideration in examining its uptake, and this informs our rationale and aims for the present research.

Rationale and Aims

The reciprocal influence of society and technology (as evident in the economy) has brought us to the point where on one hand, economic pressure and technological advancement have contributed to the development of automated assessment technology, and on the other hand, utilising automated assessment technology requires change in academic culture regarding assessment. This change is currently slow. Automated assessment technology has been developed due to socio-economic pressures. The reality that universities are run as businesses leads to certain factors that challenge educators, including that large classes are common, workloads are increasing, and the importance of quality assurance (i.e., quality management) of education and assessment. For automated assessment technologies to be utilised, a change is required in the academic culture surrounding assessment practices. Indeed, automated assessment has the power to beneficially change the socio-technological process of assessment in educational organizations. However, currently such change is resisted.

To be more contextually specific, this study emerged from our observation that at Curtin University the academic community tends not to avail itself of the state-of-the-art in automated assessment technology and thus misses some of the benefits available. Specifically, we established a university-based automated essay grading service and research centre, the Automated Assessment Laboratory (AAL) at Curtin Business School (H. Dreher, Dreher, & Reiners, 2008). We subsequently noticed that the majority of academics were reluctant to make use of the opportunity to trial the automated essay grading service provided by the AAL.

At Curtin, we observe that the use of fixed-choice format questions (multiple-choice and true/false questions) is commonplace. They are inexpensive and relieve the workload of human markers. They also offer a restricted range of learning outcomes that can be assessed (surface

learning, recognition and/or recall of facts, rather than explication of reasoning and communication of understanding, discussion of topics, and other means of showing that higher-order learning has taken place). While multiple-choice tests are a well-established assessment practice, an emerging trend at Curtin is to allow students (before submission) or staff (after submission) to use plagiarism checking technologies (e.g., Turnitin). This plagiarism detection system uses word and sentence level text comparison methods to check for duplicates, and it provides timely and relevant feedback to students on their skills at paraphrasing and referencing, while also reducing the workload for staff by automated detection of plagiarism and assisting annotation of assignments. However, Turnitin does not make evaluations or assessments of the worth or value of a submitted assignment (file of text) which still requires human marking by the usual effortful and costly process, especially in large cohorts.

The rationale for this research is that at Curtin (and we suspect, throughout Australian universities) the usage of automated assessment currently lags behind the capacity that the technology provides, thus restricting the pedagogical benefits for students, job satisfaction for staff, and quality assurance and financial benefits for universities.

Specific aims of this research were to:

- survey the human and automated assessment practices in Australian universities – specifically, the educational roles of users (e.g., students, educators, management, IT-support, HR-administration), assessment types used, frequency and mode of use (individual vs. group), mode of awarding marks (individual vs. group), and mode of marking (human vs. computer);
- examine what ‘automated assessment’ means to respondents;
- explore respondents’ impressions of automated essay grading;
- survey respondents’ use of automated assessment technologies - in particular, the forms used, programs/services used, educational roles, contexts, and purposes of use, appraisal of the usefulness of human-assessment vs. computer-assessment (in general, and by specific types of automated assessment), perceptions of the depth of understanding assessable by each type of automated assessment, preference for use by each type of automated assessment, exploration of the pros and cons of automated plagiarism checking and automated essay grading, explication of the desired elements of automated assessment technologies by staff who have used them, and exploration of the barriers to use of automated assessment technologies by staff who have not used them.

Method of Investigation

Measure

Self-report questions were administered via an anonymous online survey using the software *EFS Survey* by GlobalPark (Location of the survey: http://ww3.unipark.de/uc/hh_fak2_treiners/3581). The questions were written by the researchers based on their knowledge of and experience with automated assessment as well as educational assessment in general. Programmed filters were used to present respondents with only those questions that were relevant to them in order to decrease response fatigue and attrition. For example, some questions were worded differently for staff and students, and some questions were omitted if respondents had no prior experience with automated assessment (either in general or with specific types of automated assessment). Therefore, all questions comprising the survey were not necessarily presented to each respondent but only those that were relevant to each individual.

Procedure

Forty universities in Australia were contacted via email requesting organisational consent from senior management to participate in the survey. Five universities consented to participate. They spanned three states (Queensland, Victoria, and Western Australia). All five universities gave consent to contact their staff, while only three of these universities consented to contact their students. Staff members were contacted via email distribution lists (at 4 universities) and online newsletters (at 1 university). Students were contacted by email distribution lists (at 1 university), postings on student websites (at 1 university), and a method not stated (at 1 university). This study was approved by the Curtin University Human Research Ethics Committee.

Participants

A total of 461 respondents began the survey, however 196 (42.5%) did not complete it. The results reported here are from the 265 (57.5%) who did complete the survey (referred to as survey completers). The 265 survey completers included 106 (40.0%) males and 159 (60.0%) females. To be eligible for the study, participants needed to be 14 years-of-age or older. Respondents' demographic descriptors were rated categorically; we report here the frequency and percentage of respondents per category for the following variables: age, highest level of education obtained/completed, country of education, educational role(s) (non-exclusive categories), educational role (mutually exclusive categorisation), Australian state/territory of residence, country of birth, and ethnicity (modal categories are given in bold).

All 265 survey completers specified their educational role (mutually exclusive categorisation) with **staff only (101; 38.1%)**, *student only* (90, 34%), *both staff and student* (74, 27.9%). This was further elaborated as they had to mark their educational roles using non-exclusive categories. Here, most hold the role **student (195, 73.6%)**, further roles are *educator* (181, 68.3%), *management* (28, 10.6%), *information and communication technology technician* (15, 5.7%), *administrative/human resources staff* (15, 5.7%), and *other* (23, 8.7%).

Details about the age was given by 258 participants, using categories as we required only an estimate about the distribution for our analysis. According to the majority being *staff only*, the modal of age is **41-50 (86, 32.5%)**; followed by *31-40* (53, 20.0%), *51-60* (49, 18.5%). Thus, with participants *>60* (5, 1.9%), we had 72.9% above 30. Furthermore, we had the following categories: *14-15* (1, 0.4%), *16-17* (19, 7.2%), *18-20* (19, 7.2%), *21-25* (22, 8.3%), and *26-30* (19, 7.2%).

With **77 (29.1%)** of 256 respondents, the highest level of education was **Ph.D.**, followed by *secondary school* (56, 21.1%), *master degree* (55, 20.8%), *bachelor degree* (36, 13.6%), *honours degree* (18, 6.8%), *2 year/associate degree* (13, 4.9%), and *primary school* (1, 0.4%). The survey was restricted to Australia, so 260 survey completers stated their country of education to be **Australia (217, 81.9%)**, *UK* (13, 4.9%), and *other* (30, 11.3%).

In addition, we collect further demographic data to learn more about the participants and their background. We do not intend to link this information to the other results in this paper as it is not relevant for our objectives. Nevertheless, we report it for the sake of completeness. The variables are country of birth (257 respondents; **Australia (183, 69.1%)**, *UK* (19, 7.2%), and *other* (55, 20.7%)), Australian state/territory of residence (256 respondents; **Western Australia (119, 44.9%)**, *Victoria* (83, 31.3%), *Queensland* (47, 17.7%), and *other* (7, 2.7%)) as well as ethnicity (252 respondents; **Australasian/Oceanian (140, 52.8%)**, *North-West European* (69, 26.0%), *Southern and Eastern Europe* (18, 6.8%), and *other* (25, 9.5%))

Analysis

This study was designed using a mixed-method approach (see Tashakkori & Teddlie, 2009), consequently analysis of the survey data comprised both quantitative and qualitative methods. Quantitative descriptive statistics (frequency charts and modes) were used to summarise fixed-response format questions (multiple-choice check box, Likert-type rating scales). A qualitative analytical method was used with the open-ended responses to the short-answer questions. Specifically, a constructivist grounded theory method of analysis was employed (Charmaz, 2000). This is an inductive (bottom-up) approach that sought to understand each respondent's meaning and conceptual understanding in order to generate an integrative, aggregate conceptual framework (a schema or theoretical model of this sample's multifaceted conception of what 'automated assessment' is). In addition, through the analyses we sought to understand respondents' conceptions and impressions of automated assessment rather than the 'objective truth' about automated assessment technologies and methodologies.

The present qualitative analysis comprised two main processes: coding themes and theory building by grouping themes into categories. The first main process, open coding (also known as, initial coding or line-by-line coding, Charmaz, 2000), aimed to discover the meanings given by respondents for each question (e.g., *Please briefly describe what the term 'automated assessment' means to you.*). Coding was applied to the open-ended responses using two steps. In the first step of coding, 'open coding' was used to generate potential themes (by the researchers reading participants' responses and listing potential themes based on the key terms and concepts within the responses). In the second step of coding, 'line-by-line coding' was used to specify which themes represent the meanings provided by each respondent. Throughout this second step, themes were modified by use of the constant comparative method by contradistinction between themes (within and between respondents) via the use of exemplar excerpts (quoted text segments that typify a theme) and negative cases (responses that highlight problems with themes that indicate a need for revision of the themes) (Ryan & Bernard, 2000). More than one theme could occur per response. Respondents' demographics, educational role, and prior use of automated assessment were considered in the analysis. When coding, the researchers distinguished the explicit meanings of responses from any tacit meanings that they identified. By identifying explicit meanings as themes, this analysis aimed to identify concepts/meanings grounded in respondents' data in order to limit bias from the researchers' sensitising concepts (i.e., preconceived notions of the field comprising "extant theories or our own beliefs" Charmaz, 2000, p. 515). By also identifying tacit meanings as themes, this analysis aimed to acknowledge broader contextual factors and avoid the limitations of a surface-level and literal representation of the data.

The second main process of this qualitative analysis was theory building. This involved grouping themes into categories and generating a theoretical analytic framework to synthesise and explain the data (Charmaz, 2000). Specifically, the theoretical framework described the interrelationships of themes to form groups (i.e., categories of themes), which themselves explained the main concepts comprising this sample's understanding of 'automated assessment'. This process was facilitated by the question '*How are these themes related to each other?*' and proceeded in two main steps: focused/selective coding (selecting the most frequently occurring themes to be considered in the subsequent generation of categories of themes), and theory building (specifying categories of themes and developing these categories into theoretical constructs by describing their qualities and interrelationships). Here, the constant comparison method was used to compare data and themes within a category (central tendency of categories) and distinctions and relationships between categories (divergence and interrelationship of categories).

The nomenclature used herein for the observed frequencies of respondents, themes, and groups of themes (i.e., categories) is as follows:

- Frequency of respondents: the number of people in the total sample (N) and sub-groups (n);
- Frequency of theme occurrence: the number of times a theme is observed/coded across all respondents; frequency of theme occurrence is also the number of respondents contributing to a given theme;
- Frequency of category-theme occurrence: the number of times each theme was observed in the total sample (see frequency of theme occurrence above) summed across all themes in a given category of themes.

In short, open-ended responses were qualitatively analysed by identifying/coding themes and grouping these into categories. For each theme, respondents were counted only once. However an individual's response could generate more than one theme. Furthermore themes could contribute to more than one category. Therefore the total frequency of respondents (N) is less than the frequency of theme occurrences summed across all themes, which in turn is less than the frequency of category-theme occurrences summed across all categories.

Results and Discussion

The results are presented and discussed according to three main topics: (a) the use of assessment in general (considering marking done by both humans and computers), (b) the use of automated assessment (regarding educational roles, contexts and purposes of use), and (c) the meaning of 'automated assessment'.

Use of Assessment in General: Marking by Humans and Computers

Respondents were asked a variety of questions regarding their assessment practices in general (i.e., considering automated and/or human assessment). Firstly, respondents were asked to indicate the types of assessments they had previously used/experienced. Secondly, they were asked to rate how often they had used/experienced each type of assessment on a 4-point Likert-type scale (ranging from *rarely* to *most of the time*). Thirdly, for each type of assessment they had used/experienced, respondents answered questions about the method of marking – by human, computer, or both human and computer. Figure 1 presents the results of these three topics (i.e., type of assessment used, frequency of use, and method of marking) by depicting the number (and percentage) of respondents endorsing each response option. In brief, *essays and reports* were used by the greatest number of respondents, and these were used *most of the time*. The most commonly used method of marking was *human grading* for each assessment type (except for multiple-choice questions, which were most often *marked by computers*).

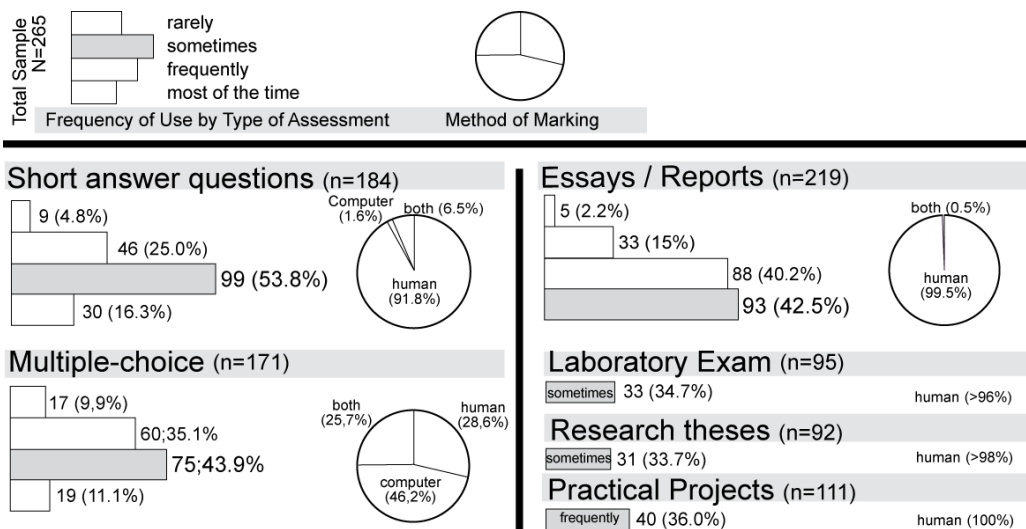


Figure 1: Type of Assessment Used, Frequency of Use, and Method of Marking by number (and percentage) of respondents.

Note. These data refer to the assessment types that respondents had previously used/experienced - n (%) = number (and percentage of the sub-sample who had used each type of automated assessment), and for the three least commonly used types of assessment (i.e., laboratory exam, research theses, and practical projects) modal categories for frequency of use and method of marking.

Additionally, staff members who had not used forms of automated assessment before were asked the question, *Might any of the following be useful in assisting educators to use and benefit from automated essay grading?* From $n = 38$ respondents who answered this question, the number of staff (and percentage of this sub-sample, $n = 38$) who endorsed each response-option were: $n = 26$ (68.4%) for *running a free trial of the automated essay grading in parallel to my normal marking*; $n = 23$ (60.5%) for *seeing results of a survey supporting the reliability/validity of automated essay grading*; $n = 20$ (52.6%) for *being aware of the benefits of automated essay grading*, and; $n = 5$ (13.2%) respondents suggested *other* options, such as providing subject-specific examples, training and support.

Use of Automated Assessment: Educational Roles, Contexts, and Purposes

Respondents were asked about the educational roles, contexts and purposes for their use of automated assessment. Figure 2 presents these data (frequency and percentage by category) for educational roles (both non-exclusive and mutually-exclusive categories), contexts, and purposes for using automated assessment in the sample ($N = 265$). For each variable (i.e., horizontal bar), the modal category is in bold. In summary, these data indicate that (a) 22.6% of the sample had not used automated assessment technology before, (b) 35.5% of the sample had used automated assessment as staff members (and not students), (c) 34.3% of the sample had used automated assessment as students (and not staff), (d) automated assessment had been used (with comparable frequency) in a wide range of educational contexts, and (e) most frequently, automated assessment was used for both summative and formative purposes, rather than either one exclusively.

Uptake of Automated Assessment Technologies

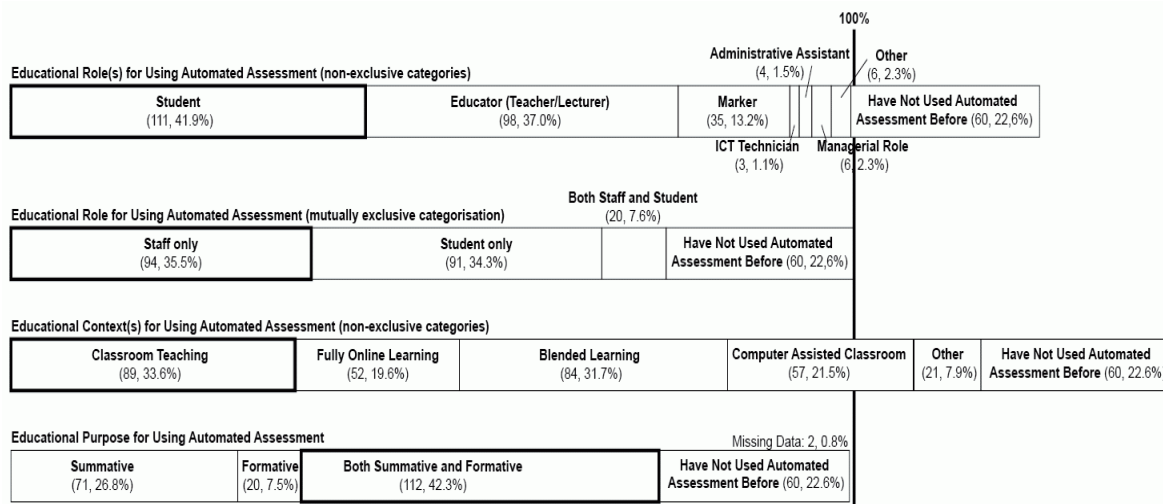


Figure 2: Educational roles, contexts, and purposes for using automated assessment.

Note. Some variables sum to > 265 (100%) of the total sample ($N = 265$) because the response options for some variables were not mutually exclusive.

The Meaning of ‘Automated Assessment’

Respondents were asked to type a short-answer response to the prompt “*Please briefly describe what the term ‘automated assessment’ means to you*”. These responses were processed using a constructivist grounded theory analysis that was conducted at two levels: themes (first-order analysis) and groups/categories of themes (second-order analysis). The results are presented below and are discussed with regard to the ‘central tendency’ of themes and their groups/categories. To consider the ‘central tendency’ of this sample’s multifaceted understanding of automated assessment, we can examine the most frequently coded themes and groups/categories of themes by specifying the most frequently coded themes (first-order level of analysis) and identifying which categories account for the majority of the sum total of category-theme occurrences across the 11 categories.

First-order level of analysis: Themes

At the first-order level of analysis, 65 themes were identified (57 explicit, 8 implicit). However, space does not permit their presentation here and no small number of themes can sum up the majority of theme occurrences. In short, when this sample of respondents were asked what ‘automated assessment’ means to them, a wide variety of features were reported, but one description (i.e., theme) was mentioned relatively more frequently than others, that automated assessment involves ‘a computer program marking an assessment’. However this description accounted for only 13.01% of the total percentage of theme occurrences. Therefore to make sense of these data it is informative to consider a higher-level of abstraction, the second-order level of analysis, groups/categories of themes.

Second-order level of analysis: Categories of themes

Table 1 presents a multifaceted schema (i.e., concept / definition / description) of ‘automated assessment’ based on the various meanings expressed by this sample, as interpreted in the present analysis. **Table 1** lists and describes the 11 groups/categories formed by the 65 explicit and implicit themes. The categories are listed in descending order of frequency (and percentage) of category-theme occurrences. The term ‘category-theme occurrence’ describes the number of times

each theme was observed in the total sample (i.e., frequency of theme occurrence) summed across all themes in a given category (hence called category-theme occurrence). Each theme could contribute to more than one category. The specific themes comprising each category are too numerous to list within the space available for this paper.

Table 1: A schema of ‘automated assessment’ comprising categories of themes

Category (i.e., group of themes)	Freq. (%)	Description - each category comprises themes describing:
Agent Performing Assessment (APA)	251 (21.7%)	The agents (described, implied, or inferred) to be performing the automated assessment.
Medium of Assessment (MoA)	237 (20.4%)	The medium in which the automated assessment process was conducted.
Action Performed (AP)	227 (19.6%)	The action (key verb) described in the process of automated assessment.
Degree of Automation (DoA)	185 (16.0%)	The degree of automation in the ‘automated assessment’.
Elements Assessed (EA)	63 (5.4%)	The elements assessed in the automated assessment process.
Appraisal of Value (AoV)	62 (5.3%)	Respondents' value appraisals of automated assessment.
Uncertainty (U)	51 (4.4%)	Uncertainty on the part of (a) the researchers (regarding the meaning of responses or appropriate grouping of themes) or (b) the respondents (regarding the meaning of ‘automated assessment’) category also includes themes reflecting confusion and misunderstandings communicated in respondents' answers.
Question / Response Format (QRF)	30 (2.6%)	The response formats cited by respondents.
Communication of Result / Feedback (CoRF)	26 (2.2%)	How results and/or feedback are communicated.
Type of Result (ToR)	23 (2.0%)	The type of result produced by the automated assessment process.
Process Description (PD)	4 (0.4%)	Respondents' descriptions of steps comprising the process of automated assessment.

Note. Freq. = frequency of category-theme occurrence, which is the number of times each theme was observed in the total sample (frequency of theme occurrence) summed across all themes in a given category (hence called category-theme occurrence). Each theme could contribute to more than one category. Parenthesised percentage values are based on the sum total of category-theme occurrences for the 11 categories.

Discussed below is the central tendency of this schema (i.e., **Table 1**) by examining the four categories that account for the majority of the variability in the data (rows 1 - 4 of **Table 1**): Agent Performing Assessment (21.7%), Medium of Assessment (20.4%), Action Performed (19.6%), and Degree of Automation (16.0%). We focus on these first four categories because together they account for 77.65% of the category-theme occurrences (see **Table 1**). Each of the remaining 7 categories accounted for less than 5.5% of the sum total of category-theme occurrences. Therefore, if we wanted to use the least number of concepts to say the most about this sample's understanding of what ‘automated assessment’ is, we would describe these first four categories, as follows.

The ‘Agent Performing Assessment’ category comprised 24 themes that were grouped in 5 sub-categories; with Computer / Technology accounting for 79.28% of the sum of theme occurrences comprising the category (i.e., 199 of 251 theme occurrences). Therefore the majority of this sample viewed ‘automated assessment’ as something that is performed by computers or technology.

The ‘Medium of Assessment’ category comprised 25 themes that were grouped in 2 sub-categories. One sub-category (Electronic Form) accounted for 97.47% of the sum of theme occurrences comprising this category (i.e., 231 of 237 theme instances). Therefore, the majority of respondents considered automated assessment as an activity that is done via an electronic medium (e.g., online or on computers).

The ‘Action Performed’ category comprised 27 themes in 9 sub-categories. One sub-category (Various Descriptions of ‘Assessing’) accounted for 73.57% of the sum of theme occurrences comprising this category (i.e., 167 of 227 theme instances). This sub-category consisted of 16 themes, amongst which there were 7 references to ‘assessment’ and 9 references to ‘mark’ (of these 9, 3 themes referred to ‘marks’, 2 to ‘marked automatically’, 1 to ‘marked immediately’, 1 to ‘marks reliably/accurately’, 1 to ‘marked/assessed’, and 1 theme referred to ‘marking/grading’). However the names of these 16 themes are themselves aggregated labels to summarise the various expressions used by respondents. In short, the ‘Action Performed’ category predominantly comprises various expressions of assessment that are summarised by two main key words – ‘mark’ and ‘assess’.

The fourth major category is ‘Degree of Automation’, which comprised 18 themes that were grouped in 6 sub-categories. Three sub-categories accounted for 86.47% of the 185 of the theme occurrences comprising this category, these were: ‘Not Specified’ (74; 40%); ‘Automatic’ (56; 30.27%), and; ‘Automation Implied by Negation of Human Involvement’ (30; 16.2%). The thematic content of the latter two sub-categories (i.e., ‘Automatic’ and ‘Automation Implied by Negation of Human Involvement’) is sufficiently well described by their names. However, the first sub-category (‘Not Specified’) deserves some discussion. A Code Memo comments that this sub-category (i.e., ‘Degree of Automation: Not Specified’) “could be used for all other themes that do not specify a degree of automation, but it seems more relevant for a few themes which tacitly imply (but do not explicitly state) automation (e.g., the theme ‘computer program marks assessment’)”. For such themes we can infer that automation is likely but the respondents did not declaratively state the process was automated. Therefore ‘automation’ may or may not feature as part of their understandings of ‘automated assessment’. In short, the most frequent meanings comprising the ‘Degree of Automation’ category, are: various descriptions where automation is likely and/or implied (e.g., those comprising the theme ‘computer program marks assessment’); a set of descriptions that include the adjective ‘automatic’, and; a set of responses that imply automation by negation of human involvement (e.g., the theme ‘human does not mark assessment’).

In summary, to use the least concepts to describe the majority of this sample’s multifaceted understanding of the term ‘automated assessment’, we can describe the content of four main categories: Agent Performing Assessment, Medium of Assessment, Action Performed, and Degree of Automation. The majority of this sample understood ‘automated assessment’ to be an activity that:

1. is performed by computers or technology,
2. is done via an electronic medium (e.g., online or on computers),
3. involves the action of ‘assessment’, communicated through various descriptions that are summarised by two key verbs – ‘mark’ and ‘assess’, and
4. involves a degree of automation, implied or explicitly referenced, whereby automation is: likely and/or implied, but not explicitly stated (e.g., ‘computer program marks assess-

ment’); explicitly referenced using the adjective ‘automatic’ (e.g., ‘computer program automatically marks assessment’), or; implied by negation of human involvement (e.g., ‘human does not mark assessment’).

Considerations regarding interpretation

A number of considerations that are relevant to the interpretation of this analysis are discussed below, namely, generalisation from sample to population, the relative importance/weight of themes and categories, and the epistemological framework (constructivism).

This schema may be unique to the present sample. Its replicability and generalisability would need to be examined using a representative sample of the population of Australian students and staff in tertiary educational organisations. The present sample was not representative of this population; therefore generalisations are not well founded.

When interpreting the relative importance of each category, it is useful to remember that the reason some categories have greater frequency of category-theme occurrence reflects two causes – the respondents’ descriptions, and the researchers’ interpretation/analysis of respondents’ descriptions.

Furthermore, the rank order of themes and categories (based on frequency of theme occurrence and category-theme occurrence) is a valid index of the importance of each element only if we subscribe to a constructivist epistemology that both attributes reality (truth) to the perception of each individual and seeks to interpret this reality (truth) according to the heuristic ‘if more people believe it, then its truth is more relevant’. While the aggregate definition of automated assessment presented above (i.e., that derived from examination of the ‘central tendency’ in the data) tells us what the majority of the sample considered to be the main elements of ‘automated assessment’, it is also important to consider the diversity in views – if Copernicus listened only to what the majority of his contemporaries took to be the case, he would never have questioned the notion that the Earth was the centre of the universe. In particular, the last category, ‘Process Description’, yields cogent insights that can inform experts in designing, implementing, using, and promoting automated assessment systems. These considerations will be the focus of a subsequent paper.

Review, Conclusions, and Future Research

The implications from the survey transfer the focal point of our research agenda to building a profound methodology for automated assessment and, in particular, to delineate a context-specific process of integration for the relevant stakeholders (e.g., students and staff). Taking into account the non-representative nature of this national sample (and also considering the various educational systems in other countries and cultures), some informative inferences can be drawn regarding the requirements and prospects of automated assessment – providing food for thought on educational and business fronts and at executive and administrative levels. Here, we glimpse behind the curtain at future developments and outline the main aspects of marketing, promoting, and refining automated assessment.

Increasing Adoption: Marketing Automated Assessment to Various Stakeholders

The constructivist grounded theory analysis suggests that this sample’s understanding of ‘automated assessment’ is multifaceted, though can be summarised as considering ‘automated assessment’ to be a process that is performed by computers or technology (i.e., as the agent of action), is done online or on computers (i.e., as the medium for action), involves ‘marking’ and ‘assessing’ (i.e., as the action performed), and involves a degree of automation or a lack of human involvement (that was explicitly referenced by some, though most often only implied). From a mac-

rosopic perspective (i.e., considering groups or categories of themes), this understanding/schema seems to be congruent with what the discipline of Information Systems refers to by the term ‘automated assessment’. However, when one examines individual themes and responses, it becomes clear that there is a great degree of diversity between responses and many respondents’ presented descriptions that were very simplistic (describing one or two characteristics such as “multiple-choice tests done online”), rarely explicated processes involved, were more congruent with fixed-response formats (e.g., multiple-choice) and summative assessment, and were less congruent with free-text response formats (e.g., automated essay grading) and formative assessment. Furthermore, while relatively few responses appraised the value of automated assessment in general, negative appraisals were more common than positive ones.

This synopsis is congruent with the quantitative data, where the majority of the sample reported using simple forms of automated assessment (i.e., multiple-choice questions were marked by computers alone for 46.2% of the 171 participants who reported having previously used/experienced multiple-choice questions). In our professional experience as university educators we see that multiple-choice questions are frequently constructed to assess recall (or recognition) of facts (e.g., regarding standardized exams for large cohorts and online materials that accompany textbooks; see also Haladyna, Downing, & Rodriguez, 2002). This is perhaps because it is easier to write multiple-choice questions that assess recall/recognition than those that assess problem solving and analytical reasoning. However, recall/recognition are skills from the lowest level of Blooms’ taxonomy of educational objectives. Indeed, when considering assessment in general (i.e., both human and computer), more people in this sample had used/experienced essays/reports ($n = 219$) and short answer questions ($n = 184$) than multiple-choice questions ($n = 171$). So it may be the case that the academics who chose the formats of these assessments are aware of the limitations of multiple-choice questions (despite their ease of administration and marking). Despite the fact that essays/reports were the most commonly used type of assessment in this sample, 99.5% of the sample said their essays/reports were marked by humans. If this observation generalises beyond this sample (i.e., holds true in replication studies), this presents an unmet opportunity for stakeholders (staff and students) to benefit from automated forms of essay grading. This begs the question of how to increase adoption of emerging assessment technologies.

Based on respondents’ impressions of automated assessment and our professional experience over the years, we outline a set of common needs among stakeholders that we consider to be important in promoting the use of automated assessment for educational benefit. Any marketing of products and methodologies for automated assessment will need to captivate a critical mass of users before the technology can move out of the lab and into common use. Hence, the marketing portfolio calls for the following to be demonstrated across various educational domains: robustness regarding marking accuracy and time (compared to human marking), ease of use for unexperienced users (students as well as teachers), and economic savings. In addition, effectively addressing stakeholders’ concerns involves illustrating a favourable ratio of benefits against potential risks; immediate, accurate, and reliable summative assessment; educationally useful formative assessment and feedback for students; reduced marking load for educators and, therefore, more time for research and individual student mentoring; increased reputation for the university through student and staff satisfaction; and quality control of the assessment process for managerial/administrative staff.

The Ethics and Pragmatics of Promoting Automated Assessment

From an ethical point of view, automated assessment contradicts various views and values of academic culture and implicitly creates an impression that students are stuffed with facts rather than being inspired to analyse, think, and create new knowledge with potential impact on society. Sub-

optimal outcomes like this can occur when automated assessment (or indeed any form of assessment) is not integrated into the educational process using sound pedagogical principals (e.g., when students and assignments are isolated from the teaching process and outcomes).

Pedagogical benefit

Below, we isolate the core educational benefits of automated assessment (as discussed above in relation to marketing considerations) that are available through applications of this technology, which are informed by sound pedagogical principles. Through the use of automated assessment technologies, teachers (and therewith markers, whom we consider here to be equivalent for the sake of simplicity) can be freed from certain tasks (e.g., marking hundreds of assessment items) and consequently have increased time/energy to provide individual guidance to students (e.g., by providing more formative feedback and time for interaction). It follows that teachers can increase the frequency of formative self-assessments with higher-order learning outcomes and with improved (automated) formative feedback to enhance the overall quality and experience of learning. In this way students can gain more freedom to explore their own learning paths along defined milestones and to assess their learning with rewards available for successful performance on self-assessment opportunities. The immediate feedback that is provided on well-designed automated assessments can enhance the learning process and can also activate intrinsic motivation for students within their learning setting. Via automation, assessment outcomes can be returned in *real-time* after a submission deadline (with immediate feedback improving the affect of formative assessment on learning). In high-stakes summative assessment, both automated and human assessment procedures can be implemented as methods of improving accuracy and reliability (i.e., expert human markers can calibrate the accuracy of automated procedures, and, in turn, automated assessment can gauge the continuing reliability of human markers, either across time or between raters).

Commercial benefit

On the whole, reputation is the quintessence of universities' activities. The quality of education is judged in society by successful graduations and post-graduation job performance. High quality education is the harbinger of new student enrolments and the attractor of research and business projects, including endowments and sponsorship, in total increasing the financial scope. Thus, pedagogical benefits of improved assessment methods and outcomes can affect the overall university performance as a business enterprise.

Automated assessment comprises an improvement for the administration and curriculum planning in various ways, including (a) precise calculations of financial costs based on the number students because marking does not depend on the efficiency of the human marker, (b) shorter time spans between exams and results, and (c) well-defined processes can result in synergy effects by, for example, merging decentralized technology in a common (automated) assessment lab in the 'cloud'.

Refining Automated Assessment Technology to Meet the Needs and Address the Concerns of Stakeholders

Automated assessment, beyond objective test scoring, is still the subject of research and is in a nascent stage regarding commercialisation. Short answer questions and essays can be marked by computers with minimal labour required from human markers (Dikli, 2006; Williams & Dreher, 2004). However the status quo still requires experts to accompany setup and evaluation processes. Thus, a useful way forward, informed by the survey data and our professional experience, includes the following steps: (a) constructing a detailed (advanced) process description including scenarios from various educational institutions, (b) defining and implementing the (business)

processes for a centralised automated assessment lab, (c) integration with the universities' systems to access learners' profiles and update their grades, and (d) adding domains by specific extensions like thesauri, terminology databases, and ontologies - tools to automatically 'grasp' further field specific characteristics. This long-term plan needs to be accompanied by particular documentation for different stakeholders in multiple scenarios demonstrating the processes, advantages, and expected outcomes. Fundamental to the envisioned success are elements such as near-intuitive and barrier-free accessibility to complete functionality and comprehensible reports about the reasoning (i.e., the logic and functioning of automated assessment technology), internal statistics, and outcomes of the assessment.

Future Research and Conclusions

Innovations in automated assessment such as automated essay grading are a key to unlocking unused potential for students (e.g., by improving formative feedback), for staff (e.g., by freeing their resources to teach more than to mark), and for universities (e.g., by providing quality assurance). Various automated essay grading technologies exist and can function as accurately as expert human markers; however the 'educational market' is yet to realise the benefits of this.

On one hand, the automated assessment research agenda has identified areas for continued technological innovation, including improvements in robustness, adaptation to new fields, integration into the curriculum and administrative systems, intuitive interfaces, and above all, achieving a critical mass to benefit from the 'pull' that Web 2.0 has demonstrated and continues to demonstrate for us. Web 2.0 and social networks augment the conscious activity of humans in their daily use of web-based applications, especially the contemporary emergence of tools to automate the handling, understanding, interpretation, and answer generation of the content that has been created by society to date. In the same way that search engines and discussion boards apply intelligent algorithms to extract 'right' answers, we need to shift assessment tools to a '2.0' level.

On the other hand, the lack of mainstream use of extant innovations such as automated essay grading can be understood by considering human-related factors such as pedagogical, economic, and socio-technological factors. The data suggest that the majority of educators prefer essays/exams but mark them manually; automated essay grading was used by only a small minority of the sample. In comparison, a large proportion of respondents had used/experienced automated assessment of multiple-choice tests. While essays/reports can assess higher forms of learning (e.g., analysis) their grading is costly in time and effort. Conversely, computer assessed multiple-choice tests are cost efficient but are often used to assess lower-level learning outcomes (e.g., recall/recognition). These results highlight a large gap between current praxis and potential benefits. The qualitative grounded theory analysis suggests that this may be due in part to stakeholders' current understanding of the technology. However, these results are from a non-representative sample, and, while they may indicate important issues or trends that can be investigated in future research, these data do not generalise to the Australian academic population. Future research could seek to replicate or further investigate these findings in representative samples. In this research we have aimed to explore the largely uncharted terrain of automated assessment in Australia and to identify issues that can inform hypothesis testing in subsequent studies.

Proof of concept has illustrated the advantages of automated essay grading, even in comparison with expert human markers, but still there is limited industry application. However all groups of stakeholders realise – motivated by varying arguments – the imperative to adapt and respond to the changing influences of technology and society. Each university, as a multifaceted pedagogical and economic entity, strives to be a leader in research and/or teaching in order to enhance their reputation and attract future students and external funding. Consequently our continuing program of research involves testing these findings against replication samples and via interviews with a range of stakeholder groups. Extant innovations in automated assessment such as essay grading

are strong foundations for us to benefit from. In order to increase adoption of these technologies we need to learn more about stakeholders' concerns, desires, and speak their language in order to have them move-in and feel at home using educational innovations such as automated essay grading.

References

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, 7-74.
- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Book 1: Cognitive domain*. London: Longman.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245-281.
- Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509-535). London: Sage.
- Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5, 1-35.
- Dreher, C., Reiners, T., Dreher, N., & Dreher, H. (2009). Virtual worlds as a context suited for information systems education: Discussion of pedagogical experience and curriculum design with reference to Second Life. *Journal of Information Systems Education (JISE)*, 20, 211-224.
- Dreher, H. (2006). Interactive on-line formative evaluation of student assignments. *Issues in Informing Science and Information Technology*, 3, 189-197.
- Dreher, H., Dreher, N., & Reiners, T. (2008). Design and integration of an Automated Assessment Laboratory: Experiences and guide. *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008* (pp. 2858-2863). Chesapeake, VA: Association for the Advancement of Computing in Education.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15, 309-334.
- Hulsart, R., & McCarthy, V. (2009). Educators' role in promoting academic integrity. *Academy of Educational Leadership Journal*, 13, 49-60.
- Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An overview. *Theory into Practice*, 41, 212-218.
- Livingston, K., & Condie, R. (2006). The impact of an online learning program on teaching and learning strategies. *Theory into Practice*, 45, 150-158.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199-218.
- Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon*, 9, 1-6.
- Ramsden, P., Masters, G., Stephanou, A., Walsh, E., Martin, E., Laurillard, D., et al. (1993). Phenomenographic research and the measurement of understanding: An investigation of students' conceptions of speed, distance, and time. *International Journal of Educational Research*, 19, 301-316.
- Ryan, G. W., & Bernard, H. R. (2000). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 769-802). London: Sage.

Uptake of Automated Assessment Technologies

- Tashakkori, A., & Teddlie, C. (2009). Integrating qualitative and quantitative approaches to research. In L. Bickman & D. L. Rog (Eds.), *The Sage handbook of applied social research methods* (2nd ed., pp. 283-317). Los Angeles: Sage.
- Thompson, D., & McGregor, I. (2009). Online self- and peer assessment for groupwork. *Education & Training, 51*, 434-447.
- Williams, R. (2006). The power of Normalised Word Vectors for automatically grading essays. *The Journal of Issues in Informing Science and Information Technology, 3*, 721-730.
- Williams, R., & Dreher, H. (2004). Automatically grading essays with Markit©. *Journal of Issues in Informing Science and Information Technology, 1*, 693-700.

Biographies



Carl Dreher holds a PhD in Psychology, a Masters in Clinical Psychology, and a Graduate Certificate in Research Commercialisation. Inter alia, Dr Dreher enjoys conducting research regarding mindfulness-based interventions in health care and emerging technologies in Information Systems. He is a “digital native” who finds emerging socio-technological developments to be an excitingly innovative confluence between people and technology. Based on his own empirical observations, he has been known to claim that “smart phones are about just as much fun as you can have while being alone.”



Torsten Reiners is a postdoctoral researcher at the University of Hamburg, Germany, and University Associate with the Curtin University of Technology in Perth, Australia. His research and teaching experiences are in the areas of operations research (meta-heuristics/simulations models for container terminals), fleet logistics, information systems and several topics in eLearning and software development. His PhD-thesis *Simulation and OR with SmartFrame* demonstrated concepts for didactical models. Besides scientific publications, he conducts research in semantic networks to improve cross-border communication, (e)learning and machine translation. Dr Reiner’s interest also includes virtual worlds and their interconnectivity / exchange without barriers. This research includes the development of adaptive systems, automatic processing, analysis, and evaluation of documents, innovative platforms in combination with emerging technologies like mobile devices. Torsten Reiners is co-founder of the Second Life Island University of Hamburg and *Students@work*, an initiative to promote education in Web 3D as well as the value of students' work.



Heinz Dreher is a Professor in Information Systems at the Curtin Business School, Curtin University, Perth, Western Australia. He has published in the educational technology and information systems domain through conferences, journals, invited talks and seminars; is currently the holder of Australian National Competitive Grant funding for a 4-year E-Learning project and a 4-year project on Automated Essay Grading technology development, trial usage and evaluation; has received numerous industry grants for investigating hypertext based systems in training and business scenarios; and is an experienced and accomplished teacher, receiving awards for his work in cross-cultural awareness and course design. In 2004 he was appointed Adjunct Professor for Computer Science at TU Graz, and continues to collaborate in teaching & learning and research projects with European partners. Dr Dreher's research and development programme is now supported by Curtin Business School Area of Research Focus funding – Semantic Analysis and Text Mining for Business and Education (www.eaglesemantics.com) in addition to other competitive funding obtained for individual projects.