



## An evaluation of semi-automated, collaborative marking and feedback systems: Academic staff perspectives

Steven Burrows and Mark Shortis  
RMIT University

Online marking and feedback systems are critical for providing timely and accurate feedback to students and maintaining the integrity of results in large class teaching. Previous investigations have involved much in-house development and more consideration is needed for deploying or customising off the shelf solutions. Furthermore, keeping up to date with the state of the art from both academia and industry is essential. This paper is motivated by a project aiming to identify a marking and feedback system for deployment at the authors' university. A detailed investigation is described which is open minded towards adopting or modifying an existing product, or the implementation of a new solution, with key features and shortcomings described in detail. *Moodle Workshops*, *Turnitin GradeMark*, *Waypoint* and *WebMark* were shortlisted and carried forward for user analysis testing. The outcomes have not only provided key conclusions concerning the suitability of existing solutions, but resulted in a comprehensive collection of functional requirements that leaders of new projects should consider. This paper should be of interest for anyone considering the adoption or upgrade of any marking and feedback system at their home institution.

### 1. Introduction

The provision of feedback to students on assessment tasks is an essential component of their self-monitoring of academic progress. Educational research is consistent on the value of such feedback to students (Higgins et al., 2002), especially for formative assessment, and course evaluation surveys regularly demonstrate that students are dissatisfied with the amount and quality of feedback they receive (Nicol, 2010; Price et al., 2010). Providing timely, personalised and detailed feedback to students, especially in large class teaching (Burrows & D'Souza, 2005), is a significant demand on academics. The approach to marking and assessment is often based on inefficient manual handling and distribution of electronic documents, and assessment for large classes has the additional complication of the coordination of a team of assessors drawn from lecturers, tutors and demonstrators.

Many systems have been developed to support semi-automated and collaborative assessment of student submissions. Semi-automated systems are characterised by the fundamental features of an online submission process by students and distribution of assessed material back to the students in a coordinated fashion. Collaborative systems allow all assessors, including demonstrators, tutors, lecturers and coordinators, to interact with one another on matters concerning assessment without a requirement to use other communication systems such as email. In a fully featured system, markers should be able to assess the submissions online and annotate the submissions with the

system providing, for example, the ability to pre-load customisable marking schemes, extract comments from a database of 'stock phrases' commonly used as feedback, and permit customised feedback on specific issues. The clear intent of semi-automated and collaborative systems is to help provide timely, detailed feedback to individual students, facilitate more efficient and more effective assessment for large classes, and identify areas of academic excellence or weakness.

Based on an identified need at enterprise level, RMIT University internal project grants were awarded in 2007 and 2008 to evaluate semi-automated, collaborative assessment systems. The primary aim of the overall project was to make a recommendation to the University on the optimal system to adopt at enterprise level, and then generalise and share the findings. The project was designed, implemented and managed by the authors, under the guidance of a project steering committee comprising six leading academics with experience in online systems. The investigation was conducted in an open and objective manner, considering all options such as adopting a commercial product or modifying an existing open source solution. Subsidiary aims of the overall project were the identification of the critical functionality of semi-automated, collaborative marking and feedback systems, the weaknesses of current systems, and necessary modifications or customisations that are needed for the implementation of such systems in universities.

During 2007 twenty-nine products were evaluated against a list of six basic criteria for an online marking and feedback system. On the basis of the evaluation, the systems were categorised into a taxonomy that grouped like functionalities. This work resulted in forty-one features and functions of interest being identified, which were used to evaluate four short-listed systems from the initial evaluation in detail: *Moodle Workshops*, *Turnitin GradeMark*, *Waypoint*, and *WebMark*. The four systems represent the full range of possible options as they comprise two commercial products, one open source system and one in-house development. The initial results were inconclusive, so the four systems were then carried forward for a user analysis testing phase in 2008 involving thirteen participants and a series of set tasks for each system, comprising marking guide design, actual marking, and processing results. The evaluation culminated in a focus group meeting to verify the project findings and formulate recommendations for future work. This paper concentrates on the methodology and outcomes of the evaluation process.

The project continued in 2009 with the goal to update the review based on the latest releases and updates to existing systems. Following that, the fourth phase in 2010 involved a further review in order to make a recommendation to RMIT University for an enterprise-wide deployment. The review included recent developments such as *LightWork* (Heinrich, 2009), and enhanced products such as *RemarksXML* (Colbran, 2009), *Blackboard Grade Centre* (Blackboard, 2011) and *Turnitin 2 GradeMark* (iParadigms, 2011). The recommendation of the 2010 review was that RMIT University should retain *Blackboard Grade Centre* as the enterprise system, relying on promised feature enhancements, incrementally implemented by new service packs. Further detail on the 2009 and 2010 phases is beyond the scope of this paper; however, all literature has been updated to represent the "state-of-the-art".

To provide the context and environment of the project, RMIT University is a dual sector institution offering programs ranging from apprenticeship training to PhDs. Programs are offered in the discipline areas of science, engineering, health, architecture

and construction, education and social work, business and economics. In 2011 RMIT has a total student population of 74,000, of which one third are international students, and almost 4,000 staff.

## 2. Previous work

A comprehensive literature review of semi-automated, collaborative assessment and feedback is beyond the scope of this paper, however a review of the development of online assessment and feedback systems is available in previous work by the authors (Shortis & Burrows, 2009), and Hepplestone et al. (2011) provide a ten year catalogue of the literature associated with assessment feedback. Less recent publications provide further background, for example Valenti et al. (2003) provide a review of ten systems and Heinrich et al. (2006) provide a comprehensive review of a wide variety of tools to support essay-type assessment. However, both the technology and the practice continue to advance, and there are many examples of recent thinking around the effectiveness and utility of assessment feedback (Nicol, 2010; Price et al., 2010). On the technology, a consortium of six universities in Australia has recently published a functional requirements specification for the management of online submission and assessment (Electronic Assignment Management Project, 2011).

The initial literature review by the authors focused specifically on operational products that provided online feedback to students, and delivered a list of twenty-nine systems to commence the evaluation. Initial recommendations were also provided by the steering committee. Six criteria were developed and used to scan some essential requirements of the systems, comprising acceptance of electronic submissions, online assessment of submissions, annotation of submissions, collaboration between markers, return of submissions, and return of results (Shortis & Burrows, 2009). The tools were organised into the following taxonomy of seven categories (Shortis & Burrows, 2009) that was developed to accommodate the systems as they were reviewed, and is summarised below. Then descriptions are given of the leading candidates that were carried forward for the user analysis testing.

### 2.1 Feedback-only tools

Feedback-only tools lack convenient functionality to quantify marks. Instead, there is generally more focus on staff-student interaction in the form of comments for incremental drafts, for example. These systems could be used as part of an assessment system if separate feedback proformas are in use, but instructors would generally be better served using a feedback and marking tool described in Section 2.7. *CommentManager* (UTS, 2009), *e-Journal*, and *My Journal* fall into this category. The latter two do not have publicly available documentation. Other systems in this category, released or identified after this study was completed, are *PebblePad* (Willis et al., 2006) and *RemarksPDF/RemarksXML* (Colbran, 2009). Whilst *PebblePad* is primarily an e-portfolio tool, it can be adapted to provide feedback to students. *Remarks* is a purpose-built feedback tool that permits annotation of PDF and other document formats.

### 2.2 Marking-only tools for automated programming assessment

Opposite to feedback-only tools are marking-only tools, which typically focus on automated solutions. Automated tools are popular for programming assessment tasks where a software functional specification can be tested against automated test cases

with agreed inputs and expected outputs. *ASSYST* (Jackson & Usher, 1997), *Auto Test* tool (Radcliffe & Rudolph, 2007), and *BOSS* (Joy & Luck, 1998) are examples in this category.

### 2.3 Self and peer assessment

Self and peer assessment tools allow students to reflect on their own work and that of their peers. They share many of the basic functions of staff-driven assessment tools, but with more emphasis on learning and feedback, and less emphasis on marks. These systems come in many forms such as self or peer assessment by itself, both, or both plus staff assessment functionality. *Calibrated Peer Review* (Chapman & Fiore, 2001), *iPeer* (McAlpine et al., 2006), and *ReView* (Thompson, 2007) fall into this category. Another system in this category, again not considered in the review, is *SPARKPlus* (Willey & Gardner, 2009), a "Self and Peer Assessment Resource Kit".

### 2.4 Test tools

Test tools (also known as quiz tools or survey tools) provide facilities for tabulating answers to a series of questions for summary and assessment. Several question types are typically supported, including select one, select many, and short, free form answers. *WebLearn* (Fernandez, 2001), which provides both tests (for assessment) and quizzes (for practice runs), is an example of this category. Other solutions include *abc.test* and *Questionmark Perception*.

### 2.5 Plagiarism detection tools

Plagiarism detection tools are mostly stand alone, for example see *CopyCatch* (Culwin & Lancaster, 2000) and *EVE* (Bull et al., 2001) for natural language, and *JPlag* (Prechelt et al., 2002) and *MOSS* (Schleimer et al., 2003) for source code. However there are a few tools that also integrate with assessment. For example, the *Turnitin* plagiarism detection module (Savage, 2004) is one part of a larger package that includes *Turnitin GradeMark* (iParadigms, 2009). *SafeAssign* (Blackboard, 2007) is an example of *Blackboard* integration that can complement other learning management system components.

### 2.6 Learning management system components

A learning management system is "a software application that automates the administration, tracking and reporting of training events" (Ellis, 2009). Learning management systems implement many related components to support comprehensive solutions for education and training. Complete learning management systems include grade management, submission or assignment management components, and usually most or all of these. For example, *Blackboard* has *Blackboard Gradebook* for grade management, *Blackboard Digital Drop Box* as a submission tool and *Blackboard Assignments* as an assignment management tool. In version 9.1 of *Blackboard* these functions have been integrated into the new *GradeCentre*. Other learning management systems such as *Moodle* and *Sakai* include similar sets of tools.

### 2.7 Marking and feedback tools

Marking and feedback tools should ideally provide comprehensive features for the whole assessment process, from designing a marking guide, accepting submissions,

conducting marking, to returning feedback and results. Some of these are offline (or desktop) tools, which generally require additional work to synchronise results manually. Some examples are *Assessment@yourfingertips* (Campbell, 2005), *CAFAS* (Computer Aided Feedback and Assessment System) (Freney & Wood, 2005), *Electronic Feedback* (Denton, 2003), *MarkIt* (Dingsdag et al., 2000), *e-TMA* (Electronic Tutor Marked Assignments) (Thomas & Taylor, 2000), *MarkIn* (Creative Technology, 2009), and *Mindtrail* (Stevens & Jamieson, 2002).

The remaining systems reviewed are online, such as *Marking Assistant* (Sondergaard, 2009), *OMAR* (Online Marking and Results) (McKenzie, 2004), *WebCoDe* (Mason et al., 1999), and do not suffer from the synchronisation problem described above. Another system in this category, not considered in the study, is *Assess by Computer* (ABC) marketed by Assessment21 (Sheader et al., 2006), which also provides online functionality for both marking and feedback to students.

Marking and feedback tools is the desired category for this evaluation. Four systems stood out from the others: *Moodle Workshops*, *Turnitin GradeMark*, *Waypoint*, and *WebMark*.

## **2.8 Moodle, Turnitin, Waypoint and WebMark**

*Moodle Workshops* (Cole, 2005) is an assessment module included in the *Moodle* open source course management system. The marking form allows many types of questions from simple Boolean yes/no criteria, upwards to many options with each criterion allowing accompanying free-form comments. The marking interface is highly configurable and allows staff assessments, peer assessments and self assessments all from the one form. Students can submit their work through the module and receive their results once released.

*Turnitin GradeMark* (Henderson, 2008) is an online marking module offered by the *Turnitin* academic integrity vendor. The *GradeMark* package is part of the larger *WriteCycle* suite which also includes grade book, peer review and plagiarism detection modules. *Turnitin GradeMark* can be used as a stand alone module or complementary with any or all of the other modules. The module has an advanced marker interface which allows direct annotation onto the student submissions, irrespective of the file format of the submission. *Turnitin* version 2 has been released since the initial study, but there has been little change in the essential functionality associated with marking and feedback.

*Waypoint* (Skeele et al., 2007) is a commercial *Blackboard* building block or standalone tool for online assignment marking offered by Subjective Metrics. The marking interface supports both select one and select many marking criteria with each contributing a relative weight towards a final score. Results can be "pushed" back into *Blackboard* for viewing by the students.

*WebMark* (Tahaghoghi & Williams, 2004) is a stand alone system developed for use within the School of Computer Science and Information Technology at RMIT University. The marking form uses an XML schema to dictate marking form sections, questions, parts and comments. Other useful features include the ability to manage group assignments and allow different sections of the same assignment to be independently evaluated by different markers.

These four systems were shortlisted for detailed evaluation by the project officer from the results of the initial scan of the systems using the six-point criteria (Shortis & Burrows, 2009) (the data for this component is given in Appendix A). A lengthy list of 41 criteria concerning system-wide features, administrative features, lecturer features, marker features and student features was drawn up in the previous work (Shortis & Burrows, 2009) for this task, but the results for the 17 mandatory features that were used were not conclusive (the results for this component are given in Appendix B with additional detail in Appendices C.1 to C.4). Therefore, it was decided that user analysis testing was needed to advance the project, as described in the next section.

Finally, it is stressed that ongoing work is always required to monitor new work in the currently active systems and the emergence of new projects. In particular, commercial vendors like Blackboard and Turnitin issue regular releases which must be closely monitored.

### **3. User analysis testing methodology**

A full user experiment was designed to have participants to perform mock evaluations of student assignments on *Moodle Workshops*, *Turnitin GradeMark*, *Waypoint* and *WebMark*. Participants were informed initially about the project in a short briefing meeting; participants then completed the mock evaluation tasks at their own pace and recorded their findings in feedback forms; then results were collated for discussion at a concluding focus group meeting. The expected outcome of the evaluation was a recommendation on systems that were feasible for future deployment at RMIT University, or indeed any institution with similar needs.

#### **3.1 Participants**

Thirteen RMIT University staff participated in the user analysis testing. Four of the participants were specifically invited to participate because of their specialist knowledge and experience of educational technology. Care was taken to ensure that these experts represented the three colleges of RMIT University; accordingly, the invited evaluators included two experts from the College of Science, Engineering and Health, one expert from the College of Business, and one expert from the College of Design and Social Context.

The remaining nine participants were volunteers. These participants were recruited through an announcement in the RMIT University staff newsletter and by word of mouth. These evaluators provided a good cross-section of staff experience, technical skills and discipline breadth. The range of disciplines represented was accounting, architecture and design, computer science, educational technology, engineering, the library and medical science, with the majority of the representation being from computer science and engineering. The group had a 60-40 gender balance and a range of age and experience from early career to senior, very experienced academics.

Most of these participants were intended users of the systems, whilst others had other motivations, such as research interests in educational technology and web accessibility. Some of the participants had prior experience with similar marking and feedback systems.

### 3.2 Briefing meeting

The participants were invited to an initial thirty minute briefing. The meeting commenced with a short presentation about the progress of the project to date, project aims, a high level description of the tasks, and a timeline. Participants were asked to complete the tasks during a two week window and spend approximately two hours completing the evaluation tasks for each system. It was then explained that all participants would later be invited to a one hour focus group meeting at a convenient date, once the evaluations had been completed. User documentation and feedback proforma participant materials were distributed at the meeting. Each component of these materials was discussed in turn with the participants.

The briefing concluded with a demonstration of the homepages for the evaluation sandbox websites. The authentication component was explained so that the participants could get started quickly. The authentication screens were not considered part of the evaluation.

### 3.3 User documentation content

Two key documents were provided to each participant - the "user documentation" and the "feedback proforma". The user documentation contained an introduction, task descriptions, participant information, authentication details, system documentation, and tips for getting started, as outlined below.

- The introduction provided a summary of the project background, aims, systems under evaluation and the purposes of the user documentation and feedback proforma documents.
- The three task descriptions for each of the four systems (twelve tasks in total) were described in full to the participants plus an optional fourth task for each system (see Appendix D). The tasks were not designed to exhaustively test all system functionality. Instead, the focus of the evaluation was on the most commonly used components. For each system, participants were first asked to create a marking guide for a fictitious assignment (see Appendices E.1 to E.4, for example), mark a fictitious submission, view the results, and then optionally explore other aspects of the system in time remaining.
- The participant information described an individualised pathway for the participants to evaluate the systems to mitigate bias. That is, the participants were divided into four equal-sized groups and each group started evaluating a different system first to stratify the evaluation.
- The authentication details comprised a fictitious staff account and a fictitious student account for use with each of the four systems. Usernames, passwords and web addresses for each system were documented.
- The system documentation provided links to the system user manuals. Additional documentation was provided to avoid common pitfalls. For example, it was important that users did not fail the tasks simply because they were using the wrong web browser or web browser version. Furthermore, some of the systems allow for multiple types of assignments, so participants were directed to select the assignment type most appropriate to a "marking and feedback tool" as per the taxonomy in Section 2.

- Tips for getting started described how to quickly get started considering that the evaluations were expected to be time consuming. Lastly, contact information for technical and other support was included.

### 3.4 Feedback proforma content

The feedback proforma was provided for recording all aspects of the evaluation. A Microsoft *Word* form was used with fixed spaces to record feedback in the form of checkbox answers and free-form comments. The feedback proforma contained an introduction, and spaces for recording individual system feedback, other aspects, comparison of the systems evaluated, ranking of features and functions, and general comments as described below:

- The introduction provided a welcome message, and a reminder that the form is not necessarily to be completed in the order presented, but that participants needed to complete the system evaluations in the individualised orders described in the user documentation.
- The individual system feedback section of the evaluation required participants to log positive and negative remarks about each task completed for all four systems and record whether the task was completed, aborted due to problems or not attempted due to time constraints (for the volunteer participants).
- The optional section for recording other aspects was provided for recording other feedback for functionality outside of the primary tasks where participants could afford extra time to explore these areas.
- After completing the tasks, participants were then asked to compare the systems evaluated by providing a ranking of the functionality / features and usability / navigation on a five-point Likert scale, summarise the best and worst features of each system in point form, and finally provide an overall view and ranking of the four systems. The overall ranking was to be determined for all systems relative to one another and include reasons for the recommendations. Participants made these judgements based upon their overall impression of the systems.
- Participants were next given the opportunity to rank the full forty-one features and functions mentioned in Section 2 to validate the rankings given by the project steering committee. A five-point Likert scale was used with scores from one (very unimportant) to five (very important).
- The proforma included space for participant general comments that did not fit in other parts of the form.

### 3.5 Focus group meeting agenda

The purpose of the focus group was to validate the user analysis testing findings as compiled from the participant feedback. All aspects of the evaluation were on the agenda including features/functions, task completion rates, and qualitative/quantitative feedback. For the features/functions, it was planned to compile the participant feedback prior to the meeting side by side with the steering committee feedback and flag anomalies for discussion. The goals were to cover general and



system specific themes, aggregated quantitative evaluation responses and draft recommendations for consideration by the focus group participants.

#### **4. User analysis testing results**

Unfortunately, the overall length of the evaluation tasks for the four systems meant that some participants were not able to complete the evaluation due to time constraints, which resulted in some incomplete feedback forms from the volunteer participants. In addition, one volunteer was not able to provide any feedback. This limited the usefulness of the quantitative findings of the project, and in some cases participants were subsequently asked to provide some additional feedback in cases where they were already close to being finished, and may have missed a question. The marking guide creation task was the most onerous, and it is recommended that time commitments of this type of task should be closely scrutinised in marking and feedback system evaluations. When performing a similar experiment in the future, it is recommended that participants are also given a prompt to record the actual amount of time spent on each system. This data would provide another measure of the ease of use.

##### **4.1 Individual system feedback**

The task completion data showed that the completion rates degraded as the participants worked through consecutive tasks. This occurred because of the sequential nature of the tasks, for example participants could not mark an assignment without successfully implementing a marking guide. Task completion data is not reported here, as the participant's view of a 'task completion' was somewhat subjective. For example, there was no discrimination on whether participants marked nearly completed tasks as 'completed' or 'aborted'.

The participant comments provided the most comprehensive feedback on the evaluation. This feedback yielded a substantive catalogue of positive remarks, negative remarks and problems encountered for each system. This feedback has been given its own coverage in Section 5.

##### **4.2 Comparison of systems evaluated**

The best and worst feature summaries were short and concise, and in some cases similar responses from participants could be aggregated as given in Table 1. It would be expected that more responses could be aggregated if the project involved more participants or had a list of predefined comments been provided. The participants were asked to spend eight hours on the evaluation, so there was some frustration generated by misunderstanding of work flows and the impact of known problems with some of the systems, leading to the critical nature of some comments. Again, the discussion of the overall qualitative findings has been reserved for Section 5.

Concerning the quantitative questions, eight responses were received for the functionality/ features rating, and nine responses were received for the usability/ navigation and overall ratings. The remaining participants did not complete enough of the evaluation to gauge these ratings. Results are presented in Tables 2, 3 and 4 respectively.

Table 1: A summary of the best and worst features in the four short-listed systems as given by the participants. Repeated comments are aggregated as indicated.

System	Best features	Worst features
<i>Moodle</i>	<ul style="list-style-type: none"> <li>• Screen layout (x3).</li> <li>• Comprehensive.</li> <li>• Uncluttered.</li> <li>• Easy to understand.</li> <li>• Assignment management page.</li> <li>• Workshop creation process.</li> <li>• Clear screens.</li> <li>• Navigability.</li> </ul>	<ul style="list-style-type: none"> <li>• Lacks flexibility to be able to mix different types in the same assignment.</li> <li>• Only accepts integer marks.</li> <li>• Setting up and editing new workshops.</li> <li>• Unable to go back and modify the assessment part.</li> <li>• Not clear on emailing all students or individuals.</li> </ul>
<i>Turnitin</i>	<ul style="list-style-type: none"> <li>• Annotation/comment/markup feature (x4).</li> <li>• Plagiarism detection (x2).</li> <li>• Documentation (x2).</li> <li>• Seems to be very comprehensive according to the documentation.</li> <li>• Grading rubric.</li> <li>• Statistical data.</li> <li>• Menu system.</li> <li>• Colours.</li> <li>• Clarity.</li> <li>• Familiarity with other components.</li> </ul>	<ul style="list-style-type: none"> <li>• Inflexibility of grading design (x2).</li> <li>• Inflexibility.</li> <li>• Cumbersome.</li> <li>• Navigation.</li> <li>• Getting stuck in loops where its not obvious how to get out and back to the main menu, not sure if information will be saved or lost if one does get 'out'.</li> <li>• Slow interface.</li> <li>• Not user friendly (e.g. the rubric scoreboard).</li> <li>• Stuck with things you may not need or want for this assignment.</li> <li>• Instructions.</li> </ul>
<i>Waypoint</i>	<ul style="list-style-type: none"> <li>• Familiarity with <i>Blackboard</i>.</li> <li>• User friendly.</li> <li>• Good basic functionality.</li> <li>• Ease of use and flexibility of marking schemes.</li> <li>• Documentation.</li> <li>• Assignment creation process.</li> </ul>	<ul style="list-style-type: none"> <li>• Switching between <i>Blackboard</i> and <i>Waypoint</i> was confusing.</li> <li>• Very poor usability.</li> <li>• Cryptic.</li> <li>• Hard to use.</li> <li>• Lacked advanced features.</li> <li>• Slow moving around screens.</li> <li>• Menus are not intuitive.</li> </ul>
<i>WebMark</i>	<ul style="list-style-type: none"> <li>• Flexibility of marking guide design (x2).</li> <li>• Flexibility.</li> <li>• Simple.</li> <li>• Easy to use.</li> <li>• Reasonable organisation.</li> <li>• Does its job well.</li> <li>• Ability to have assignments marked by multiple assessors (on a per question basis for example).</li> <li>• Powerful.</li> <li>• Result report.</li> <li>• Ability to email results to students either individually or as a whole class.</li> <li>• Easy to view results.</li> </ul>	<ul style="list-style-type: none"> <li>• XML requirement (x4).</li> <li>• Marking guide filename convention.</li> <li>• Dreadful looks.</li> <li>• Requires training.</li> <li>• Cryptic.</li> <li>• User interface needs improvement.</li> <li>• Interface.</li> <li>• Menu and navigation.</li> <li>• Documentation.</li> </ul>

The results show preference towards *Moodle Workshops*, *Turnitin GradeMark*, *Waypoint* and *WebMark* in that order for all three questions, and possible reasons are discussed in Section 5. It is noted that the alphabetic ordering of these system names is coincidental as participants were given clear instructions to evaluate the systems in customised sequences.

Table 2: Results of the participant features/ functionality ratings of the four short-listed systems using a five-point Likert scale. A higher score indicates higher preference.

System	Terrible 1	Poor 2	Average 3	Good 4	Excellent 5	Count	Average	Std dev
<i>Moodle</i>	0	1	1	3	3	8	4.0	1.1
<i>Turnitin</i>	0	0	3	4	1	8	3.8	0.7
<i>Waypoint</i>	1	1	3	3	0	8	3.0	1.1
<i>WebMark</i>	3	1	1	3	0	8	2.5	1.4

Table 3: Results of the participant usability/ navigation ratings of the four short-listed systems using a five-point Likert scale. A higher score indicates higher preference.

System	Terrible 1	Poor 2	Average 3	Good 4	Excellent 5	Count	Average	Std dev
<i>Moodle</i>	0	1	3	3	2	9	3.7	1.0
<i>Turnitin</i>	0	3	4	1	1	9	3.0	1.0
<i>Waypoint</i>	2	1	4	1	1	9	2.8	1.3
<i>WebMark</i>	4	1	2	2	0	9	2.2	1.3

Table 4: Results of the participant relative rankings of the four short-listed systems. A lower rank indicates higher preference, unlike Tables 2 and 3.

System	Rank 1	Rank 2	Rank 3	Rank 4	Count	Average	Std dev
<i>Moodle</i>	6	2	1	0	9	1.4	0.7
<i>Turnitin</i>	1	4	3	1	9	2.4	0.9
<i>Waypoint</i>	2	0	5	2	9	2.8	1.1
<i>WebMark</i>	0	3	0	6	9	3.3	1.0

Statistical significance cannot be obtained from these small result sets, so no unequivocal conclusions are offered suggesting that one system is strictly 'superior' to another. Given that there were not enough participants willing and able to commit sufficient time to complete the written feedback, validation of the findings in the focus group meeting was essential (see discussion in Section 6). Correlation of written feedback and focus group findings provided a higher level of confidence in the conclusions than could be drawn from the evaluation alone.

### 4.3 Ranking of features and functions

All features except for two were ranked with a score higher than the middle score on the five-point Likert scale, indicating preference for a highly functional solution that 'does everything'. The evaluation suggests agreement with this need as most systems in the high level system evaluation performed well only on a subset of the required tasks (such as comment management, or peer review tools), leaving the need for a comprehensive solution undiminished. However, it is not valid to overstress this point with just eleven participants having completed this part of the evaluation.

## 5. User analysis testing qualitative findings summary

This section summarises the detailed qualitative feedback provided by the participants. Common themes that were identified for multiple systems are discussed first followed by system-specific themes for *Moodle Workshops*, *Turnitin GradeMark*, *Waypoint* and *WebMark* respectively.

## 5.1 Common themes

Prior experience with online marking and feedback systems was low for many participants and some participants didn't understand the fundamental workflow that such systems provide. For example, the first step was to 'create an assignment module', and in some cases participants tried to submit an assignment first which is not possible in some systems. Workflow wizards would be helpful here in guiding inexperienced users beginning with an empty and default system account.

The design of the commercial systems (*Turnitin GradeMark* and *Waypoint*) could be more tightly integrated in some areas. For example, the main entry points to the *Turnitin GradeMark* component are the small 'assess me' icons next to each assignment submission and it would be helpful to have additional entry points to highlight the availability of this module. With *Waypoint*, the software is completely separate from *Blackboard*, and the system model requires some data to be manually pushed by the click of a button between systems at the discretion of the user. It can be easy to forget to push pending results from *Waypoint* to the *Blackboard GradeBook* module in this manner.

The ability to switch between user roles can speed up the workflow considerably. *Moodle* and *Turnitin* have functionality for this without requiring additional accounts or user login, which means that re-authentication is not necessary and student views of the system can be previewed quickly when logged in as a member of staff.

The split-screen features in *Moodle Workshops* and *Turnitin GradeMark* were particularly favoured by the evaluators. This functionality allows the user to view the submission in the top portion of the screen and manage the marks and feedback in the bottom portion. Viewing both components at once eliminates the need to toggle between multiple windows.

Familiarity with existing systems or supporting tools was appreciated with *Turnitin* and *Waypoint*. Having an institution-wide licence for the *Turnitin* plagiarism module and *Blackboard* (with which *Waypoint* is integrated) provided familiar starting points.

Negative feedback concerning update rates and responsiveness were reported for both *Turnitin GradeMark* and *Waypoint*. *Turnitin GradeMark* appears to have the most bandwidth-expensive interface with its advanced features that facilitate assignment annotations. Also, responsiveness of *Waypoint* was noted as slow from the experience of the evaluators, which may be attributed to the geographic location of the *Waypoint* server administered by Subjective Metrics.

Participants strongly disliked working with multiple software packages for marking online assessment, as much toggling between the packages was required for regular activities. *Waypoint* was problematic here as *Waypoint* and *Blackboard* were used together in the evaluation. For example, *Waypoint* users need to manually push assignment results to the *Blackboard Gradebook* module, as noted above. This concern is not applicable, however, when using *Waypoint* as a stand alone tool, which was not tested in the evaluation. *WebMark* was also problematic as the software only manages marking and a second tool was required to accept the submissions.

## 5.2 Moodle Workshops

*Moodle Workshops* was most applauded for non-functional requirements such as usability, layout, navigation and accessibility. However, documentation was found to be incomplete as there are user-contributed placeholders that need completion. Also, a *Moodle Workshops* marking guide is relatively complex to set up, as options are included that can facilitate self-assessment and peer-assessment, as well as staff-assessment, in the one interface.

## 5.3 Turnitin GradeMark

Participants appreciated the comprehensive assignment annotation functionality which was not present in other marking and feedback tools considered in this evaluation. Also, the unrelated plagiarism detection and peer review packages that are available with an extended licence were considered desirable. Moreover, *Turnitin GradeMark* was the only solution that allows annotation of submissions by means of its split-screen interface containing an annotatable copy of the submission in the top region, and the marking form at the bottom. Other positive feedback identified the detailed documentation and statistical analyses of class performance.

Remarks about 'ease of use' were provided about marking guide design as there are indeed good cues for key steps. However, *Turnitin GradeMark* is the least flexible in terms of marking guide design because the marking guide criteria can only be organised into a single grid, which means that each item must have an identical number of pre-defined feedback options. Additionally, the relative weightings of each criterion must be identical. Marking guides that require a "select many" design are not currently supported, therefore only "select one" designs are currently supported.

Evaluation of *Turnitin-Blackboard* integration features remains for future work. The evaluation used a stand alone *Turnitin* set-up due to resource constraints.

## 5.4 Waypoint

*Waypoint* was commended for the presentation of its reports in particular. Also, *Waypoint* allowed for more flexible marking guide design than *Turnitin GradeMark* but was not as flexible as *Moodle Workshops* and *WebMark*.

One participant remarked on the unfamiliar American terminology used, such as essay elements and superior grades, and therefore some internationalisation considerations may be helpful for other regions. Also, *Waypoint* uses pop-up windows extensively and some participants recorded difficulty in managing the web browser pop-up blocker, which sometimes interfered with access to the system.

## 5.5 WebMark

The strongest features of *WebMark* are the flexibility in marking guide design and the per-question marking feature whereby submissions can be easily shared between assessors on both a per-question and per-student basis. This feature is not available in other marking and feedback system considered in this evaluation.

*WebMark* does not include a submission upload screen, but this allowed a simple implementation of group-based assignment marking. The interface allowed multiple

student usernames to be entered for a received assessment, unlike the other systems that require a one to one relationship between the submission and a student. Also, *WebMark* was the only system of the four reviewed that manages hurdle requirements as there are options to group assessments together when preparing final results.

The most difficult aspect of *WebMark* is the requirement to use marking guides based on XML code. This requirement was problematic for many participants who did not have sufficient XML knowledge. If an XML format is retained in future iterations of this software, an XML editor would be essential as suggested by one participant. Otherwise a complete redesign of the marking guide creation process may be necessary for users without the necessary technical background.

Some participants attempted to create the XML marking guide, but couldn't operate the upload interface which required a specific file naming convention including the subject identifier, offering information, assignment number and case-sensitive input. Additionally, the entry point to the marking guide upload interface was independent of the list of subjects and the filename information would not be required if the upload feature was better integrated.

Usability and look and feel were generally criticised as non-intuitive and clumsy. In particular, there are awkward radio buttons on the main menu which require selection before clicking the 'next' button. A more effective solution would allow the menu options to be selected directly. *WebMark* is an in-house project designed for and used exclusively by the School of Computer Science and Information Technology at RMIT University. Substantial changes would be needed before use by other schools and universities would be feasible.

Finally, *WebMark* was the only solution where unmarked submissions cannot be easily identified, as *WebMark* does not accept submissions directly from students.

## 6. User analysis testing focus group meeting outcomes

Nine participants were present for the focus group meeting and discussions were led by the authors. These participants were all original evaluators and the remaining four were unable to attend.

The advantages and disadvantages of commercial, open source and in-house solutions were discussed in the context of options for future deployment. Concerns were raised with the commercial systems because the host institution would have least creative control. However, any implementation effort using an open source or in-house approach could be potentially superseded by a commercial system at any time.

Not surprisingly, a commercial solution such as *Turnitin* was well liked because of the familiarity with the system and the ease with which this option could be made available within any institution with licences for other modules, such as the plagiarism detection module. There was some support for a wider trial of *Turnitin GradeMark* to follow this project, whilst still recognising the limitations of the current marking guide structure. Similarly, *Waypoint* is attractive given its ability to operate as a *Blackboard* building block or an independent tool.

An open source solution such as *Moodle* may not be difficult to establish within a university, but ongoing support would be required. The formation of an 'expert user

group' would be valuable in answering academic and operational questions; however the provision of ongoing technical support for development and maintenance would have to be established. Concerns were expressed regarding information security issues as well as development of suitable data feeds and connections with enterprise systems. There is also the question of economics and how the expenditure on licensing fees of a commercial product compares to the additional internal support required for an open source product.

Whilst an in-house development would enable the most creative control over system features, doubts were raised concerning the ability to support the development of a new system. In the case of RMIT University, the Information Technology Services group sets a priority for available resources in supporting commercial systems and integration of enterprise level systems, rather than the development of new, technically demanding applications. Therefore the most favoured option for an in-house development was the use of a contractor or a student project to develop an initial prototype that could be subsequently adopted and integrated later. Irrespective of the source of the prototype, there would have to be significant support the university in terms of re-engineering, integration and ongoing maintenance. However the clear advantage of this pathway is that the priority of desirable features identified would be respected.

## 7. Systems summary

Each of the four systems considered have strengths and weaknesses. In the case of RMIT University, for example, none of the four systems would be directly compatible with enterprise systems and will require modification, customisation or the provision of data feeds. Beyond this requirement, each system has individual issues to be considered:

*WebMark* was least favoured by the evaluators, largely because of the poor interface design and the requirement to understand XML to be able to create and modify the marking guide. The system is designed around the premise that it will be used by staff in the School of Computer Science and Information Technology at RMIT University and will deal only with students enrolled in courses from this school. Clearly *WebMark* would require significant modification to be viable more widely.

*Waypoint* has the advantages of flexible marking, good reports and a link to *Blackboard*, but the evaluators found the interface and terminology confusing. Evaluators also felt that the link to *Blackboard* could be more tightly integrated. Furthermore, knowledge of web-browser pop-up blocker use is needed.

In the case of RMIT University, *Turnitin* is a known system with current deployment of the plagiarism detection module and has the advantage of familiarity. In general, like *Waypoint*, there are *Blackboard* integration options, and the *Blackboard* integration manual (iParadigms, 2009) suggests a more seamless integration than the stand alone *Waypoint* interface. One of the strongest features of *Turnitin GradeMark* is the ability to directly annotate student submissions, a feature not available in any of the other systems considered in this project. The substantive drawback of this system is the restrictive grid-like structure of the marking guide creation interface, which may inhibit the widespread adoption of the assessment function.

*Moodle Workshops* was the system most favoured by the evaluators, notwithstanding strong reservations on the support needed for open source systems and the need for integration with enterprise systems. *Moodle Workshops* has the most intuitive interface and a good level of flexibility in marking guide creation. The open source nature of the system allows for a high level of customisation and potential to be integrated with enterprise systems, but would require a sustained investment of resources.

## 8. Recommendations for future work

Of the commercial systems, *Turnitin GradeMark* was identified as a strong candidate. In the case of RMIT University, facilitating a fixed term trial would be straightforward to set up, given the current licensing arrangement for the plagiarism module. The new module would only need to be 'turned on' by *Turnitin* account administrators after purchase of a suitable licence, if this option is pursued. However, an additional investigation is needed to evaluate the *Blackboard* integration options.

*Moodle Workshops* was ranked highest in all quantitative questions. In the case of RMIT University, this system has least potential for enterprise system integration given the investment with *Blackboard*. Customisation and support possibilities need to be investigated. Institution user guides would address the incomplete documentation problem.

A final possibility is the commencement of an in-house project incorporating an agreed subset of the best features reviewed in this project plus integration. Care would need to be taken to avoid the problems observed in the *WebMark* in-house project, and results of this project should not be extrapolated to suggest that all in-house projects will have similar problems. An in-house project could be implemented by a contractor or as a student group project with future enterprise system integration with a central information technology services group. This was the most favoured option as indicated by the focus group participants and steering committee.

Having made the above recommendations, ongoing work is always required to monitor new and updated marking and feedback systems. The rapid development of new systems and the enhancement of existing systems with new features is such that several that were not considered in the project have since come to the attention of the authors. Development of commercial and open source systems will continue and it is therefore imperative that any new project remains flexible to accommodate the latest innovations.

## 9. Conclusions

This paper has reported on a project investigating options for a semi-automated, collaborative marking and feedback system for deployment or creation at RMIT University, with the aim to generalise and share all recommendations for the use of other institutions. The initial work involved a breadth-first evaluation of twenty-nine systems using six basic criteria. This phase resulted in forty-one features and functions of interest, and the features identified as mandatory were used to evaluate the four short-listed systems: *Moodle Workshops*, *Turnitin GradeMark*, *Waypoint* and *WebMark*. These systems were then carried forward for a user analysis testing phase involving thirteen participants, which culminated in a focus group meeting to verify the project findings and formulate recommendations for future work.



Surprisingly, the four most promising systems have limited functionality in the desired criteria of a semi-automated, collaborative assessment system. Whilst all four systems will automatically distribute assessment results to students, only *GradeMark* returns feedback and only *WebMark* supports collaborative assessment. In fact, across all of the twenty-nine systems surveyed, very few have these desired features. Innovation, both educational and technological, is not an inhibitor, yet the most frequently observed automatic function is the simple distribution of results. Whilst systems will continue to be developed, it must be concluded that fully featured systems that address the majority of the 'must have' functionality of semi-automated, collaborative assessment are not yet widely available.

This paper has presented the detailed background work needed for the adoption or development of a marking and feedback system for any institution. The paper has identified the essential features required for such systems to be adopted by any institution, and has also identified the clear weaknesses of available systems. There is an undiminished need for semi-automated, collaborative assessment systems to respond to the need for timely, personalised, detailed feedback to students, as well as to provide an efficient and effective system for academics to manage assessment. Continuing research is required to track the development of new and updated systems in order to monitor progress toward the optimum system that has all of the required functionality.

## Acknowledgments

The authors acknowledge the following people and organisations for their involvement and support with this project:

- RMIT University for supporting this project with the 2007 and 2008 Learning and Teaching Investment Fund (LTIF) grants.
- Project steering committee members Jim McGovern, John Milton, Garry Allan and Seyed Tahaghoghi.
- Robyne Lovelock and iParadigms for their support with the *Turnitin* evaluation.
- Andrew McCann, Christopher Heisen, Peter Heisen and Subjective Metrics for their support with the *Waypoint* evaluation.
- Nigel Williams, Jim McGovern and the RMIT University Information Technology Services group for their support with the *Moodle* and *WebMark* evaluations.
- User analysis testing participants John Benwell, Daryll Cahill, George Fernandez, Heiko Rudolph, Iza Bartosiewicz, Meaghan Botterill, Andrea Chester, Ferry Dharmawan, Margaret Hamilton, Margaret Jollands, Pauline McCabe, Tom Molyneaux and Stephen Penman.
- Paper reviewers John Benwell, George Fernandez and Heiko Rudolph from RMIT University, Andrew McCann from Subjective Metrics, and the anonymous reviewers.

## Appendices

Appendices A, B, C.1, C.2, C.3, C.4, D, E.1, E.2, E.3 and E.4 are contained in the accompanying file 'burrows-appendix.pdf'. URL <http://www.ascilite.org.au/ajet/ajet27/burrows-appendix.pdf>

## References

Blackboard (2007). *Blackboard SafeAssign Building Blocks Instructor Manual*. [viewed 17 Apr 2009 at [http://www.cmu.edu/blackboard/how-to/gradcenter/Blackboard\\_SafeAssign\\_Building\\_Blocks\\_Instructor\\_Manual.pdf](http://www.cmu.edu/blackboard/how-to/gradcenter/Blackboard_SafeAssign_Building_Blocks_Instructor_Manual.pdf), verified 6 Nov 2011 at <https://behind.blackboard.com/s/faculty/refcenter/docs/details.Bb?DocumentID=3911&pid=100&rid=-1&dt=>]

- Blackboard (2011). About the Grade Center. [viewed 26 Oct 2011] [http://help.blackboard.com/instructor/Content/\\_instructor\\_grade\\_center/grade\\_center\\_about.htm](http://help.blackboard.com/instructor/Content/_instructor_grade_center/grade_center_about.htm)
- Bull, J., Collins, C., Coughlin, E. & Sharp, D. (2001). *Technical review of plagiarism detection software report*. Luton, England: University of Luton Computer Assisted Assessment Centre. [http://www.jisc.ac.uk/uploaded\\_documents/luton.pdf](http://www.jisc.ac.uk/uploaded_documents/luton.pdf)
- Burrows, S. & D'Souza, D. (2005). Management of teaching in a complex setting. In *Proceedings of the Second Melbourne Computing Education Conventicle*. Melbourne, Australia. [verified 6 Nov 2011] <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.71.6199&rep=rep1&type=pdf>
- Campbell, A. (2005). Application of ICT and rubrics to the assessment process where professional judgement is involved: The features of an e-marking tool. *Assessment and Evaluation in Higher Education*, 30(5), 529-537. <http://dx.doi.org/10.1080/02602930500187055>
- Chapman, O. L. & Fiore, M. A. (2001). *Calibrated peer review - a writing and critical thinking instructional tool - the white paper: A description of CPR*. Los Angeles: University of California.
- Colbran, S. (2009). *The ReMarks PDF Markup Editor*. Canberra, Australia: Australian Learning and Teaching Council.
- Cole, J. (2005). Workshops. In *Using Moodle: Teaching with the popular open source course management system*, 95-114. Sebastopol, California: O'Reilly Media.
- Creative Technology (2009). MarkIn homepage. <http://www.cict.co.uk/software/markin>
- Culwin, F. & Lancaster, T. (2000). A review of electronic services for plagiarism detection in student submissions. In *Proceedings of the Eighth Annual Conference on the Teaching of Computing*. Edinburgh, Scotland: LTSN Centre for Information and Computer Science. [http://www.ics.heacademy.ac.uk/events/presentations/317\\_Culwin.pdf](http://www.ics.heacademy.ac.uk/events/presentations/317_Culwin.pdf)
- Denton, P. (2003). Returning feedback to students via email using Electronic Feedback 9. *Learning and Teaching in Action*, 2(1), 1-9. <http://www.celt.mmu.ac.uk/ltia/issue4/denton.shtml>
- Dingsdag, D., Armstrong, B. & Neil, D. (2000). Electronic assessment software for distance education students. In *Learning to Choose - Choosing to Learn. Proceedings ASCILITE Coffs Harbour 2000*. [http://www.ascilite.org.au/conferences/coffs00/papers/don\\_dingsdag.pdf](http://www.ascilite.org.au/conferences/coffs00/papers/don_dingsdag.pdf)
- Electronic Assignment Management Project (2011). Functional Requirements: Electronic Assignment Management Project: La Trobe University, Flinders University, Massey University, University of Canberra and University of Southern Queensland. Personal communication, 22 September 2011.
- Ellis, R. K. (2009). A field guide to learning management systems. In *Learning Circuits*. Alexandria VA, USA: American Society for Training and Development. [http://www.astd.org/NR/rdonlyres/12ECDB99-3B91-403E-9B15-7E597444645D/23395/LMS\\_fieldguide\\_20091.pdf](http://www.astd.org/NR/rdonlyres/12ECDB99-3B91-403E-9B15-7E597444645D/23395/LMS_fieldguide_20091.pdf)
- Fernandez, G. (2001). WebLearn: A common gateway interface (CGI)-based environment for interactive learning. *Journal of Interactive Learning Research*, 12(2), 261-276. <http://www.editlib.org/p/8422>
- Freney, M. & Wood, D. (2005). CAFAS: An automated computer aided assessment tool for providing student feedback and managing assessment information. In *Proceedings of the Fifth Evaluations and Assessment Conference*. Curtin University of Technology. <http://otl.curtin.edu.au/eac2006/papers/freneywood.pdf>
- Heinrich, E. (2009). Lightwork: Managing marking effectively. In *Same places, different spaces. Proceedings ascilite Auckland 2009*. <http://www.ascilite.org.au/conferences/auckland09/procs/heinrich-poster.pdf>
- Heinrich, E., Milne, J., Crooks, T., Granshaw, B. & Moore, M. (2006). Literature review on the use of learning tools for formative essay-type assessment. In *e-Learning Collaborative Development*. Wellington, New Zealand: Tertiary Education Commission. <http://etools.massey.ac.nz/documents/LitReview101006.pdf>

- Henderson, P. (2008). Electronic grading and marking: A note on Turnitin's GradeMark function. *History Australia*, 5(1), 11.1-11.2.  
<http://journals.publishing.monash.edu/ojs/index.php/ha/article/download/275/288>
- Hepplestone, S., Holden, G., Irwin, B., Parkin, H. J. & Thorpe, L. (2011). Using technology to encourage student engagement with feedback: A literature review. *Research in Learning Technology*, 19(2), 117-127. <http://dx.doi.org/10.1080/21567069.2011.586677>
- Higgins, R., Hartley, P. & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, 27(1), 53-64.  
<http://dx.doi.org/10.1080/03075070120099368>
- iParadigms (2009). Turnitin GradeMark. [viewed 6 May 2009; not found 6 Nov 2011].  
<http://turnitin.com/static/grademark.html>
- iParadigms (2011). GradeMark. [viewed 26 Oct 2011].  
<https://turnitin.com/static/products/grademark.php>
- Jackson, D. & Usher, M. (1997). Grading student programs using ASSYST. In *Proceedings of the Twenty-Eighth SIGCSE Technical Symposium on Computer Science Education*. San Jose, California: ACM Press. <http://dx.doi.org/10.1145/268084.268210>
- Joy, M. & Luck, M. (1998). Effective electronic marking for on-line assessment. In *Proceedings of the Sixth Annual Conference on the Teaching of Computing and the Third Annual Conference on Integrating Technology into Computer Science Education: Changing the Delivery of Computer Science Education*. Dublin, Ireland: ACM Press. <http://dx.doi.org/10.1145/282991.283096>
- Mason, D., Woit, D., Abdulla, A., Barakat, H., D'Souza, M. & Pires, C. (1999). Web-based evaluation for the convenience of students, markers, and faculty. In *Proceedings of the Fifth International Conference on Web-Based Learning*. Fredericton, Canada: WWW Courseware Developers. <http://naweb.unb.ca/99/proceedings/mason/>
- McAlpine, I., Reidsema, C. & Allen, B. (2006). Educational design and online support for an innovative project-based course in engineering design. In *Who's Learning? Whose Technology? Proceedings ASCILITE Sydney 2006*.  
[http://www.ascilite.org.au/conferences/sydney06/proceeding/pdf\\_papers/p168.pdf](http://www.ascilite.org.au/conferences/sydney06/proceeding/pdf_papers/p168.pdf)
- McKenzie, S. (2004). Assessing quality of feedback in online marking databases: An opportunity for academic professional development or just Big Brother? In *Beyond the comfort zone. Proceedings ASCILITE Perth 2004*.  
<http://www.ascilite.org.au/conferences/perth04/procs/mckenzie.html>
- Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment and Evaluation in Higher Education*, 35(5), 501-517.  
<http://dx.doi.org/10.1080/02602931003786559>
- Prechelt, L., Malpohl, G. & Philippsen, M. (2002). Finding plagiarisms among a set of programs with JPlag. *Journal of Universal Computer Science*, 8(11), 1016-1038.  
<http://dx.doi.org/10.3217/jucs-008-11-1016>
- Price, M., Handley, K., Millar, J. & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment and Evaluation in Higher Education*, 35(3), 277-289.  
<http://dx.doi.org/10.1080/02602930903541007>
- Radcliffe, P. J. & Rudolph, H. (2007). Plagiarism: A survival strategy or recipe for disaster? In *Proceedings of the Eighteenth Annual Conference of the Australasian Association for Engineering Education*. Melbourne, Australia.  
[http://www.aae.com.au/conferences/papers/2007/paper\\_51.pdf](http://www.aae.com.au/conferences/papers/2007/paper_51.pdf)
- Savage, S. (2004). Staff and student responses to a trial of Turnitin plagiarism detection software. In *Proceedings of the Third Australian Universities Quality Forum*. Adelaide, Australia: Australian Universities Quality Agency.  
[http://www.auqa.edu.au/files/publications/auqf2004\\_proceedings.pdf](http://www.auqa.edu.au/files/publications/auqf2004_proceedings.pdf)

- Schleimer, S., Wilkerson, D. & Aiken, A. (2003). Winnowing: Local algorithms for document fingerprinting. In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*. San Diego CA: ACM Press. <http://dx.doi.org/10.1145/872757.872770>
- Shedder, E., Gouldsbrough, I. & Grady, R. (2006). Staff and student perceptions of computer-assisted assessment for physiology practical classes. *Advances in Physiology Education*, 30, 174-180. <http://advan.physiology.org/content/30/4/174.full>
- Shortis, M. R. & Burrows, S. (2009). A review of the status of online, semi-automated marking and feedback systems. In *ATN Assessment Conference 2009*. Melbourne, Australia: RMIT University. <http://emedia.rmit.edu.au/conferences/index.php/ATNAC/ATNAC09/paper/viewFile/205/46>
- Skeele, R. W., Carr, V. B., Martinelli, J. & Sardone, N. B. (2007). Innovation in e-assessment: Exploring a multidimensional tool. In *Proceedings of the 2007 World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. Quebec City, Canada. <http://editlib.org/p/26368>
- Sondergaard, H. (2009). Personal communication.
- Stevens, K. & Jamieson, R. (2002). The introduction and assessment of three teaching tools (WebCT, Mindtrail, EVE) into a post graduate course. *Journal of Information Technology Education*, 1(4), 233-252. <http://jite.org/documents/Vol1/v1n4p233-252.pdf>
- Tahaghoghi, S. M. M. & Williams, H. E. (2004). The WebMark course management system. In *International Conference on Computing in Education (ICCE)*. Melbourne, Australia.
- Thomas, P. & Taylor, D. (2000). Reducing the distance in distance education. In *Proceedings of the Fourth Computer-Assisted Assessment Conference*. Loughborough, England: Loughborough University. <http://hdl.handle.net/2134/1786>
- Thompson, D. (2007). Integrating graduate attributes with student self-assessment. In *Proceedings of the Connected 2007 International Conference on Design Education*. Sydney, Australia. <http://hdl.handle.net/10453/3008>
- UTS (2009). CommentManager: An online system for commenting on students' work. [viewed 6 May 2009; not found 6 Nov 2011]. <http://www.iml.uts.edu.au/assessment/feedback/commentmanager.html>
- Valenti, S., Neri, F. & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education*, 2, 319-330. <http://jite.org/documents/Vol2/v2p319-330-30.pdf>
- Wiley, K. & Gardner, A. (2009). Assessment for learning: Using minor assessment to promote major learning. In *ATN Assessment Conference 2009*, RMIT University. <http://epress.lib.uts.edu.au/research/handle/10453/11389>
- Willis, H., Gravestock, P. & Jenkins, M. (2006). Throwing a pebble into the pond: E-portfolios and student engagement. In *Who's Learning? Whose Technology? Proceedings ASCILITE Sydney 2006*. [http://www.ascilite.org.au/conferences/sydney06/proceeding/pdf\\_papers/p68.pdf](http://www.ascilite.org.au/conferences/sydney06/proceeding/pdf_papers/p68.pdf)

**Authors:** Steven Burrows, School of Computer Science and Information Technology  
RMIT University, GPO Box 2476, Melbourne 3001, Australia.  
Email: [steven.burrows@rmit.edu.au](mailto:steven.burrows@rmit.edu.au)

Mark Shortis, School of Mathematical and Geospatial Sciences, RMIT University,  
GPO Box 2476, Melbourne 3001, Australia. Email: [mark.shortis@rmit.edu.au](mailto:mark.shortis@rmit.edu.au)

**Please cite as:** Burrows, S. & Shortis, M. (2011). An evaluation of semi-automated, collaborative marking and feedback systems: Academic staff perspectives. *Australasian Journal of Educational Technology*, 27(7), 1135-1154. <http://www.ascilite.org.au/ajet/ajet27/burrows.html>