The Implementation of an Automated Assessment Feedback and Quality Assurance System for ICT Courses

J. Debuse
M. Lawley
R. Shibl
Faculty of Business
University of the Sunshine Coast
Maroochydore DC, Queensland 4558, Australia
jdebuse@usc.edu.au mlawley1@usc.edu.au rshibl@usc.edu.au

ABSTRACT

Providing detailed, constructive and helpful feedback is an important contribution to effective student learning. Quality assurance is also required to ensure consistency across all students and reduce error rates. However, with increasing workloads and student numbers these goals are becoming more difficult to achieve. An automated feedback system, referred to as the Automated Feedback Generator (AFG), has therefore been designed and developed with the aim of providing superior quality assurance and efficiency in both assessing student assignments and providing feedback. Unlike existing automated marking and feedback software, AFG aims to allow educators to perform the entire process of student feedback generation for any assessment type. The AFG system is investigated across two introductory ICT courses: general ICT and programming. The aim is to demonstrate that AFG provides a more effective means for providing student feedback than alternative manual and automated approaches. This is achieved by comparing AFG with these alternatives and demonstrating that it offers quality control, efficiency and effectiveness benefits whilst generating consistent feedback from a student perspective. An empirical approach is employed using attitudinal data. T tests are used to test hypotheses comparing three feedback generation approaches: AFG, manual and a more complex automated approach. The results show that feedback from AFG was perceived to be constructive, helpful and with error levels less than or equal to those for other course feedback approaches; students also found feedback to be consistent with that produced by the more complex alternatives.

Keywords: Quality assurance, Automated assessment, ICT education

1. INTRODUCTION

Numerous studies have found that feedback on assessment is important. Feedback has been consistently found to be influential in student achievement (Black and William, 1998; Hattie, 1987; Higgins, Hartley and Skelton, 2002). If students are to engage in a subject and identify areas of strengths and weaknesses (Hyland, 2000) they need feedback on their progress and performance (Higgins, Hartley and Skelton, 2002; Thurmond et al., 2002). Their motivation and self-efficacy can also be increased by providing personalized feedback on assessment rather than generic comments (Allen et al., 2003; Hyland, 2000). As a result, feedback is viewed as being very important by students of all ages (Felix, 2001) and across a range of fields (Lyall and McNamara, 2000; Sims, 2000). Student failure has even been related to an absence of feedback (Entwistle et al., 1989). However, just providing feedback is not sufficient; its quality must be high enough to be useful. This can be achieved using quality assurance, defined as "systematic management and assessment procedures...to ensure achievement of quality outputs or improved quality" (Harman and Meek, 2000, p.

vi). In order to achieve quality assurance and as a result be effective, feedback should be timely (James, McInnis and Devlin, 2002; Wiggins, 1997), informative (James, McInnis and Devlin, 2002) and detailed (Wiggins, 1997).

While the benefits of feedback are known, students and teachers often differ in their perception of these (Maclellen, 2001). A serious issue with feedback is that too little of it is received from instructors (Holmes and Smith, 2003). A further concern is feedback consistency (Holmes and Smith, 2003). Feedback has commonly been found to be vague and non-specific (Higgins, Hartley and Skelton, 2002), leading to negative perceptions by students. In addition, handwriting legibility has been identified as a common problem (Higgins, Hartley and Skelton, 2002).

Feedback must therefore possess a range of qualities to connect with students and hence impact on their achievement (Higgins, Hartley and Skelton, 2002). However, with growing workloads and student numbers it is increasingly difficult to provide such quality, personalized feedback (Higgins, Hartley and Skelton, 2002).

This study investigates the suitability from a student perspective of an educational system designed to improve

efficiency and quality assurance of feedback on assessments. The software produces this feedback in the most common form within its application context of higher education, namely written comments (Higgins, Hartley and Skelton, 2002). The aim is to demonstrate that AFG provides a more effective means for providing student feedback than alternative manual and automated approaches. This is achieved by comparing AFG with these alternatives and demonstrating that it offers quality control, efficiency and effectiveness benefits whilst generating consistent feedback from a student perspective. An empirical approach is employed using attitudinal data. T tests are used to test hypotheses comparing three feedback generation approaches: AFG, manual and a more complex automated approach.

2. BACKGROUND

The issues regarding quality can be ameliorated with the use of technology to assist instructors in providing feedback. Table 1 summarizes the key features of a number of example packages.

System	Marking	Feedback	Assessment types
Blayney and Freeman	Automated	Automated	Spreadsheet
TRAKLA2	Automated	Automated	Algorithm design
CourseMaster	Automated	Automated	Program code, circuit & OO diagrams
VIOPE	Automated	Automated	Program code
SchemeRobo	Automated	Automated	Program code
MarkIT	Automated	None	Essay with model answer
MindTrail	Manual	Automated	Any
Re:Mark	Manual	Automated	Essay
EFS	Manual	Automated	Any
Word Autotext	Manual	Semi-automated	Any
Teachers Report Assistant	Manual	Semi-automated	Any
ReportWriter	Manual	Semi-automated	Any
TutorBoard	Manual	Semi-automated	Text-based

Table 1: Key features of marking and feedback support systems

If assessment (both production of marks and feedback) is fully automated then the potential quality assurance improvements are dramatic, but the complexity of this process demands considerable development effort that limits its application to simple, narrowly defined types such as spreadsheet questions (Blayney and Freeman, 2004). Blayney and Freeman's (2004) system, developed at the University of Sydney, is restricted by both its spreadsheet question format and the need for programming expertise in developing new questions, although it provides feedback containing marks and comments automatically.

Fully automated assessment is also supported by the prototype MarkIT system, which grades essays by comparison with a model answer. The system is limited to a single assessment type, and gives overall marks only without detailed feedback (Williams and Dreher, 2004). Moreover, large differences between marks produced manually and those assigned by MarkIT can occur on individual essays, although the difference between the average marks given by both approaches is small.

The application of full assessment automation to more complex tasks has the drawback of restricting its application even further to a single topic. For example, there are a variety of packages to automatically assess programming assignments and generate feedback (Korhonen et al., 2003). These include the free TRAKLA2 (Korhonen et al., 2003) and SchemeRobo (Saikkonen, Malmi and Korhonen, 2001), together with commercial systems such as the Ceilidh successor CourseMaster (Foxley et al., 2001; Higgins, Symeonidis and Tsintsifas, 2002) and VIOPE (Carver and Henderson, 2006; Vihtonen and Ageenko, 2002).

Limiting the automation of the assessment process to feedback production places the additional burden of grading upon educators; however, flexibility is dramatically improved, allowing application to the enormous range of assessments that are beyond marking automation. The commercial MindTrail package achieves such feedback automation by allowing educators to record marking information within knowledge trees, from which standardised feedback may be generated (Blayney and Freeman, 2004). A key drawback of MindTrail is that its efficiency gains are mitigated by preparation requirements that can exceed four hours (Stevens and Jamieson, 2002). The feedback produced by MindTrail has also not been found superior to that produced manually, although it appears to improve consistency across markers (Cargill, 2001). MindTrail have now ceased trading (Cargill, 2001).

More limited feedback automation is supported by the commercial Re:Mark package ("Re:Mark Online Grading and Markup Solution for Blackboard", 2006). The system is restricted to adding in-line comments to essay-type assessments and therefore unsuitable for many ICT assignments such as databases or programming. Re:Mark allows educators to pre-define comments and organize these into folders, with associated marks that may be customized. The package integrates with the Blackboard course management system, and can assign grades based on comments directly into its student results database.

The automation of feedback in the form of marks and comments is performed by the freely available Electronic Feedback system (EFS), based on Word and Excel VBA (Denton, 2001). The system is very flexible and powerful, but its principal drawback is a steep learning curve; the array of system components is likely to take a significant time for users to learn (Figure 1), and the complexity of entering marks and comments for each student is significant (Figure 2). EFS allows standard comments to be defined for each marking criterion, each with an associated score, and can use the combination of comments to produce the marks and feedback. EFS stores the results in a spreadsheet, and offers

492

options such as emailing results to students, checking for plagiarism through patterns in results and validation to ensure marks are within preset limits.

Increasing the requirement for manual intervention in feedback production reduces the scope for efficiency improvements. In addition, quality assurance mechanisms such as mark validation and plagiarism detection cannot be included. However, such approaches require minimal software or technical knowledge; for example, feedback

generation can be semi-automated by storing and retrieving frequently used comments using the Autotext feature of Microsoft Word (Krucli, 2004). The free Teachers Report Assistant (Denton, 2001) uses a similar approach, in which educators load banks of pre-written comments, edit them and select combinations that are customised with a student's name and gender. These are then copied to the clipboard for insertion in report documents. Similarly, the Apple Macintosh ReportWriter system allows educators to select
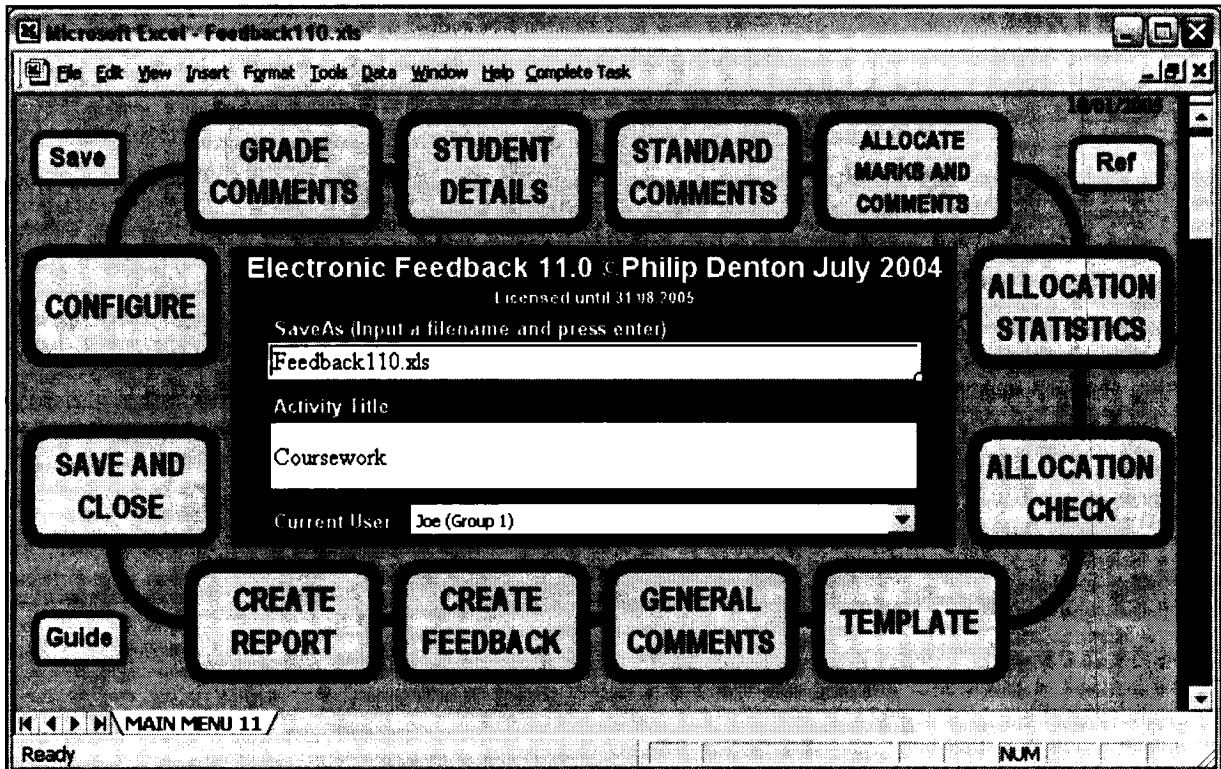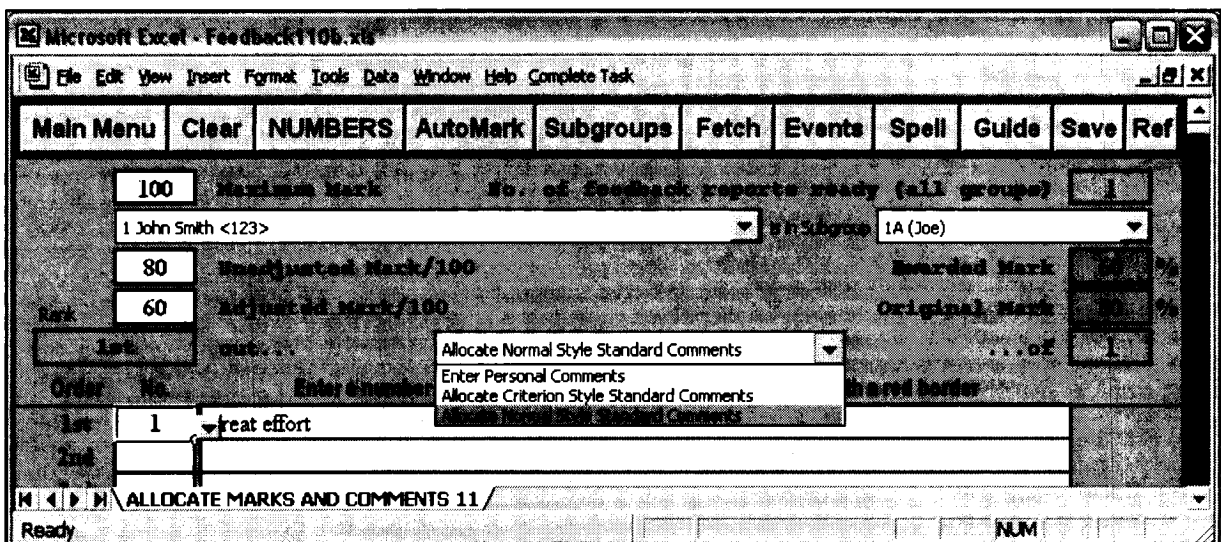

Figure 1: The EFS main interface

Figure 2: The EFS mark and comment allocation interface

493

pre-defined comments prior to editing to produce end of year reports (Robinson, 2002). The free TutorBoard package extends this theme, allowing pre-defined comments and free-hand drawing to be added to students' work (Heaney, 2005; Heaney and Daly, 2003).

This study presents a new fully automated feedback generation system (AFG), which has been designed to overcome many of the limitations identified in existing semi-automated and manual approaches. AFG can be applied to a broader range of topics and assessment types than systems that automate grading. Furthermore, its operational simplicity and efficiency, feedback detail levels and assessment breadth cannot be matched by existing fully automated feedback systems. AFG is evaluated in this study by measuring the quality of its feedback from a student perspective and benchmarking this against an existing marking system and manual approaches.

A description of AFG follows, before the research method and results; these are then discussed and conclusions presented.

3. THE AFG SYSTEM

AFG was conceived and developed by the lead author using Microsoft Word VBA; its operation is illustrated in a simplified process diagram within Figure 3, and Figure 4 shows its main interface. The system was developed for academic use, and there are currently no plans to commercialise the technology. The software can be obtained by contacting the authors. Contact details are provided at the start of the paper. AFG contains an embedded spreadsheet (Figure 5), where users enter marking criteria details in the columns followed by student marks and comments in the rows. The feedback, which is similar to that generated by Mindtrail (Cargill, 2001), is then automatically produced at the click of a button. AFG informs the user and refuses to produce feedback if any of the following quality assurance checks fail: every mark awarded must be numeric, non-empty and below the maximum specified by the user; maximum mark values and student IDs must be non-negative numbers; all criteria must have an associated title and maximum possible marks; the total marks for each student must be calculated correctly; at least one criteria group must exist; and student IDs, names and surnames must not be missing. Users are also forced to save the program before feedback generation to guard against loss of data through system failure.

Comments can be added to the feedback given on each criterion where marks are deducted, using the standard Excel spreadsheet comment facility. Alternatively, double clicking a cell opens the comment entry interface (Figure 6), which extracts the comments for the criterion from all other students and presents them in descending order of prevalence. Any comments that already exist for the cell are selected, and the user can switch each available comment on or off by clicking on it. New comments can also be typed in and added to the list; when the interface is closed, all selected comments are added to the cell. When feedback is generated, the user can choose to either ignore the comments or add them to the feedback documents produced.

AFG is highly scalable, since increases in the number of students do not affect the set-up time and only impact marginally on feedback generation. Indeed, the time taken to mark each student is likely to decrease with increasing student numbers, since the likelihood of having to write a new comment from scratch will diminish as the total number of available comments increases. Educators can also write more detailed and lengthy comments if they are for reuse, reducing the common problems of feedback insufficiency (Holmes and Smith, 2003) and vagueness (Higgins, Hartley and Skelton, 2002).

Entering results directly into a spreadsheet rather than writing on the assignments submitted by students reduces the problem of feedback consistency (Holmes and Smith, 2003) in a number of ways: there are reduced ambiguities as a result of detailed instructions for each criterion; individual marks can have justifying comments attached to them for consistency checking; and if multiple markers are used then uniformity across them can be improved, since each may see at a glance all marks awarded for every criterion. The common problem of handwriting legibility (Higgins, Hartley and Skelton, 2002) is also avoided.
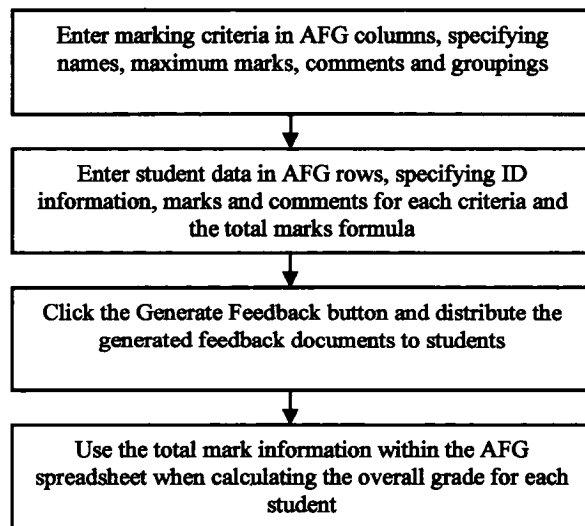
```
┌─────────────────────────────────────────────┐
│ Enter marking criteria in AFG columns,        │
│ specifying names, maximum marks, comments     │
│ and groupings                                 │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│ Enter student data in AFG rows, specifying ID │
│ information, marks and comments for each       │
│ criteria and the total marks formula           │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│ Click the Generate Feedback button and         │
│ distribute the generated feedback documents    │
│ to students                                    │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│ Use the total mark information within the AFG  │
│ spreadsheet when calculating the overall grade │
│ for each student                               │
└─────────────────────────────────────────────┘
```

Figure 3: A simplified process diagram illustrating the operation of AFG

In summary, AFG has been designed to provide superior quality assurance and efficiency when assessing student assignments and providing feedback. The aim of the remainder of this study is to determine whether students perceive AFG to deliver this superiority.

4. METHOD

A survey of students' attitudes to the feedback produced was conducted using an instrument adapted from the University's student evaluation of teaching questionnaire, along with additional sources (Blayney and Freeman, 2004; Hede, 2005). The questions were designed to determine how useful students found the feedback from AFG, and how this compared to manual feedback from other courses. Standard software evaluation instruments could not be used within this study, since the respondents had no direct usage of the soft-

494

Automated Student Feedback Generator

Instructions

The spreadsheet referred to in these instructions can be found on the following page of this document; it contains a number of rows containing marking criteria information, followed by a row for each student. Columns exist for student ID, surname and first name, followed by a column for each marking criteria. The definitions that apply to this spreadsheet may be found in the Appendix.

To use the automated student feedback generator, the following steps must be followed; before doing so, you should create a copy of the original version of this document, so that it is available to be customised for future assessments. Note also that to edit the spreadsheet it must be double clicked.

1. Go to Tools...Macro...Security - confirm that security level is medium. Open Excel and do the same. If you change either level close and reopen this document.
2. For each marking criterion that exists for the assessment, perform the following steps. Note that the first marking criteria column in the spreadsheet should be used for the first criterion, the second for the second and so on.
    a. Place the criterion description in the criteria description row of the spreadsheet.
    b. Place the maximum marks that may be awarded for the criterion in the criteria marks row of the spreadsheet.
    c. Place any comments for the criteria, such as exactly how the criterion is to be applied, in the criteria comments row of the spreadsheet.
3. In the column following the last marking criterion (total marks column), place the text "Total marks" and "Total" in the criteria description and criteria group title rows of the spreadsheet respectively; the total marks available for the assessment should also be placed in the criteria marks row of the spreadsheet in this column.
4. For each group of marking criteria that exists, place the group title in the criteria group title row.
5. For each student that is to be assessed, place the following information in each of the student rows (each row corresponds to a single student)
    a. Place the ID, surname and first name in the ID, surname and first name columns.
    b. Place a formula in the column that computes the sum of all the marking criteria columns in the total marks column.
    c. Enter the marks awarded for each marking criterion in the marking criteria columns. Errors in mark entry such as values below zero or above the maximum marks available for the criterion are highlighted by the cell changing colour; such colouring applies to up to 100 marking criteria and 2000 students.
    d. Enter any comments regarding each mark awarded by double clicking on its cell and adding and selecting the comments within the form that is shown. If the comments are to be included within the feedback documents then click the following check box so that a tick appears within it:
    e. ☑ Include Comments in Feedback
6. Save this document; the contents of the spreadsheet can be copied and used in Excel at a later date if required.
7. Enter the full path (eg C:\feedback\) that the feedback documents are to be saved in within the following textbox:

8. Enter the title of the assessment (eg Individual Paper 1) within the following textbox:

9. Click the following button to begin feedback generation: **Generate Feedback**
10. When generation is complete, a feedback document will be generated for each student in the specified path, with name set to surname followed by first name and student ID.

Figure 4: The AFG Main Interface

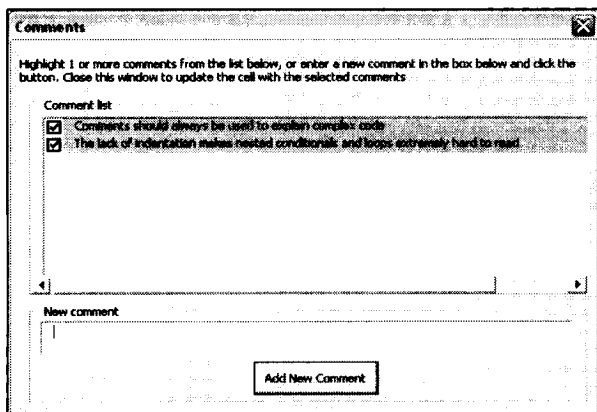**Figure 5: The AFG Spreadsheet Interface**

Figure 6: The AFG comment entry interface

ware; however, the suitability and accuracy components of the functional quality measure defined by Ortega, Perez and Rojas (2003) were used

Five point Likert scales were used to measure agreement with the following statements; in addition to a response of 'not applicable', the values ranged from one ('strongly disagree') through three ('neutral') to five ('strongly agree'):

S1. The feedback in the printed document containing individualized comments and marks for my "BUS108 individual research paper 1" was constructive & helpful.

S2. The feedback in the printed document containing individualized comments and marks for my "BUS108 individual research paper 1" was more constructive & helpful than written feedback received for assessment items within other USC courses this semester.

S3. The feedback in the printed document containing individualized comments and marks for my "BUS108 individual research paper 1" contained more errors than written feedback received for assessment items within other USC courses this semester.

The feedback for the BUS108 course referred to in the statements was produced by AFG. BUS108, referred to as Course 1 within this study, was an introductory ICT course, and the research paper was an essay on the use of ICTs within an organisation. The aim of statement S1 was to assess student attitudes to the utility of automated feedback. Statements S2 and S3 were designed to assess how students viewed the utility and reliability respectively of the automated feedback compared to feedback provided within other courses at the University.

The results for these statements were used to test hypotheses, using a similar approach to that presented by Stevens and Jamieson (2002). The first pair of hypotheses tested whether course participants agreed or disagreed with statement S1:

H_{1a}: Course participants perceive that the feedback produced by AFG is constructive and helpful. The response to statement S1 would need to be greater than three for this to hold.

H_{1b}: Course participants perceive that the feedback produced by AFG is not constructive and helpful. The response to statement S1 would need to be less than or equal to three for this to hold.

Hypotheses to test agreement or disagreement with statements S2 and S3 were formed in the same way.

A similar approach was used to test the overall attitude of each student towards the importance of feedback, so that this could be taken into account when analyzing the results. Students were asked 'How important to you is receiving written feedback on assessment items?' with the available responses ranging from 0 ('not at all important') through 2 ('moderately important') to 4 ('extremely important').

The survey also contained free response items; students were asked to identify the best aspects of the feedback produced by AFG, together with the areas that were most in need of improvement. The instrument also contained questions to identify respondent information such as degree type, number of degree courses taken and first language. Course 1 had many students spread across multiple educators; a question was therefore added to confirm that the correct tutor had marked their assessment and therefore feedback had indeed been produced by AFG.

A control group was created within the study, comprising students within Course 1 whose research paper feedback had been manually produced. Two separate educators produced the feedback for the control group students.

A version of the survey was also applied to students that were taught by the author within an introductory programming course, referred to within this study as Course 2. The assignment was a Java programming exercise, and the feedback was produced using AFG.

The EFS system described in Section 1 was included within the study to allow the performance of AFG to be benchmarked. Version 12 of the package was used by the author, and the survey was conducted a year after all other surveys. The first research paper assessment for Course 1 was again used, with the topic changed slightly from the previous year to examine the benefits of business information systems to organisations.

The courses examined within the study have considerable homogeneity, as illustrated in Figure 7. All Course 1 groups were identical across key areas such as modules, delivery methods and learning outcomes; they also shared the same coordinator, who has overall responsibility for designing and running the course. The tutors in Course 1 are only responsible for delivering tutorials and marking assignments; the coordinator designs the content of both of these. The only source of variation across Course 1 groups other than their feedback production approach is thus: their delivery of the tutorials prepared for them by the coordinator; their assignment marking; and, in the EFS case, the year in which the course was held, which resulted in a minor topic change within the assignment. A mark standardization meeting held by the course coordinator reduced the likelihood of variation in assignment marking. The coordinator also performed further standardization before the marks and feedback were returned to the students, by requesting a sample from each educator and suggesting modifications if they fell outside standard boundaries. Further, the EFS and AFG Course 1 groups had the same educator. This educator was also responsible for the AFG group within Course 2, thus further reducing the scope for variation. Course 2 had the same educator throughout all lectures, tutorials and marking. Thus,

496

#	Feedback production	Mean	95% Confidence Interval for Mean		Std. Dev.	N
			Lower Bound	Upper Bound		
S1	Manual	3.74	3.35	4.12	1.109	34
	AFG	3.87	3.53	4.20	.900	30
	EFS	4.19	3.85	4.53	.75	21
S2	Manual	3.22	2.92	3.52	.832	32
	AFG	3.21	2.74	3.67	1.103	24
	EFS	3.29	2.76	3.81	1.146	21
S3	Manual	2.52	2.19	2.84	.890	31
	AFG	2.13	1.59	2.67	1.254	23
	EFS	2.20	1.75	2.65	.951	20

Table 2: Responses to Statements

the delivery method, modules, learning styles and outcomes were all standardized within Course 2.

The lead author conducted all surveys within tutorials; this allowed far greater response rates to be achieved than would be possible through alternative mechanisms such as mail, web or email.

Missing values were treated on a pair-wise basis within analyses, so that responses with missing data were only excluded from tests for which the data was required. For all independent samples T tests that were performed, significance was tested at the 0.05 level and all variables were tested for normality and homogeneity of variances prior to statistical analysis; where assumptions could not be met even after transformations, an independent samples T test assuming inequality of variances was undertaken.

5. RESULTS

A total of 87 responses to the survey were received; the response rate was 95%. Four of the Course 1 responses did not identify the required marker and were therefore rejected. A respondent also gave answers to statements S2 and S3, comparing the feedback to other courses, yet responded that they were only taking a single course; their responses to S2 and S3 were therefore removed.

The most popular degree program was the Bachelor of Business (Accounting), which was taken by 25% of respondents. The students appeared to have limited time available for study; the most frequent number of hours of paid work per week during the semester was greater than ten. Moreover, 88% of students were full time, and the most frequent number of degree level courses being taken was four. Language issues seem unlikely for the students, with 87% identifying English as their first language.

Table 2 contains the mean, standard deviation, 95% confidence intervals and N values for responses to statements S1, S2 and S3 presented in Section 3. The 95% confidence intervals for statements S1, S2 and S3 are shown in Figures 8, 9 and 10 respectively.

The responses appear to be very similar regardless of whether the feedback production is produced automatically by AFG or manually. Indeed, an independent samples T test shows no significant difference between the mean value for AFG automated and non-automated feedback for each statement (S1: $t_{0.05, 62}$ = -5.16, P = 0.608; S2: $t_{0.05, 54}$ = 0.040, P = 0.968; S3: $t_{0.05, 52}$ = 1.323, P = 0.192). The two automated approaches also yielded similar responses; an independent samples T test shows no significant difference between the mean value for EFS automated and AFG automated feedback for each statement (S1: $t_{0.05, 49}$ = -1.352, P = 0.182; S2: $t_{0.05, 43}$ = -0.231, P = 0.819; S3: $t_{0.05, 41}$ = -0.202, P = 0.841).

For all three approaches, hypothesis H_{1a} holds, since the lower bound of the mean in each case exceeds three. Hypothesis H_{1b} holds for none of the approaches, since the upper bound of the mean in each case is not three or less. Hypotheses H_{2a}, H_{2b} and H_{3a} do not hold for any approach, but H_{3b} holds for all. Respondents therefore appeared to find AFG automated, EFS automated and manual feedback
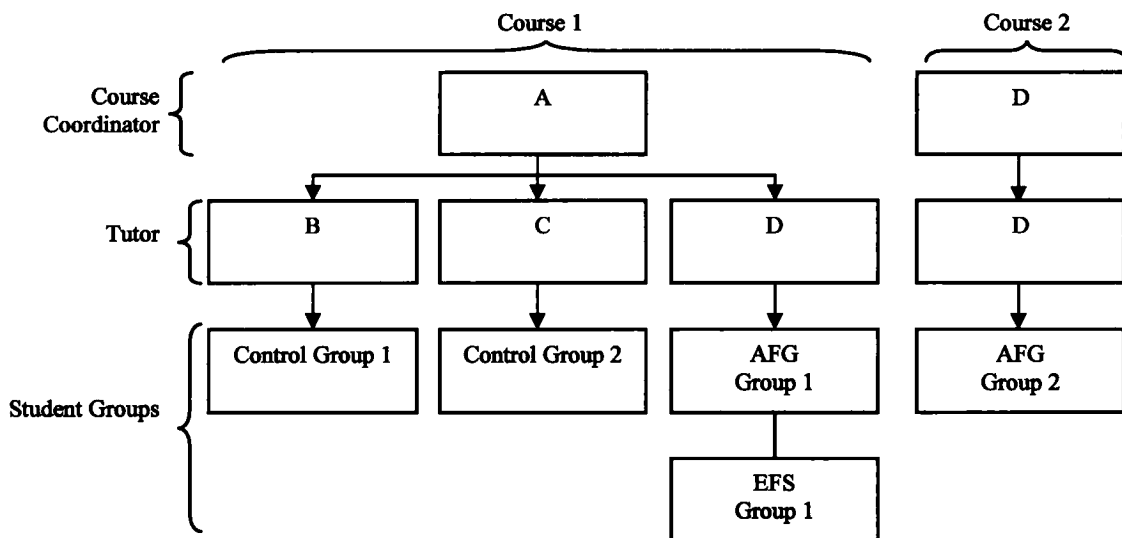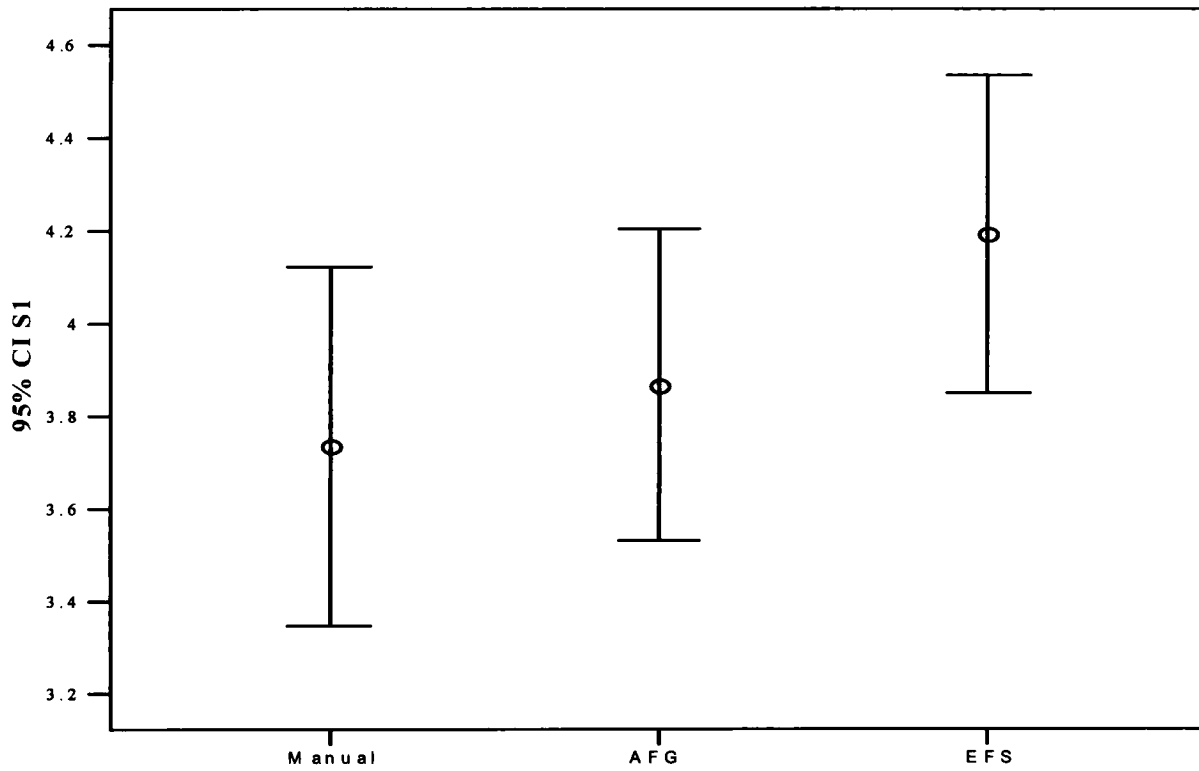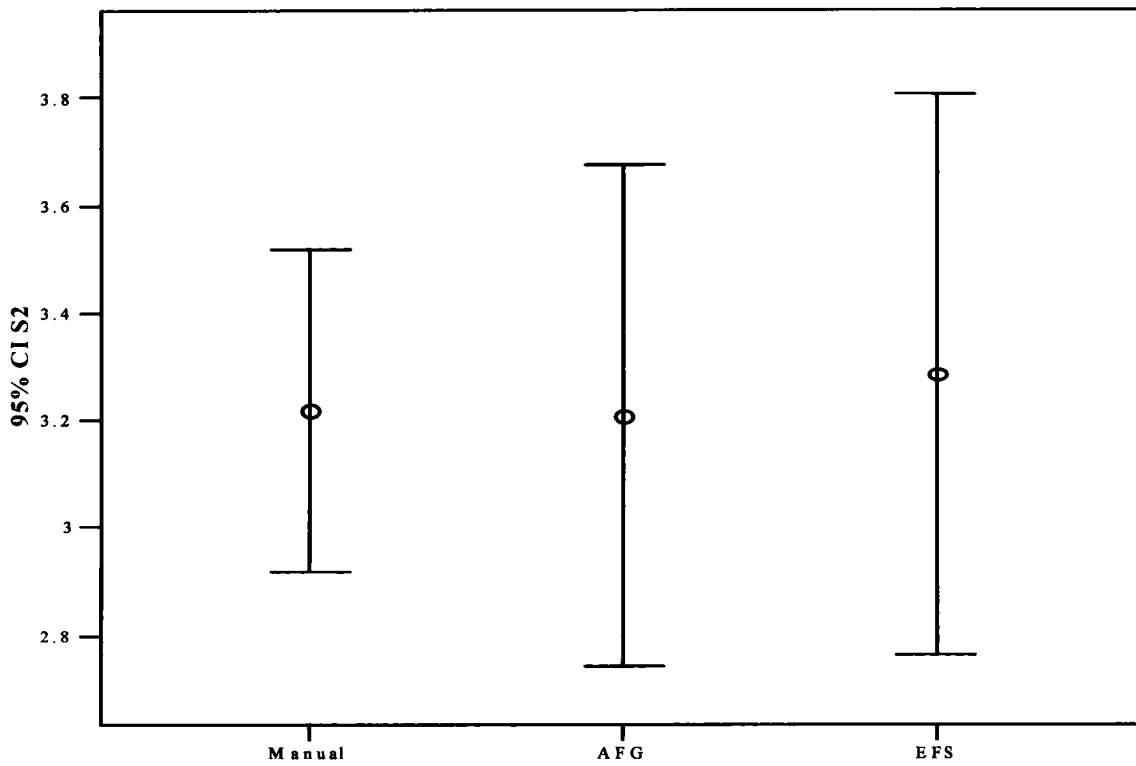

Figure 7: The coordinator, tutor and student group structures for the courses

497

Figure 8: 95% Confidence Intervals for S1

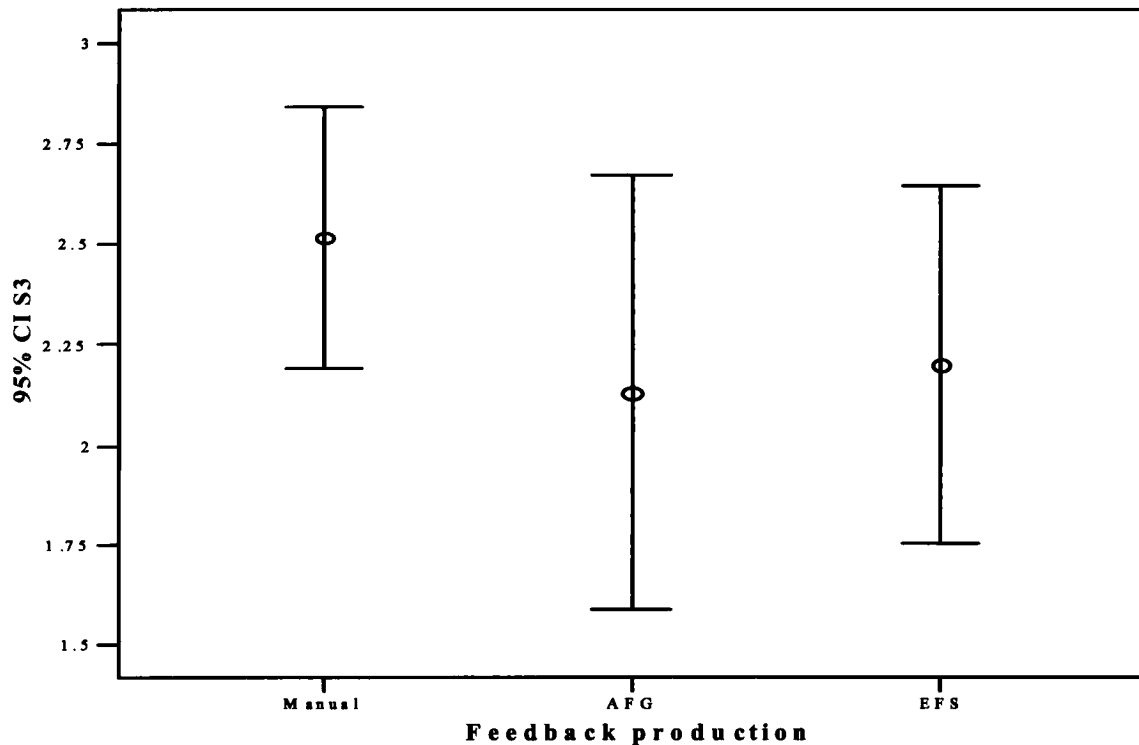**Figure 9: 95% Confidence Intervals for S2**

498

Figure 10: 95% Confidence Intervals for S3

constructive and helpful, and with no more errors than from other courses.

A potential source of variation in the data was that AFG automated feedback was produced across two separate courses. This variation was investigated by performing independent samples T tests for statements S1, S2 and S3 across the two courses for students that had received automated feedback. No significant differences between the courses were found for the mean statement values (S1: $t_{0.05, 15.739} = -2.109$, P = 0.051; S2: $t_{0.05, 22} = -1.407$, P = 0.173; S3: $t_{0.05, 18.035} = 0.834$, P = 0.415). However, there was a tendency for difference in statement S1, since its P value is below 0.1.

A further potential source of variation in the data was the use of two separate educators for the non-automated feedback data. However, independent samples T tests showed no significant differences between the mean value of the first and second educator for the statements (S1: $t_{0.05, 32} = 1.850$, P = 0.074; S2: $t_{0.05, 30} = 0.118$, P = 0.907; S3: $t_{0.05, 29} = -1.333$, P = 0.193). Again, there was a tendency for difference in statement S1, since its P value is below 0.1.

Free responses identifying the best aspects of the AFG automated feedback included the breakdown of marks into categories, and students appeared to gain a good understanding of where they had gone wrong and how to improve. Some respondents also identified fast turnaround, improved readability compared to handwriting and email delivery as positive aspects. Greater detail and personalisation was however requested, along with presenting the marks as positive scores obtained in each category rather than negative amounts lost. EFS produced very similar free responses, although its email delivery was not used and so turnaround was less rapid.

6. DISCUSSION

The students surveyed appear to have a strong positive disposition towards receiving written feedback on assessment items; the large numbers of hours worked and courses being taken by many of them are likely to contribute to this, since they may not have sufficient time to discuss their results with tutors. They appear to find the AFG feedback to be constructive, helpful and with error levels less than or equal to those for other course feedback. Their perception of manually produced feedback appears very similar, suggesting that AFG can be used without detrimentally affecting students. AFG gave comparable results to EFS from a student perspective, despite its simpler functionality. Table 3 summarises the hypotheses from which these results are drawn.

The study has a number of potential limitations. The students were queried across ten sections (tutorial groups) over two years, using three educators and two separate courses, which is likely to introduce greater variation than if a single educator and course had been investigated within a single section over one year. Also, the two tutors that did not use the automated approach used different forms of feedback; one produced word-processed documents, whilst the other gave hand-written feedback on a word-processed sheet. However, the variations between feedback across all three tutors were likely to have been reduced through the overall course coordinator checking that a sample set of

499

feedback fell within standard boundaries, together with a mark standardisation meeting. Indeed, the course coordinator ensured that all Course 1 groups were identical in areas such as course content, delivery and learning methods, and only a minor topic change differentiated the EFS assignments from the remainder. The EFS feedback was also generated by the same educator as all AFG feedback. Most importantly, the differences between courses and educators were not found to be significant.

Hypothesis	Feedback		
	AFG	EFS	Manual
H1a: Feedback is constructive and helpful	Holds; its inverse (H1b) does not	As AFG	As AFG
H2a: Feedback is more constructive and helpful than written feedback for other assessment items within courses at the same university during the semester investigated	Neither this hypothesis nor its inverse (H2b) holds	As AFG	As AFG
H3a: Feedback contains more errors than written feedback for other assessment items within courses at the same university during the semester investigated	Does not hold, although its inverse (H3b) does	As AFG	As AFG

Table 3: Hypotheses for AFG, EFS and manual feedback

Although the restriction of the study to a single university may reduce the variation in the results, it also restricts the extent to which these can be generalised. There was also a reasonable chance that statement S3 was misinterpreted by some students; this corresponded to the relative error levels in the feedback, and it would not be clear to students whether the errors were due to the marker, the system or even the course itself. Students' responses to the other two statements may also be coloured according to how well the student has performed and how diligently each marker has produced feedback.

The use of AFG by a single educator has the potential to bias the study, although the lack of significant differences between educators when manual study was used suggests that this is likely to be relatively minor. The restriction of EFS usage to the same educator offers the advantage of removing the effect of educator from the comparisons between this system and AFG.

The AFG system has a number of undeniable quality control benefits, and has been designed to maximise grading efficiency. Indeed, the effort required to set up the package compares favourably with the traditional manual approach of setting up a document containing a table of marking criteria, since the user only has to set up the mark spreadsheet with criteria columns. This contrasts strongly with the

complexities of comparable systems such as EFS and MindTrail. The performance of AFG is likely to be even better within larger class sizes, where its scalability, consistency and modification mechanisms will have greater importance.

7. CONCLUSIONS

An innovative approach to automated feedback generation has been presented and shown to give comparable performance from a student perspective to manual methods and an existing marking system; moreover, it provides quality control, efficiency and effectiveness benefits that increase with class size. The breadth of assessment types that it can be applied to exceeds that of the Re:Mark package and all systems that automate grading. Finally, it is far less complex than systems such as EFS and MindTrail, and is therefore likely to offer superior usability and efficiency.

8. REFERENCES

Allen, A., Montgomery, M., Tubman, J., Frazier, L. and Escovar, L. (2003), "The effects of assessment feedback on rapport-building and self-enhancement processes," Journal of Mental Health Counselling, Vol. 25, No. 3, p. 165.

Black, P. and William, D. (1998), "Assessment and classroom learning," Assessment in Education, Vol. 5, No. 1, pp. 7-74.

Blayney, P. and Freeman, M. (2004), "Automated formative feedback and summative assessment using individualised spreadsheet assignments," Australasian Journal of Educational Technology, Vol. 20, No. 2, pp. 203-31.

Cargill, M. (2001), "Enhancing essay feed-back using 'MindTrail' software: exactly what makes the difference in student development?," Proceedings of the Changing Identities: language and academic skills conference, University of Wollongong, Australia.

Carver, J. and Henderson, L. (2006), "Viope as a Tool for Teaching Introductory Programming: An Empirical Investigation," Proceedings of the 19th Conference on Software Engineering Education & Training (CSEET '06), Hawaii, pp. 9-16.

Denton, P. (2001), "MS Office software for returning feedback to students via email," The Higher Education Academy, Retrieved 30 May, 2005, from http://www.heacademy.ac.uk/embedded_object.asp?id=21653&filename=Denton.

Entwistle, N. J., Hounsell, D. J., Macauley, C., Situnayake, G. and Tait, H. (1989), "The Performance of Electrical Engineers in Scottish Higher Education," Report to the Scottish Education Department, Centre for Research on Learning and Instruction, Department of Education, University of Edinburgh.

Felix, U. (2001), "Beyond Babel: Language learning online," Language Australia, Melbourne.

Foxley, E., Higgins, C., Symeonidis, P. and Tsintsifas, A. (2001), "The CourseMaster Automated Assessment System – a next generation Ceilidh," Retrieved 3 November, 2006, from http://www.ics.heacademy.ac.uk/events/presentations/520_foxley.pdf.

Harman, G. and Meek, V. K. (2000), "Repositioning quality assurance and accreditation in Australian higher education," in Evaluations and Investigations Program Higher Education Division report 00/2, Department of Employment, Education, Training and Youth Affairs, Canberra.

Hattie, J. A. (1987), "Identifying the salient facets of a model of student learning: a synthesis of meta-analyses," International Journal of Educational Research, Vol. 11, No. 2, pp. 187-212.

Heaney, D. (2005), "TutorBoard - The System," Retrieved 17 October, 2005, from http://www.computing.dcu.ie/~dheaney/tutorboard.html.

Heaney, D. and Daly, C. (2003), "Mass Production of Individual Feedback," Proceedings of the Second International Conference on Principles and Practice of Programming in Java, Ireland, Vol. 42, p. 93.

Hede, A. (2005), Personal Communication.

Higgins, C., Symeonidis, P. and Tsintsifas, A. (2002), "The marking system for CourseMaster," Proceedings of the 7th annual conference on innovation and technology in computer science education, pp. 46-50.

Higgins, R., Hartley, P. and Skelton, A. (2002), "The conscientious consumer: Reconsidering the role of assessment feedback in student learning," Studies in Higher Education, Vol. 27, No. 1, p. 53.

Holmes, L. E. and Smith, L. J. (2003), "Student evaluations of faculty grading methods," Journal of Education for Business, Vol. 78, No. 6, p. 318.

Hyland, P. (2000), "Learning from feedback on assessment," in The Practice of University History Teaching, P. Hyland and A. Booth (eds.), Manchester University Press, Manchester, pp. 233-47.

James, R., McInnis, C. and Devlin, M. (2002), "Assessing Learning in Australian Universities: Ideas strategies and resources for quality in student assessment," Centre for the Study of Higher Education, University of Melbourne.

Korhonen, A., Malmi, L., Nikander, J. and Tenhunen, P. (2003), "Interaction and Feedback in Automatically Assessed Algorithm Simulation Exercises," Journal of Information Technology Education, Vol. 2, pp. 241-56.

Krucli, T. (2004), "Making Assessment Matter: Using the Computer to Create Interactive Feedback," English Journal (High school edition), Vol. 94, No. 1, pp. 47-52.

Lyall, R. and McNamara, S. (2000), "Learning tool or potplant stand? Students' opinions of learning from a CAL program in a distance education context," Australian Journal of Educational Technology, Vol. 16, No. 2, pp. 126-46.

Maclellen, E. (2001), "Assessment for learning: the different perceptions of tutors and students," Assessment and Evaluation in Higher Education, Vol. 26, No. 4, pp. 307-18.

Ortega, M., Perez, M. and Rojas, T. (2003), "Construction of a Systemic Quality Model for Evaluating a Software Product," Software Quality Journal, Vol. 11, No. 3, pp. 219-42.

"Re:Mark Online Grading and Markup Solution for Blackboard," (2006), Sciworth, Retrieved 22 May, 2006, from http://www.mydropbox.com/info/ReMark_for_Blackboard.pdf.

Robinson, P. (2002), "ReportWriter," Retrieved 7 June, 2005, from http://www.pmpfr.co.uk/reportwriter.html.

Saikkonen, R., Malmi, L. and Korhonen, A. (2001), "Fully automatic assessment of programming exercises," Proceedings of the 6th Annual SIGCSE/SIGCUE Conference on Innovation and Technology in Computer Science Education, Canterbury, United Kingdom, pp. 133-136.

Sims, R. (2000), "An interactive conundrum: Constructs of interactivity and learning theory," Australian Journal of Educational Technology, Vol. 16, No. 1, pp. 45-57.

Stevens, K. and Jamieson, R. (2002), "The Introduction and Assessment of Three Teaching Tools (WebCT, MindTrail, EVE) into a Post Graduate Course," Journal of Information Technology Education, Vol. 1, No. 4, pp. 233-52.

Thurmond, V., Wambach, K., Connors, H. R. and Frey, B. B. (2002), "Evaluation of student satisfaction: Determining the impact of a web-based environment by controlling for student characteristics," American Journal of Distance Education, Vol. 16, No. 3, pp. 169-90.

Vihtonen, E. and Ageenko, E. (2002), "Viope-computer supported environment for learning programming languages," Proceedings of the International Symposium on Technologies of Information and Communication in Education for Engineering and Industry, Lyon, France, pp. 371-2.

Wiggins, G. (1997), "Feedback: How learning occurs," in Assessing Impact Evidence and Action, American Association of Higher Education, pp. 31-9.

Williams, R. and Dreher, H. (2004), "Automatically Grading Essays with MarkIT," Journal of Issues in Informing Science and Information Technology, Vol. 1, pp. 693-700.

AUTHOR BIOGRAPHIES

Justin Debuse is a Lecturer in Information Systems at the University of the Sunshine Coast. He received his PhD from the University of East Anglia. His research encompasses topics such as data mining, automated assessment, optimisation and visual programming. He has published in journals such as Systems Research and Behavioral Science, Computer Journal, Journal of Applied Intelligence and Journal of Intelligent Information Systems.

Meredith Lawley is an Associate Professor of Marketing at the University of the Sunshine Coast. She received her PhD from the University of Southern Queensland. Her research interests include consumer behaviour, international education, e-business, e-marketing and services marketing.

501

Rania Shibl is an Associate Lecturer in Information Systems at the University of the Sunshine Coast. She was previously an Assistant Lecturer in Information systems at Massey University. Her research interests are in the areas of information systems/technology in health and decision support systems. She has published in the area of health information systems, including research in decision support systems, information overload and IT adoption. She is currently undertaking a PhD in Information Systems at the University of the Sunshine Coast.

Information Systems & Computing
Academic Professionals

STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the Journal of Information Systems Education have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.