Investigating sharable feedback tags for programming assignments

This paper presents an investigation into the usage of sharable feedback tags as a way of delivering feedback to three different cohorts of programming students. A series of research questions are investigated which include investigating any perceived benefit from students using feedback tags and investigating how students interact with their feedback. Results indicate that students with both the lower and higher marks in a cohort are more likely to opt to share their feedback and programming work than students with mid-ranged marks. A variety of reasons for and against sharing given by students are discussed. Six categories of student behaviour exhibited during interaction with their feedback have been identified in this paper. This paper has shown that feedback tags can be used successfully as a form of sharable feedback and that a number of future research possibilities exist that can extend this topic.

Keywords: programming; feedback; feedback tags; sharable feedback; Web 2.0

1. Introduction

Learning and teaching computer programming is widely recognised as being a challenging undertaking within Higher Education (DuBoulay, 1989; Robins, Rountree, J. & Rountree, N. 2003; Winslow 1996) and does not appear to have become any easier over time. This notorious problem, when considered along with the belief that feedback is "... is the life blood of learning" (Rowntree, 1987), leads the authors to believe that by changing programming feedback we may be able to better support students who are learning to program. Regardless of whether the feedback is generated internally from the learners past experiences or externally from a lecturer assessing a software project, feedback is an exceptionally important aspect of the learning process (Laurillard, 1993).

Feedback has been delivered using a variety of media for example, audio and video podcasts, e-mail and Virtual Learning Environments. All of these attempt to illicit positive changes to students learning. Many of these media deliver feedback in a way which is conceptually or physically isolated from the students' original source code. It follows from cognitive load theory (Sweller, 1994) that if feedback is separated from the students' original work a higher cognitive load could be required in its interpretation. This is because not only do students have to interpret the feedback but they also have to try and relate it back to aspects of their original work in order to take corrective action.

This article describes an investigation into the use of sharable tag based feedback delivered alongside students' original source code submissions and extends the investigative work initially presented in (Cummins, Burd and Hatch 2009). Student interactions with their feedback and the shared feedback of their peers have been recorded and investigated in this paper with the aim being to identify possible benefits associated with using tag based feedback. As an overview, this paper aims to investigate the answers to the following five research questions;

- RQ1: Do individual students perceive benefit from receiving feedback in the form of tags annotated throughout their software?
- RQ2: Do students opt-in to share their code and associated feedback?

- RQ3: Which students tend to opt-in to share their code and feedback?
- RQ4: Do students perceive benefit from having access to other students code and feedback?
- RQ5: How do students use feedback delivered in tag form?

Student perceived benefit, as mentioned in RQ1, is an important consideration when evaluating new forms of feedback. This is because inevitably the students who receive the feedback are in the best position to determine how well they were able to use it to aid in their learning. For the purpose of this investigation, perception is measured by analysis of student responses to questionnaires and focus groups. Benefit is defined as each student's perceived ability to improve based on the feedback given and their ability to understand the feedback. Furthermore, students will be asked whether they have received enough feedback in their opinion and whether it is of high enough quality. These aspects of the feedback will determine how much benefit the new form of feedback has provided in this investigation.

Exploring how students interact with their feedback and their motivation for or against sharing is a key focus of this paper. This is the motivation for including RQ2, RQ3 and RQ4 as research questions.

1.1. Literature review: tools to support assessment and feedback in programming

There are three general approaches to using technology to handle assessment feedback. These are summarised as: using software to alter an existing document by insertion of comments, using software to simulate writing in ink over the top of students work, and by delivery of a separate document that contains comments relating to the piece of work (Plimmer & Mason, 2006).

Issuing a separate document containing the feedback is by far the weakest of the systems because if references are to be made to specific aspects of the students original work they must be made with a navigational commentary as well (Plimmer & Mason, 2006) e.g. ‘In file X, line 654, you should ...’. This increases the cognitive load required for the student to understand their feedback, as they must refer to two documents simultaneously, keeping both in working memory.

A variety of systems exist to support electronic assessment of student programming work and they often adopt one of two broad strategies, known as fully automated or semi-automated.

Fully automated systems often focus on testing program correctness by running students’ submissions through predefined unit tests and returning a value of either pass or fail for each test. An example of one such system, Scheme-Robo is designed to assess students’ submissions written in the Scheme functional programming language (Saikkonen, Malmi & Korhonen, 2001). The main limitation of this and many other automated systems of programming assessment is the quality of the feedback presented to students. In many cases it is rigid and impersonal, with students finding it difficult to relate exactly what the feedback means in the context of their original work. Furthermore, the approach adopted by many automated systems ignores aspects such as style and elegance of the programming work submitted. These aspects are important for software comprehension as they inevitably determine how maintainable the software is. These aspects are important indicators of how well students are mastering the skills necessary to become competent programmers.

Some automated systems have attempted to assess the more subjective aspects of students programming work including comprehensibility (Berry & Meekings, 1985). However, these approaches are not always guaranteed to be aligned with

examiners ideas of how code comprehensibility should be assessed. For example, automated tools also are unable to determine how meaningful comments and variable names are in the context of the software project.

Due to the restrictive nature of automated systems they are often used in conjunction with examiner assessment, resulting in a more semi-automated process. Semi-automated systems attempt to leverage the speed and convenience of automated systems and combine it with examiner experience and intuition. These systems are more likely to be used in practice because they have fewer limitations and are not as rigid as some automated approaches to programming assessment. An example of one semi-automated system is the BOSS system for electronic assessment of Java programming code (Joy, Griffiths & Boyatt, 2005). This system operates by running the students code through pre-specified test cases and automatically assigning marks based on these results. However, the system does not aim to replace the examiner; on the contrary, the examiner is still an integral part of the system as they must judge the quality and style of the work submitted. A limitation of the BOSS system is that it presents student feedback as a separate, virtual summary sheet, isolated from the students' original work. As previously mentioned, this could cause a cognitive overhead for students trying to map feedback to aspects of their own work.

The Environment for Learning to Program or ELP system enables delivery of feedback in the form of a dynamic discussion that appears annotated within students programming work (Bancroft & Roe, 2006). This system is a particularly good example of how to provide feedback to programming work that is personalised, traceable and presented within the context of the students' originally submitted work. ELP preserves the context of the feedback by storing it as a discussion overlaid on the student's original submission. This reduces the cognitive overhead and allows the student to see exactly what aspect of their work is being discussed. The ELP system has demonstrated the positive impact of providing personalised, in context feedback for programming work.

2. The SWATT system

The SoftWare Assessment Through Tagging (SWATT) system, used in this paper, utilises techniques usually found in Web 2.0 systems to provide a different approach to feedback delivery for programming students. The SWATT system has been developed in conjunction with the research project presented, as a tool to support the generation and dissemination of feedback tags. These tags are then presented to students within the context of the original programming work, as shown in Figure 1.

Figure 1. A screenshot of the SWATT system showing tags annotated throughout source code

A tag is a short, fragment of human readable text which can act as a form of searchable metadata when it is attached to a resource, which in this case, is a student programming submission. Tags are keywords that are used to describe a particular resource in some way. A key distinction between a tag and traditional notions of metadata is that there are few, if any, formal restrictions imposed on the format of tags. As a result, tagging resources is a particularly easy and flexible process that requires almost no training or instruction. An example feedback tag is "good javadoc", which praises a student's usage of source code documentation. Typically, this tag would be placed at a location where the student has included clear or comprehensive documentation for a Java class or method.

The use of tags as a form of feedback leads to a number of benefits including the easy generation of visual summaries of the feedback data, for example tag clouds. A tag cloud is a weighted list which displays each tag in various font sizes according to the tags frequency within the dataset. That is, the more often the tag is appears in the dataset, the bigger the tag will be rendered in the weighted list. These visualisations may be useful for students as a summary of their feedback, and can act as a starting point for students in exploring their feedback in greater detail. Figure 2, shows an example of feedback tags visualised as a tag cloud. It is clear from this that the feedback tag "use javadoc" was frequently occurring in the data set.

Figure 2. A feedback tag cloud generated using the Wordle.net online tool

The analysis of a cohort's tag cloud may be useful for lecturers who may wish to identify areas of focus for remedial teaching based on weaknesses highlighted from students' assignments.

The SWATT system is similar to the ELP system in that the feedback is embedded within the students' original submission. However, there are two important differences; the first is that the feedback generated is in tag form and not full sentences. This means that the feedback generated is potentially reusable and is composed of two or three words or a very short phrase. The reusability stems from the fact that feedback tags are often generic and can be applied to different students work. The reuse of the feedback tags in different situations provides a greater context from which students can understand the tags. The second difference is that these feedback tags along with the students' submitted work can be shared, providing the opportunity for students to be exposed to significantly more feedback and programming code than they would be normally. Feedback tags clouds, such as the one shown in Figure 2, are also able to provide a quick and intuitive summary of students' feedback.

A negative aspect of the SWATT approach is that the feedback tags used are often shorter than traditional feedback comments. This may mean they are less detailed and potentially ambiguous. This is especially the case when considering the sentiment of feedback, that is, whether one perceives feedback as being positive, negative or neutral. These perceptions may vary between examiners and students causing a level of ambiguity (Cummins, Burd & Hatch, 2010). However, the reusability of feedback tags and the ability for them to be shared between students mitigates this problem by putting the feedback into a richer context. This additional contextual information may support students in being able to infer additional meaning from their own feedback and as a result reducing the ambiguity associated with shorter feedback.

The SWATT process for feedback is summarised as;

(1) Students submit their completed source code to the online SWATT system.
(2) Examiners use a plugin specially developed for a popular programming development environment to download feedback tag suggestions, annotate students work and upload the annotations to the SWATT online system.
(3) Examiners can make the feedback visible after the marking process has completed. At this moment students can view and explore their feedback online.
(4) Students then can opt to share their feedback and associated work anonymously and in so doing are allowed to view the shared feedback of their peers.

In addition to providing feedback in tag form the SWATT system also facilitates the sharing of this feedback information along with the students' programming work.

2.1. Sharing feedback tags and source code

As with many Web 2.0 systems, the more users who contribute and share within SWATT, the more useful the system can become. Students' feedback and work defaults to being private in the SWATT system. This means students must actively opt-in to the sharing aspects of the system. The SWATT system operates a blanket all or nothing sharing model. This means when students opt to share their work, it is shared anonymously with the entire cohort and individuals have no further control over which of their peers are able to see their feedback or work. The act of sharing is therefore permanent within the SWATT system. This is to prevent students from briefly sharing their work in order to view the other shared resources and then un-

sharing it again. Students are made aware of this restriction and must acknowledge it before they can opt to share. Only students, who have opted to share, are able to view other students' shared feedback or use any analysis features provided by SWATT which require information from other students' feedback or work.

The act of sharing within the SWATT system increases the information available to the individual user but also because that user has had to share their feedback, the amount of information available to the sharing community also increases. This additional information enables more comprehensive analysis capabilities as more data is available. Strategies such as: co-occurrence analysis, tag clustering, and frequency analysis can be used to identify links between individual feedback tags. This may be useful for supporting students' understanding of their feedback and programming work.

A simple similarity metric has been developed to encourage students to share and explore the feedback of their peers. The system provides a percentage similarity value for each submission irrespective of whether the student has opted to share. This value shows how similar the feedback of the student who is currently using the system is to all other submissions. This simple similarity metric is calculated as follows;

$$MySimilarityPercentage = \frac{|(MyTags \cap OtherPersonsTags)|}{|MyTags|} 100$$

The enhanced access given to sharers enables them to perform a number of additional functions including: viewing other shared feedback in comparison to their own (see Figure 3), and performing searches on the corpus of shared feedback tags and source code information.

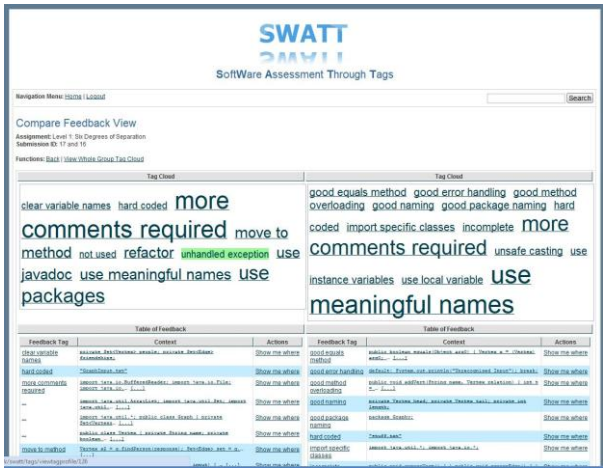

Figure 3. Screenshot showing students comparing feedback summaries using SWATT

3. Investigation design

In order to answer the research questions posed by this paper a number of research methods are employed including questionnaires, focus groups and observation data collected electronically from student usage of the SWATT system.

The research presented in this paper is conducted using an iterative investigation design. The results of three investigations conducted with three different cohorts of students, each at different stages of their undergraduate degree programme are considered. The first two investigations (I_1 and I_2) are used as preliminary or dry-run investigations to help direct and validate the planned research methods for I_3. The final investigation (I_3) provides detailed results which make use of the experience and data gained from I_1 and I_2. Table 1 summarises the key aspects of each investigation.

Table 1 Summary of investigations

	I_1	I_2	I_3
Number of Students Involved	67	21	45
Year of Study	2nd Year	1st Year	1st Year
Summative / Formative	Summative	Formative	Summative
Group Project or Individual	Group Project	Individual Project	Individual Project
Feedback Release Process	Marks released before feedback tags	Delay over summer holidays between submission and feedback release	Feedback released first and then marks followed 2 weeks later.
SWATT Data Collection	✓	✓	✓
Questionnaires	✓	✓	✓
Focus Groups			✓

One of the main differences between the three investigations was the order and timing that the feedback was released. In the first instance, Investigation 1 (I_1), the feedback tags generated from the SWATT system were released approximately two weeks after the summative marks were given to the groups participating in the assignment. In I_2, the students submitted their formative assignment at the end of an academic year and were given their feedback after the 3 month summer holiday. In I_3, the students were given their feedback tags quickly and their assessment marks followed two weeks later due to the time required for the mark verification process. These differences will undoubtedly impact how students in each cohort interact with their feedback.

Another difference between I_1, I_2 and I_3 is the purpose of the assessment. In I_1 and I_3 the assessment was designed to be summative with marks generated to measure how well students have met particular learning outcomes. I_2 was formative and therefore the assessment was designed to provide constructive feedback only and not provide a quantitative measure of how well students have met learning outcomes.

Another equally important difference between the three investigations is the mode of completion. That is whether the assignment required group work or was intended as an individual piece of work. I_1 was a group project involving 12 groups of between five and six students. All other investigations were individual assignments requiring no group work.

Whilst all investigations involved undergraduate computer science or software engineering students, each one involved a cohort that was at different stages of their programme of study. In I_1 the cohort was towards the end of their second year of study. I_2 involved a group of students who were at the end of their first year of study and I_3 involved students who had completed one third of their first year of study.

The timing, purpose and mode of completion differences described, were largely outside the control of the researcher. However, for I_3, the researcher was able to influence the assessment activity to ensure that the conditions were as appropriate as possible for using the SWATT system. This simply meant using the results from I_1 and I_2 to direct I_3. For example, the decision to use an individual project instead of a group project was motivated from the results and student feedback from I_1. The decision to trial the SWATT approach using a summative assignment was based on the results from I_2. This resulted in an iterative investigation design that culminated with the results from the final investigation I_3.

4. Investigation method

This study has used an iterative model of investigation such that the improvements gained from each investigation are applied to the subsequent one. Each investigation was successive and occurred sequentially. For each investigation, the same core method was followed which involved examiners annotating students source code submissions with feedback tags and then releasing them for student interaction via the SWATT system. In all three investigations student interactions with the SWATT system were recorded electronically.

At the time of feedback release, students were given a questionnaire to complete which focused on recording their perceptions of the new feedback system and their interaction with their own and other students' feedback. In I_1 and I_2 the questionnaire was delivered without any incentive. However, due to consistently low response rates threatening the validity of the study, a prize draw to win gift vouchers was included in I_3. The questionnaires were all administered electronically and were anonymous. This means that individual questionnaire data cannot be linked to data collected from an individual's interactions with the SWATT system.

After I_1 it was clear from student responses that the SWATT approach to feedback was less useful for a group project. It became clear that students were happy to discuss the feedback within their groups and had little need to use the SWATT system as a means of sharing feedback. Students commented that they would have found the sharing aspects of the system more useful in individual assessment where they would not have had a team to share and discuss their feedback with. This suggestion was applied in I_2 and an individual formative assessment was used instead of a summative group project.

In I_2 it appeared the formative nature of the assignment caused fewer students to submit work at all. This meant that whilst more students viewed their feedback, not all students in the cohort submitted something and as such did not receive any feedback. As a result of the low engagement with formative work, it was decided to in I_3 to utilise a summative project and to make sure it was individually assessed.

After I_1 and I_2, it was decided to include focus groups as an exploratory research method in order to triangulate the data gathered using the other research methods and to better explore some of the reoccurring issues appearing in the data. One particular issue to be explored in focus groups was the motivation of why students did not opt to share their feedback and work. Questionnaires provided an insight into why students did choose to share but not as much detail on why they did not. The focus groups were selected by random invitation according to whether students had shared or not, depending on the topics to be considered in the group. One focus group focused exclusively on why students opted not to share and as such only non-sharing students were invited. The other focus group had a mixture of sharers and non-sharers and investigated the broader topics of the SWATT approach to feedback.

5. Results and discussion

The results of this paper are presented with emphasis on I_3 due to the higher response rate and the higher validity associated with the results. However, data from I_1 and I_2 which is relevant to discussion may be included when appropriate. Table 2 summarises the results for each of the three investigations.

Table 2 Table summarising results for all investigations I_1, I_2 and I_3

	I_1	**I_2**	**I_3**
Questionnaire Response Rate	21% (14/67)	38% (8/21)	71% (32/45)
Number of Sharers	42% (5/12) groups	43% (9/21)	42% (19/45)
Average Mark (Sharers)	80% (SD=11.22)	N/A	62% (SD=14.12)
Average Mark (Non-Sharers)	76% (SD=9.14)	N/A	58% (SD=10.44)
Students viewing their feedback via SWATT	58% (39/67)	95% (20/21)	100% (45/45)

5.1. Student perceptions of feedback tags

As previously mentioned, students perceived benefit is evaluated in terms of: how easy it was to understand the feedback; the perceived quality of the feedback; how sufficient the amount of feedback received is; and the perceived ability for students to improve based on the feedback received.

The questionnaire results for I_1 and I_2 have been only briefly considered in this paper due to the low response rates. The general trend is that students' perceived benefit increased with each successive investigation. The results for the I_3 questionnaire concerning students' perceived benefit are summarised below;

- 94% of students thought that the tag based feedback was "Easy" or "Very Easy" to understand.
- 72% of students indicated that the feedback they received was of a "Very Good" or "Good" quality. 22% of respondents reported that it was of an "Average" quality and the remaining 6% indicated that they perceived it as being "Poor" in quality.
- 56% of students indicated that the quantity or amount feedback they received was "About Right", 41% reported that it was "Not Quite Enough" and finally one respondent (3%) reported that the feedback they received was "Far From Enough".
- 81% of students reported that receiving this type of feedback was "Very Useful" or "Useful" in helping them to improve their work. 13% indicated a neutral response and the remaining 6% reported it as being "Not Very Useful".

The most frequently reported benefit associated with the SWATT approach to feedback is summarised in a response from a student questionnaire, "You can see a general theme to how you've done instantly but then drill into certain areas to get more information". This student has identified that by using the feedback tag cloud they were able to view a high level summary of the assignments feedback. This reportedly enabled them to navigate their feedback more easily. The ability for students to be able to focus on where the feedback was used specifically within the

original programming work is highlighted as being a positive aspect of the SWATT approach.

Students reported, especially in the questionnaires, that feedback tags had less inherent meaning and were only really meaningful in the context of where they were associated within students submitted source code. This led to some ambiguity especially when technical terms were used that students were unfamiliar with for example “high coupling”. High coupling describes a situation where more than one objects in a program are highly dependent on one another, which could make maintenance problematic.

Within one of the focus groups in I_3, the students reported that having the feedback tags released before the summative marks resulted in all participants engaging in a process of estimating their marks based solely on their feedback tags. This yielded a side-effect which was some students actively engaged with their feedback tags and even researched the meaning of some of the more vague or technical terms in order to understand them. The motivation behind this was reported as being the students had to understand the technical terms in order to factor them into their estimate of their summative marks. As a result, participants in the focus group reported that they felt they had remembered more of their feedback and had learnt more from it overall.

5.2. Sharing feedback

In all investigations the proportion of students opting to share their work is consistently around 43%. This is despite the use of different cohorts and assignments.

Throughout the questionnaires and focus groups, detailed reasons were given as to why students opted into the sharing scheme. These include the following key motivations;

- **Checking up on examiners** – Students, especially in preliminary investigations, reported they wanted to see what the examiner was commenting on in other peoples work and to check for examiner consistency.
- **Competition** - Some students desired to see how well they had done in comparison to others.
- **Confidence** - Some students opted to share their work for no apparent benefit to themselves. They reported that they did not actually look at any other students work but felt as though they wanted to help other people by sharing theirs. Automatic collection data confirmed that some students did share and did not look at anyone else’s work / feedback.
- **Curiosity** - Some students reported that they were just curious as to how their peers had approached the same problem using different solutions.
- **Learn From Others Mistakes** - Some students reported that they had a desire to learn from other peoples mistakes and ensure they did not make them in future assessments.
- **Understand Feedback Better** - At least one student in the questionnaire reported they had shared so that they could see how other students feedback was similar to their own, for the purpose of better understanding their own feedback.

Many of these motivations were reflected consistently through the preliminary investigations in the questionnaires.

The questionnaires on their own however, did not provide detailed information as to why students did not share and, as a result, a dedicated focus group was run to explore this topic in I_3. The results provide an insight into why some students would not want to share their work and feedback. These are summarised below;

- **Distrust of Anonymity** - Some students reported that they did not trust that their peers would be unable to identify them through their code.
- **Fear of Being Discontent** - At least one student, who completed the questionnaire, was concerned that they would realise their work was significantly inferior to that of their peers, if they could see other peoples' work and as such did not want to know how well others had done.
- **Forgetfulness** - At least two of the participants selected in the focus group reported that they had actually intended to share their work but had forgotten to login and select the option.
- **Lack of Confidence** - Both in questionnaires and focus groups at least one student reported that they did not think their work was good enough to share and were worried about the standard of their source code. Some participants suggested that they no one could possibly gain benefit from seeing their work and so decided to not share.
- **Lack of Interest** - One participant reported that they were uninterested in other peoples work or feedback as they could not see how it would help them in their learning.
- **Paranoia** - There was a concern expressed that a student could share their work and if it was regarded as being inferior by a group of peers and this inferiority was discussed in public the owner of the work could possibly overhear and would feel victimised personally, even if the peers did not know who it was, the owner would.
- **Social / Informal Sharers** - At least two participants in the focus groups confirmed that they had shared their programming source code informally and outside of the SWATT system. They said they preferred discussing face-to-face their feedback and work with their peers. In some cases students simply logged into the system at the same time as a friend to look through each others feedback and work.

It is unclear how many individuals in each cohort held the same motivations for and against sharing. The data that these ideas have been extracted from has come from focus groups and individual questionnaire responses. These, unfortunately, do not indicate how widespread the feelings are, only that they exist within a sample of the participants.

The most surprising finding was that of ‘paranoia’ as described above. This fear was relayed via questionnaire and was completely unexpected by the researchers. In most cases a majority of these fears would be alleviated if students were able to select exactly who was able to view their feedback. During the focus group held with exclusively non-sharer participants it was suggested that a social networking style sharing, for example that employed by ‘facebook’, would encourage more sharing between individuals. All participants in the focus groups, who did not intend to share, agreed that they would have liked to select individuals to share their work with and were largely apprehensive of blanket sharing across the whole cohort.

It is interesting to note that 19% of questionnaire respondents, when asked if anonymity made a difference to their decision to share, stated that they would not share their work or feedback, no matter what. A majority of respondents reported that they would still have shared their work irrespective of whether it was anonymous or not. This could indicate a desire to learn from each other or perhaps it could indicate that the students were proud of their work. A total of 28% of respondents confirmed that they would not have shared, had the system not provided some degree of anonymity.

As previously mentioned, 43% of a cohort appears to be a common proportion across all investigations conducted. It is clear from this that on average less than half of all students in a cohort have shared their feedback and programming work with their peers. However, based on questionnaire data and focus groups, the comments from those who did opt to share their work were largely positive. All students who shared their feedback and work and who viewed other student's feedback reported a benefit. This was mainly from seeing the different ways of implementing similar projects and the associated feedback from examiners. This opinion was also reflected in both of the preliminary investigations.

5.3. Investigating sharers and non-sharers

Statistical analysis using an independent sample T-test for I_1 shows that there is no statistically significant difference between the marks of those who share (M=80.60%, SD=11.22) and the marks of those who do not share (M=76.29%, SD=9.14). The assignment set in I_2 was formative and as such no summative marks were generated. This means that the results from I_2 cannot contribute to this statistical analysis. In I_3 there is also no statistically significant difference between the marks of those who share (M=62.05%, SD=14.12) and the marks of those who do not share (M=58.15%, SD=10.44). However, the low significance is likely due to the small sample sizes of the groups of sharers and non-sharers.

Despite the low statistical significance, it is clear from Figure 4 and Figure 5 that there is some pattern to the assessment marks students achieve and whether or not they share their work and feedback. Sharers appear to have a slightly higher mean assessment marks in both I_1 and I_3 which indicates that on average students with higher marks are more likely to share. However, the low statistical significance means that the average marks between the two groups must be interpreted with caution.

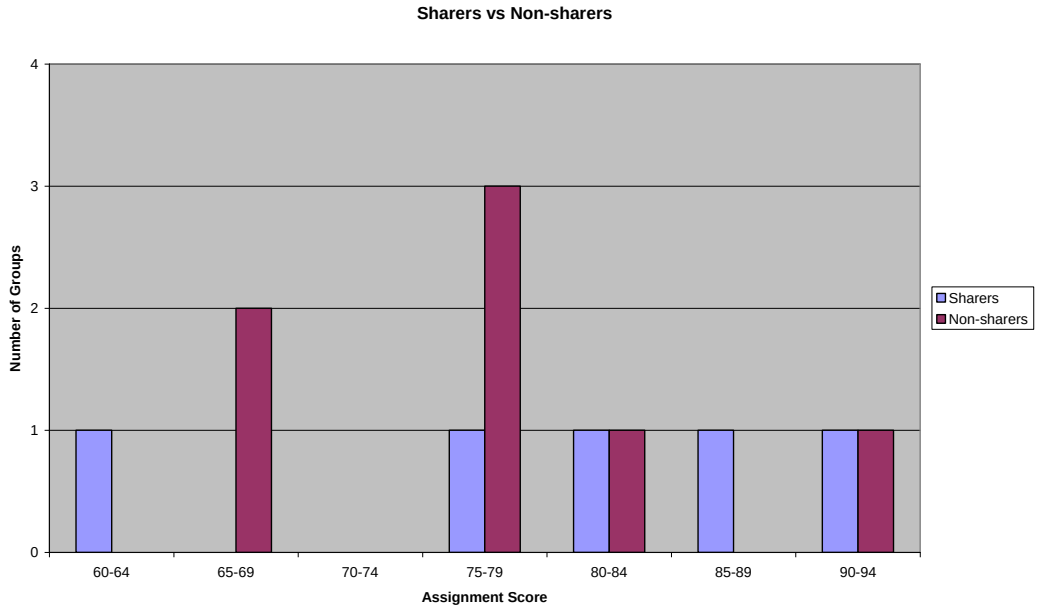

Figure 4 Graph showing sharers vs non-sharers and assignment results (I_1)

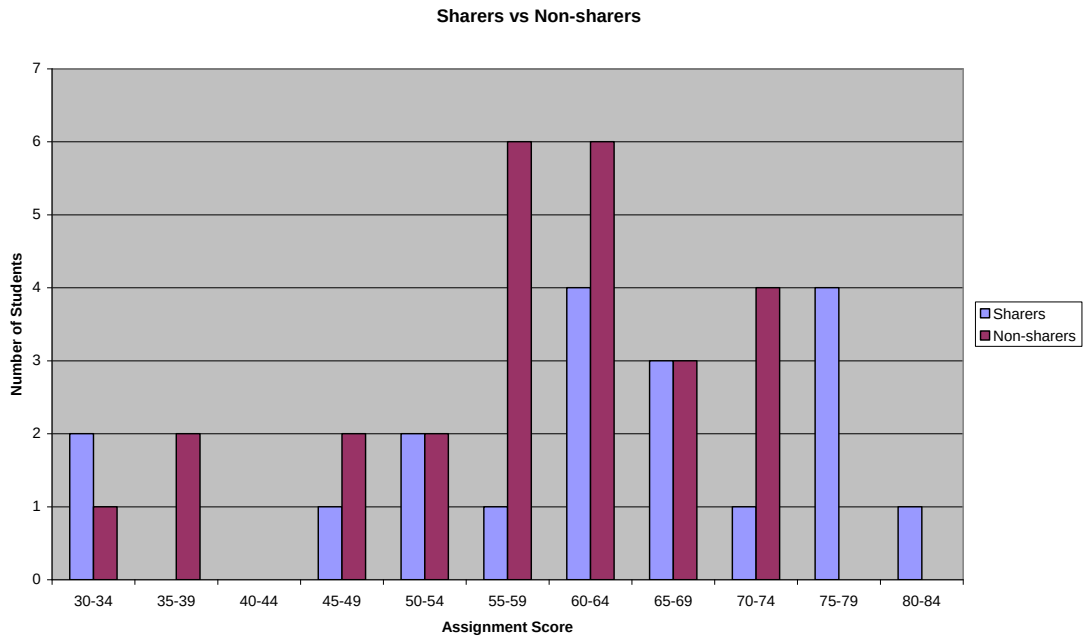

Figure 5 Graph showing sharers vs non-sharers and assignment results (I_3)

It appears from Figure 4 and Figure 5, as though a majority of students who do not opt into the sharing scheme achieve mid ranged marks in both I_1 (75%-79%) and I_3 (50%-59%). It therefore, follows that students who share in these investigations are more likely to have achieved either the higher or the lower marks. This is interesting that this trend is maintained in I_3 since the students had all made their choice on whether to share before they received their assessment score. In I_1 the students were aware of their assignment score in advance of deciding whether or not they should share.

5.4. Investigating student interaction with tag based feedback

The results collected, especially from I_3, show that students interact with their feedback in a variety of different ways. The information collected has lead to the identification of a number of different categories of students which are shown in Table 3. The frequency of students in some of these categories is unknown since many of the categories were identified from focus group discussion and individual questionnaire responses. It is therefore unclear how widespread the behaviour is the categories that were not detected through exploration of the system usage data.

The group 'Explorers', introduced in Table 3, was detected by analysing the automatic data collected from the system usage and students within this group shared immediately. They then appeared to over the course of the month systematically explore the feedback that was shared by their peers. It also appeared that they took interest in reviewing the source code submitted by other students. This group was equally small with only two apparent cases where this type of prolonged usage of the system occurred.

Table 3 Categories of behaviour observed from student feedback interaction

Category Name	**Description**
Explorers	This group of students appeared to repeatedly login to the system over a wide spread of dates and times and on each occasion they explored one of the projects shared by their peers. Often comparing it to their own work. This group of students viewed both feedback tags and the associated source code.
Informal Sharers	These students decided not to share their work using the SWATT environment; instead they informally discussed and shared their work and feedback with their friends.
Librarians	Some students, who did not share their work, reported that they had used the SWATT system as a personal library of source code that they could reuse or look at to make improvements to their future work.
One-off Viewers (Non-Sharers)	Students in this group logged in once and explored their own feedback and explored it in the context of their source code but did not use the system more than once.
One-off Viewers (Sharers)	Students in this group logged in once and explored their own feedback and other people's feedback and source code but did not use the system more than once. In this case it was clear students were more interested in viewing the feedback tags of their peers not necessarily the associated source code.
Surface Users	Students in this group simply logged in once, looked at their feedback tag cloud and did not at any point explore the system or view their tags alongside their own work.

This process of informally sharing was quite common and was confirmed through the results from the focus groups. It was mentioned that students preferred showing their friends and discussing their work and feedback in a face-to-face environment. Students in this cohort have admitted to using the SWATT system to share their work and feedback but the sharing occurred by simply showing their friends the screen where they had logged on instead of using the sharing functionality provided and recorded by the system. This type of sharing was unmonitored and would not have been detected using automated data collection methods.

It became apparent from both data collected from the usage of the system and the focus groups that some students used the SWATT system purely to view their own code on a regular basis. After further investigation, it appeared that the students were using the system as a central point where they could access their code to be reused in different programming work. These students have been labelled 'librarians' as they seem to have used the system to keep a personal library of their work and feedback. Thus the students had adapted the system to suit their own purposes as a central source code repository of their work and attached feedback. Two participants in the focus groups reported that they had used the SWATT system on multiple occasions in order to make sure they were not making the same mistakes again in their current programming work.

The two most commonly noticed groups of students are the One-off Viewers (Non-Sharer) and (Sharer) groups. Students who have been classified as being apart of these two groups used the system once to view their feedback and/or the feedback of their peers. It is clear that there is a subset of students apart of this group who

viewed their own feedback and then opted to share their work but at no point viewed any other students work in exchange. One such student was a participant in a focus group and simply stated he was happy for other people to see how he had approached the problem but had no need to explore other peoples' solutions. However, upon further discussion it became apparent that the student had in fact shared their programming work informally with their friends. This group of users appeared to explore their own or others feedback in detail but they only did so once.

The 'Surface Users' group is a small group of students detected through reviewing the automatic data collected from the systems usage. Two students out of the cohort appeared to login and view their feedback tag cloud and summary page but did not perform any other interactions with their feedback. This includes not opting to share their feedback. Due to the logging being anonymous it is unclear as to what other factors may have influenced this behaviour. The term `Surface Users' has been borrowed from educational literature specifically that of Deep and Surface learning (Marton and Säljö, 1976) as it implies students in this group have only glimpsed the surface of their feedback and have not fully explored the meaning of it.

It is clear from this investigation that students adapt and use the SWATT system in different ways to try and effectively augment their personal learning style. For some of the students, it is clear that they treated the tag based feedback as any other type of feedback and looked at it once and did not look at it again. However, for a majority of students whilst they only used the system on one occasion they did appear to interact with their feedback and explore it thoroughly during that one session.

It is clear from researcher observation as well as student opinions from focus groups, that the timing and order of release of the feedback has a significant effect on interaction with feedback. The timeliness of feedback is identified in the literature as being crucial to enabling students to use it effectively (Rowe & Wood, 2007). The research presented in this paper confirms this, especially in the results of I_2 and I_3 where there was a significant delay in delivery of feedback in I_2 and very little delay in I_3. Releasing the assignment marks in advance of the feedback resulted in less than 60% of students in I_1 viewing their feedback when compared to 100% in I_3. However, it should be noted that this comparison may be misleading since it is expected that I_1, being a group project, some students may have viewed the feedback in a group situation and not logged into the system using their individual account.

It was reported in focus groups that students thought they had engaged more with their feedback because the feedback tags were released before the marks. A majority of participants agreed that they had engaged in the process of estimating their marks based only on their feedback tags. In so doing, it forced students to really try to understand the feedback so they were better able to estimate from the feedback what the examiners marks may be.

5.5. Exploring the research questions

The research questions outlined at the beginning of this paper are addressed in this section, in the light of the results and discussion presented.

RQ1: Do individual students perceive benefit from receiving feedback in the form of tags annotated throughout their software?

Largely, students in the I_3 were satisfied with their feedback when it was delivered using the SWATT approach. The amount of feedback received was the only aspect which yielded a mixed opinion between students with only 56% of students in I_3 reporting it as being adequate, the remaining thought that more was needed. This

outcome is very similar to the results gained in I_1 and I_2. It should however, be noted that students in focus groups unanimously agreed that they prioritised high quality feedback which they could use to improve their learning over quantity.

The primary benefit of the SWATT approach to feedback tagging as reported by students was the ability for the SWATT system to present a high level overview of the student's feedback in the tag cloud. As well as the ability of the SWATT system to enable focused exploration of the feedback or "zooming in" on specific feedback from the tag cloud and seeing it in the context of the students original work.

The primary disadvantage as reported by students is the inability for some feedback tags to provide feedback without additional metadata or external research being required by the student. A few students complained that they could not take immediate corrective action because they had to research the meaning of a feedback tag. However, it was noted in the focus group that half of the participants found the activity of researching the technical feedback tags as being constructive to their learning and increased their overall engagement with the feedback.

It was clear from the preliminary investigations that students prefer this type of feedback when delivered for summative individually assessed projects in contrast to group projects or formative assignments. The reason that there is more engagement in summative projects could be that there is a higher perception of importance associated with it from students since its result directly contributes to their qualification result.

Overall, it can be concluded that students surveyed, especially in the I_3, did perceive a significant benefit in receiving feedback in the form of tags allocated throughout their source code, this is especially true as 81% reported they were able to improve their work using the feedback tags. The ability for students to improve their learning in someway from the feedback is the most important criteria for success as reported from students in focus groups.

RQ2: Do students opt-in to share their code and associated feedback?

It is clear that in I_1, I_2 and I_3, roughly the same proportion of students, 43%, have opted to share their feedback and work. The reasons identified by students for and against sharing are interesting and give an indication as to why some students were reluctant to opt into the sharing scheme. It is expected that blanket sharing to potentially the entire cohort was daunting and caused some reluctance.

RQ3: Which students tend to opt-in to share their code and feedback?

Whilst statistical analysis does not provide a statistically significant result, there is a visible pattern in students who opt to share their work. It appears as though more students who have achieved lower or higher marks within the cohort tend to opt to share. This is less so with student achieving a mid-range mark. This is in terms of the proportion of students opting to share vs those opting not to in each mark range.

RQ4: Do students perceive benefit from having access to other students code and feedback?

In the final investigation, 73% of students who shared their work reported that they did find benefit in seeing the feedback and work of their peers. The remaining 27% stated that they did not find benefit because whilst they had shared they had not at the time of completing the questionnaire looked at anyone else's work and so were unable to comment. These results suggest that on the whole those who did share their work and looked at their peers work and feedback did find some benefit in doing so.

RQ5: How do students use feedback delivered in tag form?

It appears that different students have used the SWATT system in different ways depending on their personal learning styles. The behaviour observed from the

data and identified from discussions in the student lead focus groups has resulted in this paper defining six categories of student interaction with tag based feedback. These are listed in Table 3. This list may not be exhaustive as only two focus groups were held and as such some student behaviour may not have been captured.

6. Threats to Validity

One of the threats to validity in this study is the low questionnaire response rate for I_1 and I_2. This however, has been mitigated by the paper focusing more on the results from I_3 which yielded a higher response rate.

Since each investigation had different circumstances associated with it, for example the timing of release and order of release of the feedback varied. It is difficult to make direct comparisons between the investigations, for example when considering the difference in assessment marks between those who shared and those who did not; I_1 was a group assessment and I_3 was an individual assessment. This means that there are added complications in I_1. An example could be the case that an individual within the group may not wish to share and it would be morally difficult for a group to go against an individual's privacy wishes. However, in I_3 it was entirely an individual decision on whether to opt into the sharing functionality of the SWATT system. The different circumstances surrounding each investigation mean that comparison between them is difficult. However, the results presented can be considered independently and provide a foundation for future research.

7. Conclusions and Further Work

This paper has outlined the key findings of a three year research project investigating the use of sharable feedback tags as a means of delivering feedback to programming students. The SWATT approach exploits the popularity and familiarity of Web 2.0 tagging technologies in order to deliver in-context, sharable feedback to programming students.

The ability for students to share their feedback and associated source code was consistently used by about 43% of the cohort in all three experiments, despite them involving different students. A number of different reasons for and against sharing assessment feedback in this way have been recorded and described in this paper; some of these were unexpected and provide an interesting insight into how students perceive their assessed work and feedback.

The SWATT system has been used by a number of students and many of these have adapted the system to suit their own learning needs by interacting with feedback tags in different ways. These different approaches to interacting with the SWATT system have been investigated and described in this paper. The different interaction groups that have been discovered show a range of behaviours and ways of interacting with tag based feedback, which were unexpected and warrant further experimental research. This further research could focus on quantifying how many students exhibit qualities that match to each behaviour group identified.

Further developmental work could include implementing a 'facebook' style of sharing to enable students to have finer control over who can view their work and feedback. This may result in a higher level of sharing between individual students, although may yield a reduced quantity of feedback available to each student.

This paper has provided a significant foundation for further research into tag based feedback. In addition to this, results have been presented to help classify student approaches to interacting with tag based feedback. Exploration of students' perceptions of tag based feedback has led this paper to conclude that this feedback

strategy is a new and exciting approach for delivering feedback to individual programming assignments.

References

Bancroft, P. & Roe, P. (2006) Program annotations: feedback for students learning to program. Proceedings of the 8th Austalian conference on Computing education. Hobart, Australia, Australian Computer Society, Inc.

Berry, R. E. & Meekings, B. A. E. (1985) A style analysis of C programs, Communications of the ACM, 28(1), pp80-88.

Cummins, S., Burd, L., and Hatch, A. (2009) Tag Based Feedback for Programming Courses, SIGCSE Bulletin, 41(4) (Dec. 2009), pp 62-65

Cummins, S., Burd, L. & Hatch, A. (2010) Using Feedback Tags and Sentiment Analysis to Generate Sharable Learning Resources, Proceedings of the 10th IEEE International Conference on Advanced Learning Technologies, Sousse, Tunisia.

Duboulay, B. (1989) Some Difficulties of learning to program. In Soloway, E. & Spohrer, J. C. , Studying the Novice Programmer (pp 283-299), Sussex, England, Lawrence Erlbaum Associates Inc. Publishers.

Joy, M., Griffiths, N. & Boyatt, R. (2005) The boss online submission and assessment system. Journal on Educational Resources in Computing (JERIC), 5(3).

Laurillard, D. (1993). Rethinking university teaching: a framework for the effective use of educational technology. Routledge, London.

Marton, F. and Säljö (1976) Qualitative differences in learning: 1 - outcome and process. British Journal of Educational Psychology, 46, 4-11

Plimmer, B. & Mason, P. (2006) A pen-based paperless environment for annotating and marking student assignments, Proceedings of the 7th Australasian User interface conference, Hobart, Australia, Australian Computer Society, Inc.

Robins, A., Rountree, J. & Rountree, N. (2003) Learning and Teaching Programming: A Review and Discussion, Computer Science Education, 13, pp137-72.

Rowntree, D. (1987) Assessing Students: How Shall We Know Them?, Nichols Pub Co.

Rowe, A. D. & Wood, L. N. (2007) What feedback do students want?, The Australian Association for Research in Education (AARE) International Educational Research Conference. Fremantle, Austrailia.

Saikkonen, R., Malmi, L. & Korhonen, A. (2001) Fully automatic assessment of programming exercises, ACM SIGCSE Bulletin, 33(3), pp133-136.

Sweller, J. (1994) Cognitive load theory, learning difficulty and instructional design., Learning and Instruction, 4, pp295-312

Winslow, L. E. (1996) Programming pedagogy: a psychological overview. ACM SIGCSE Bulletin, 28(3), pp17-22.