

Using statement banks to return online feedback: Limitations of the transmission approach in a credit-bearing assessment

Philip Denton^{a*} and Philip Rowe^b

^aFaculty of Science, Liverpool John Moores University, Liverpool, UK. ^bSchool of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, UK.

Philip Denton, Faculty of Science, Liverpool John Moores University, Byrom St, Liverpool, L3 3AF, UK. 0151 231 2035, p.denton@ljmu.ac.uk.

Philip Rowe, School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom St, Liverpool, L3 3AF, UK. 0151 231 2069, p.h.rowe@ljmu.ac.uk.

Philip Denton is the Associate Dean (Education) for the Faculty of Science at Liverpool John Moores University (LJMU). He is interested in the development and evaluation of electronic tools to assist with assessment and to encourage student engagement with tutor feedback. Philip Rowe is a Reader within the School of Pharmacy and Biomolecular Sciences, LJMU. His main research interest is statistical methods related to pharmacy.

Using statement banks to return online feedback: Limitations of the transmission approach in a credit-bearing assessment

Electronic marking tools that incorporate statement banks have become increasingly prevalent within higher education and their advantages are considered. In an experiment, printed and emailed feedback was returned to 243 first year students on a credit-bearing laboratory report assessment. A transmission approach was used, students being provided with comments on their work but no guidance as to how they should use these remarks. A multiple choice question (MCQ) test, undertaken before and after the return of feedback, was used to measure learning. Although returned comments included model answers to the MCQs, test scores showed no overall enhancement, even when students' marks for their laboratory reports were initially hidden. A negative and significant ($p = .010$) linear trend between relative test scores and test date suggests that even modest improvements in subject knowledge were lost over time. Despite this, students could accurately guess their mark based on emailed feedback alone, estimated and awarded marks being linearly correlated ($p < .001$). It is concluded that statement banks organised according to published assessment criteria can ultimately help students to appreciate how their work was graded. However, students should be encouraged to produce a structured response to received feedback if self-assessment is to occur.

Keywords: assessment, feedback, technology, self-regulation

Introduction

One of the reasons why assessment is undertaken is to facilitate student learning, a process supported when tutors supply high quality feedback to their students (Orsmond, Merry, and Reiling 2000). Electronic marking programs that can help tutors to prepare and return online feedback have been available for a number of years (Denton 2001). A review of available technologies by Heinrich, Milne, and Moore (2009) notes a strong positive impact on staff from using tools of this type. Of particular interest are systems that allow assessors to return feedback by allocating comments from a statement bank. It has been found that students prefer this type of feedback over traditional, 'red pen'

annotations on their work (Denton et al. 2008). Using a statement bank prepared in advance can expedite marking (Case 2007). A study by McKie Bell, Smailes, and Smith (2006) reports a 30 hour time saving over two statistical assessments involving 120 students. This is of benefit to staff in an era of resource constraints. It is also advantageous to students, given that timeliness is a well-established trait of good feedback (Gibbs and Habeshaw 1993). In a recent large-scale survey by the National Union of Students (2012), however, two-thirds of UK undergraduates reported that they normally waited more than two weeks for feedback to be returned.

Allocating comments from a statement bank can help tutors to return task-focussed feedback. This is of particular value when marking work at the extremes of the marking scale where tutors might otherwise recourse to drafting perfunctory remarks relating to a student's ability or intelligence. A review of assessment in classroom settings found that feedback that was more focussed on the student than the assignment has a negative effect on attitudes and performance (Black and Wiliam 1998). Using statement banks also facilitates the return of detailed comments, including model answers. This is important given that Higgins, Hartley, and Skelton (2001) have found that students develop negative perceptions of feedback if it does not provide sufficient information on how they might address deficiencies. Similarly, students do not find feedback helpful if it is too general or vague (Weaver 2006). In a study by Price et al. (2010), students expressed dissatisfaction over feedback that was ambiguous and that failed to include examples of what was expected. In an experiment conducted by Huxham (2007), tutor feedback was returned in the form of printed model answers or personalised comments that were annotated on submitted work. Students were then found to perform better in subsequent examination questions associated with the model

answer feedback. In a survey conducted during this study, however, students expressed a preference for personalised feedback.

Brown (2001) identifies one of the common weaknesses in assessment as the inconsistencies in both marking and feedback that can occur when grading duties are shared across a team. This possibility can be lessened if all tutors allocate comments from an agreed statement bank. Improved student awareness of the assessment criteria has been reported when online tools are used to couch returned comments in terms of the published assessment criteria (Case 2007). Approaches of this type can help students to understand the scheme that has been used to assess performance. Accordingly, the return of feedback using printed criteria sheets incorporating performance-level definitions is recognised as good practice (Nicol and Macfarlane-Dick 2006).

Of the six broad categories in the UK National Student Survey (NSS), the one relating to both assessment and feedback has continued to exhibit the lowest satisfaction ratings, and this is acknowledged in contemporary articles (Boud and Molloy 2013). Three questions in the current NSS relate to the clarity of assessment criteria, promptness of feedback, and the level of detail in returned comments (by tutors, implicitly). In Nicol and Macfarlane-Dick's (2006) model of good feedback practice, it is argued that of parallel importance are approaches to assessment that facilitate the active engagement of students with feedback and that promote self-regulated learning. Students use external feedback to create their own internal feedback and this ultimately leads to improvements in student achievement. In transmission models of feedback, tutors pass on messages to students that are then used to make improvements. Such approaches have been criticised as they do not consider how students' motivations and beliefs will affect how they will use the tutor's comments, if at all (Nicol and

Macfarlane-Dick 2006). Boud and Molloy (2013) have emphasised that even elementary feedback should be defined by information used, and not information transmitted.

It is widely accepted that curricula that feature repeated high-stakes assessments reinforce notions that feedback is about marks attained, and this can be detrimental to student self-esteem and motivation. Butler (1987), for example, found that students pay comparatively less attention to their feedback when their mark is included. Credit-bearing assessments can still be used to facilitate self-assessment, however. Taras (2001) advocates withholding marks until students have had time to process external feedback. In her approach, students review their feedback against the agreed assessment criteria in the presence of a tutor, identify areas for improvement, and estimate their grade.

Although online marking tools that employ statement banks are being used increasingly within higher education, Nicol and Milligan (2006) point to a paucity of research over the effectiveness of this type of feedback. Such tools can be used to deliver personalised model answer feedback to students who are on their own at a computer. However, it is not clear if this technique would necessarily lead to the same improvements in attainment that have been recorded when printed generic model answer sheets are circulated in class (Huxham 2007). There is also merit in investigating whether the known deficiencies of the transmission approach can be mitigated against. In line with part of the model developed by Taras (2001), students might be encouraged to re-process their feedback by inviting them to estimate their grade, based on their online feedback alone.

Method

An experiment was conducted concerning tutor feedback on 243 laboratory reports submitted by BSc Applied Chemical and Pharmaceutical Science (ACAPS) and MPharm Pharmacy undergraduates. This credit-bearing assignment was set as part of a first year module and required students to present and discuss raw data collected in the laboratory. The six assessment criteria that were to be used by tutors when grading the assignment were published in advance of the deadline, alongside their associated weightings and guidance notes. Students were expected to follow the normal conventions for organising and writing scientific reports and, consequently, all submitted work had the same basic structure. This exercise had features in common with many other undergraduate science laboratory assessments in that students investigated the practical applications of theories that were likely to feature in the final exam.

The six assessment criteria were used to structure a statement bank of comments within the Electronic Feedback freeware (Denton et al. 2008). The range of comments available reflected the full range of marking outcomes. For example, the student's presentation of their references was worth 5% and this aspect had six associated feedback comments, worth from 0 to 5%. Some assessed criteria, e.g. 'Introduction', had only one set of accompanying comments. Other criteria could only be adequately remarked upon by sub-dividing the criterion into assessed aspects, each with its own list of standard comments. For example, the 'References' criterion was separated into two aspects; presentation and range, each being worth 5%. The electronic statement bank was divided in this manner to produce 13 assessed aspects in total.

After submitting printed copies of their reports during a supervised laboratory session, an unheralded ten-question multiple choice test was circulated alongside

participant information. All ten questions directly related to subject material covered in the practical, including the theories that could feature in the final examination. The test was conducted under closed-book and invigilated conditions and 238 students agreed to participate. Each student's MCQ performance was used only to benchmark their knowledge of this subject area for the purposes of the experiment, it being made clear that the test itself was not credit bearing.

Before marking of the submitted laboratory reports commenced, the class were divided into three groups; A, B and C, each being of approximately equal size and containing a representative distribution of students; 26-31 males, 46 females, 9-12 students from each marker, 28-32 ACAPS students, and 44-49 MPharm students. Over 95% of students were in the age range 18-19 and this factor was not used to influence group compositions. Seven assessors were involved in the marking of the reports and each one received a copy of the electronic set of comments.

During marking, student scripts were minimally hand-annotated with remarks and symbols. Detailed feedback comments were allocated from the statement bank by each assessor selecting the most appropriate comment for each of the 13 assessed aspects. After marking was complete, personalised feedback sheets were printed out and attached to the students' submitted work. A transmission approach was used and returned comments referred only to the attributes of the submitted work. They did not, for example, indicate what action was required on the part of the student, apart from emailed feedback for Groups A and B. A message at the end of these electronic feedback reports asked the recipient to reply via email to confirm that it had been read. These two Groups received feedback emails between three and eight days before hard copies were returned in class. For group B only, the student's mark for the work was not

shown and these students were invited to reply with an estimation of their mark based on online feedback alone.

To ensure that no students were disadvantaged by this experiment, all groups had an opportunity to collect marked work in class within the institution's expected turnaround time of three weeks. This was around three months in advance of the final examination and 163 students were present to receive their work. This was the first opportunity that group C had to review the assessor's comments on their assignments and they were given time in class to read through their marked scripts. In the feedback reports, comments selected from the statement bank were presented under headings that corresponded to each of the six published assessment criteria. Returned feedback also included model answers to all ten questions in the benchmarking MCQ test. For example, all returned feedback remarked on the correct units of measurements for acid dissociation constants and an MCQ based on this information was incorporated within the test. A week after receiving hard copies of their feedback, the original MCQ test was repeated under the same conditions and 243 students agreed to undertake this follow-up assessment.

A revised data set of 141 students was prepared by excluding students who did not fully take part in the experiment; those who did not undertake both the benchmarking and follow-up MCQ tests, the group A and B students who did not email to confirm that they had read their online feedback, and those group C students who did not collect their feedback in the taught session one week before the follow-up MCQ test. The number of days between students first reading their feedback and the follow-up MCQ test was recorded and this ranged from seven to eleven days.

Results and discussion

Marks awarded in the laboratory report were comparable for each group, Table 1, supporting the notion that each of the three groups is representative of the whole class. There were significant ($p < .001$) correlations between % marks for the laboratory reports and scores for both the first and second MCQ tests. This suggests that the test questions were an appropriate measure of the same qualities that were used to judge the practical scripts. It is these qualities, including the students' knowledge and understanding of this subject area, that would be expected to be enhanced if learning through feedback occurred.

For groups A and B, 28% of 163 students failed to reply via email to confirm that they had read their online feedback in the period before work was returned in class. Even asking group B students to email an estimate of their mark at the commencement of their feedback did not result in a greater number replying to confirm it had been read. It is not clear whether these students read their feedback only superficially and missed the request for a read receipt, or fully read their feedback and chose not to reply to this atypical request. The number of students collecting printed feedback represents 67% of those who submitted assignments for marking. This reflects the level of attendance at the session where hard copies of the feedback were first attempted to be returned.

(Table 1)

All but one of the 57 group B students who replied to confirm that they had read their feedback also provided an estimation of their performance, Figure 1. Overall, students demonstrated that they were adept at guessing their % mark after reflecting on their online feedback alone, the Pearson correlation coefficient for awarded and estimated % marks being .730 ($p < .001$). This significant linear trend suggests that students broadly understood the mark scheme against which they were being judged.

Those students accurately guessing their mark demonstrated similar evaluative skills to those of the assessors of their work. As Sadler (1989) has noted, this is essential if students are to be able to take action to 'close the gap' between current and good performance.

(Figure 1)

The change in % score from the benchmarking to the follow-up MCQ test ($\Delta\%$) was calculated for each student using the revised data set, average results for each group showing a decline in test performance, Table 2. No significant differences were found in two sample *t*-tests when $\Delta\%$ was compared between genders ($p = .11$) or between programmes ($p = .94$). A one-way analysis of variance for $\Delta\%$ against marker identity ($p = .21$) again detected no statistically significant differences. It is apparent from Table 2 that values of $\Delta\%$ for group C were clustered around a higher mean value than for groups A and B. However, a one-way analysis of variance showed no significant differences in $\Delta\%$ between the various student groups ($p = .32$).

(Table 2)

In the revised data set, 72 students performed less well in the follow-up MCQ test, 27 students achieved the same mark, and only 42 students improved their score. It appears that even those who read their model answer feedback did not make good use of it. This notion of superficial engagement with the returned comments is further supported by a scatter plot of $\Delta\%$ against the number of days between feedback being read and the follow-up test, Figure 2. A negative linear correlation between these two variables is the only significant relationship ($p = .010$) found within the revised data set. The gradient (\pm SE) of this correlation indicates a point estimate decrease of $11\pm 4\%$ in

MCQ test score over the five day interval examined. A significant negative correlation ($p = .025$) is also found when only data for students receiving emailed feedback (groups A and B) are analysed, representing a point estimate test score decline of $16 \pm 7\%$ over the same five day interval.

(Figure 2)

The correlation in Figure 2 suggests that any enhancements in subject knowledge from received information were transitory, a trend that has been recorded previously. For example, Pashler et al. (2005) presented factual information to 258 students and their recall was tested later on the same day and one week afterwards. A decline in test performance was observed and it noteworthy that their investigation also used a transmission approach to provide information.

Conclusions

The transmission type feedback returned in this study apparently failed to enhance students' subject knowledge. This finding further challenges both the effectiveness of the 'feedback as telling' approach and the assumptions on which it is based, recently summarised by Boud and Molloy (2013). Around one-quarter of emailed students failed to respond to a request for a feedback read receipt. This asks questions over the extent to which students actually read returned comments and further study of this area would be worthwhile. Although there was no apparent educational benefit from withholding marks, this does not discredit Taras's (2001) model for promoting self-assessment. Rather, it emphasises the importance of the other features of her approach, including the tutor-facilitated sessions where students review their feedback.

A deficiency of transmission models of feedback is that they do not naturally encourage assessor and peer dialogue around learning, one of the seven principles of good feedback practice (Nicol and McFarlane-Dick 2006). Indeed, any student in this study receiving online feedback on their own at a computer has no immediate prospect of discussions with their tutor or classmates. According to Carless (2006), students do want to learn from their feedback but are often unclear as to how to go about this. The digital delivery and presentation of feedback offers opportunities to furnish students with electronic tools to guide how they re-process tutor remarks and to facilitate dialogues. Kerrigan et al. (2009) have already undertaken some work in this respect and it is an area that warrants further investigation and development. In the absence of such support, students may not respond to written feedback in a manner that is consistent with what the tutor intended, as appears to be the case in this investigation and in other contemporary studies of feedback (Crisp 2007).

The model answer feedback returned in this experiment was delivered in a format that had previously been found to be highly rated by students for features that are emblematic of high quality feedback (Denton et al. 2008). This investigation therefore provides another example of where student perceptions of feedback and its educational effectiveness are not necessarily the same, echoing a finding of Huxham's (2007) study. The outcomes of this experiment also reinforce Nicol and Milligan's (2006) assertion regarding the need to investigate the effectiveness of technology-enhanced feedback, even when notional best-practices are observed and detailed comments from tutors are returned in a timely manner.

Although some advantages in using statement banks and returning model answer feedback have been identified, the outcomes of this experiment suggest that care should

be exercised how they are used. The number of comments returned to students in this study, for example, exceeds Lunsford's (1997) recommendation of around three substantive remarks for each piece of work. A potential development would be to design electronic marking tools that provide some means to highlight those remarks that students should pay most attention to. In this way, the feedback could then emphasise the main themes that the student should address in future work.

Self-regulated learning is fostered when there is a clarity of purpose for the assessment that is shared between both tutors and students (Price et al. 2010). The practical exercise in this study has since been revised and it links to a subsequent assessment, which were always there, have now been made much clearer. It is now a formative assignment set in preparation for an open-book examination of a series of practical activities. This provides the opportunity to gauge learning from external feedback as it invites students to use returned comments to produce another piece of identified work to a higher standard (Boud 2000). The value of returning statement bank comments electronically to facilitate learning merits further study. It would be instructive to investigate, for example, whether this approach has any positive impact on subsequent student performances in a linked assessment, thereby completing the feedback loop (Sadler 1989).

Many students in this study were able to accurately estimate their % marks by reading through online comments from the assessor. It is suggested that this occurred because the marking software employed allowed both feedback to be couched, and % marks to be calculated, in terms of the published assessment criteria. The students' apparent awareness of these criteria endorses the findings of Case (2007) when statement banks are used to return feedback. According to Cross (1996), assessment

without feedback is like archery practice in the dark. Even with the lights switched on, however, students require an understanding of the target that they are aiming for.

References

Black, P., and D. Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy and Practice* 5(1): 7-74.

Boud, D., and E. Molloy. 2013. "Rethinking models of feedback for learning: The challenge of design." *Assessment and Evaluation in Higher Education* 38(6): 698-712.

Boud, D. 2000. "Sustainable assessment: Rethinking assessment for the learning society." *Studies in Continuing Education* 22(2): 151–167.

Brown, G. 2001. *Assessment: A guide for lecturers*, York: LTSN generic centre.

Butler, R. 1987. "Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest and performance." *Journal of Educational Psychology* 78(4): 210–216.

Carless, D. 2006. "Differing perceptions in the feedback process." *Studies in Higher Education* 31(2): 219–33.

Case, S. 2007. "Reconfiguring and realigning the assessment feedback processes for an undergraduate criminology degree." *Assessment and Evaluation in Higher Education* 32(3): 285-299.

Crisp, B. R. 2007. "Is it worth the effort? How feedback influences students' subsequent submission of assessable work." *Assessment and Evaluation in Higher Education* 32(5): 571–81.

Cross, K. 1996. "Improving teaching and learning through classroom assessment and classroom research." In *Improving student learning: Using research to improve student learning*, edited by G. Gibbs. Oxford: Oxford Centre for Staff Development.

Denton, P. 2001. "Generating coursework feedback for large groups of students using MS Excel and MS Word." *University Chemistry Education* 5: 1-8.

Denton, P., J. Madden, M. Roberts, and P. Rowe, 2008. "Students' response to traditional and computer-assisted formative feedback: a comparative case study." *British Journal of Educational Technology* 39: 486–500.

Gibbs, G., and T. Habeshaw, 1993. *Preparing to teach: An introduction to effective teaching in higher education*. Bristol: Technical and Educational Services.

Heinrich, E., J. Milne, and M. Moore. 2009. "An Investigation into E-Tool Use for Formative Assignment Assessment – Status and Recommendations." *Educational Technology and Society* 12(4): 176–192.

Higgins, R., P. Hartley, and A. Skelton. 2001. "Getting the message across: The problem of communicating assessment feedback." *Teaching in Higher Education* 6(2): 269–274.

Huxham, M. 2007. "Fast and effective feedback: Are model answers the answer?" *Assessment and Evaluation in Higher Education* 32: 601–611.

Kerrigan, M. J. P., M. Clements, A. Bond A and G. Saunders. 2009. "eReflect - Making assessment count." Paper presented at the Fourth International Blended Learning Conference, Hatfield, June 17-18.

Lunsford, R. 1997. "When less is more: principles for responding in the disciplines." In *Writing to learn: strategies for assigning and responding to writing across the disciplines*, edited by M. Sorcinelli and P. Elbow. San Francisco: Jossey-Bass.

McKie Bell, F., J. Smailes, and L. Smith. 2006. *Providing Electronic Feedback: Evaluation of Electronic Feedback 100 Software*. Newcastle: Northumbria University.

National Union of Students. 2012. *Student Experience Research 2012 Part 1: Teaching and Learning*. Gloucester: Quality Assurance Agency.

Nicol, D. J., and D. Macfarlane-Dick. 2006. "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice." *Studies in Higher Education* 31(2): 199-218.

Nicol, D. J., and C. Milligan. 2006. "Rethinking technology-supported assessment in terms of the seven principles of good feedback practice." In *Innovative Assessment in Higher Education*, edited by C. Bryan and K. Clegg. London: Taylor and Francis

Orsmond, P., S. Merry and K. Reiling. 2000. "The use of student derived marking criteria in peer and self-assessment." *Assessment and Evaluation in Higher Education* 25(1): 21-38.

Pashler, H., N. J. Cepeda, J. T. Wixted, and D. Rohrer. 2005. "When does feedback facilitate learning of words?" *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31: 3-8.

Price, M., K. Handley, J. Millar, and B. O'Donovan. 2010. "Feedback: all that effort, but what is the effect?" *Assessment and Evaluation in Higher Education* 35(3): 277-289.

Sadler, D. R. 1989. "Formative assessment and the design of instructional systems." *Instructional Science* 18: 119–144.

Taras, M. 2001. "The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutors." *Assessment and Evaluation in Higher Education* 26(6): 605–614.

Weaver, M. 2006. "Do students value feedback? Students' perception of tutors' written responses." *Assessment and Evaluation in Higher Education* 31: 379–394.

Table 1. Profile of all the participants in this experiment ($n = 243$).

Group	n	Mean % mark for laboratory report	Standard deviation of % marks	n (confirmed reading of online feedback)	n (collecting hard copy feedback)
A	81	61.0	13.0	60	52
B	82	62.3	11.4	57	57
C	80	62.6	12.7	n/a	54

Table 2. Profile of participants in the revised data set ($n = 141$).

Group	n	Gender		Programme		Mean, change in % MCQ test score ($\Delta\%$) ^a
		Male	Female	ACAPS	MPharm	
A	43	15	28	14	29	-5.3
B	45	21	24	13	32	-6.4
C	53	20	33	21	32	-1.7

^aCalculated for each student by subtracting the benchmark MCQ score from the follow-up MCQ test score.

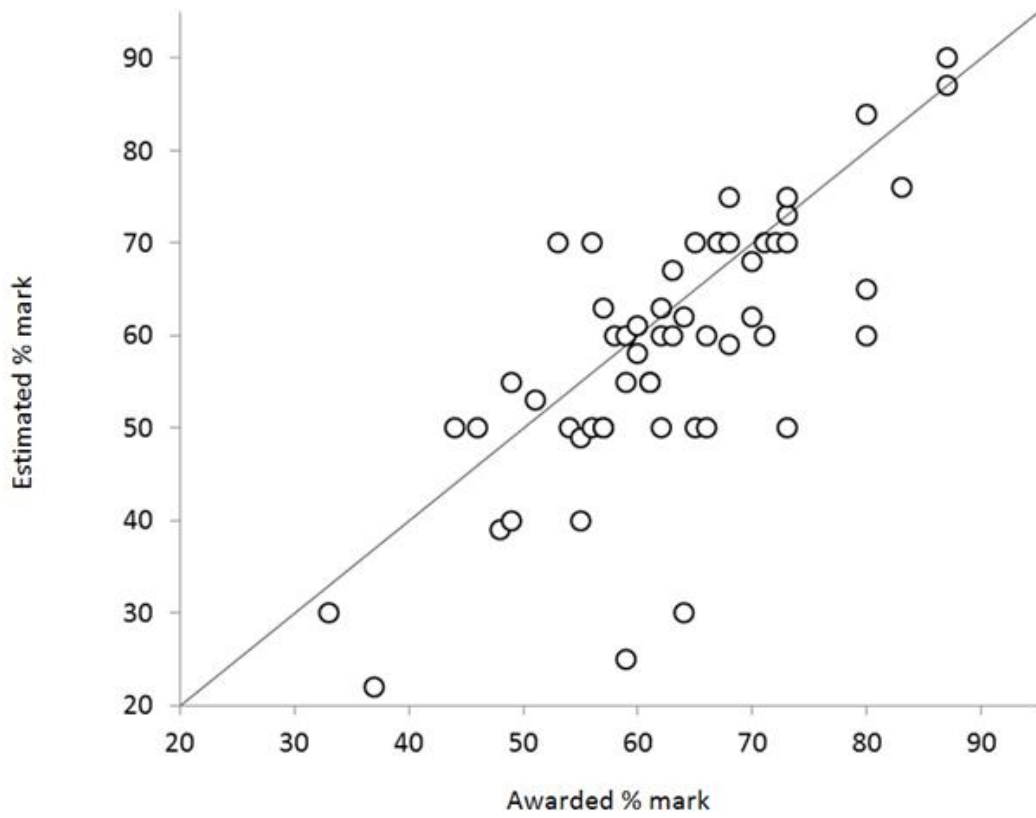


Figure 1. Correlation of estimated and actual laboratory report % marks for 56 group B students. The solid line represents $y = x$.

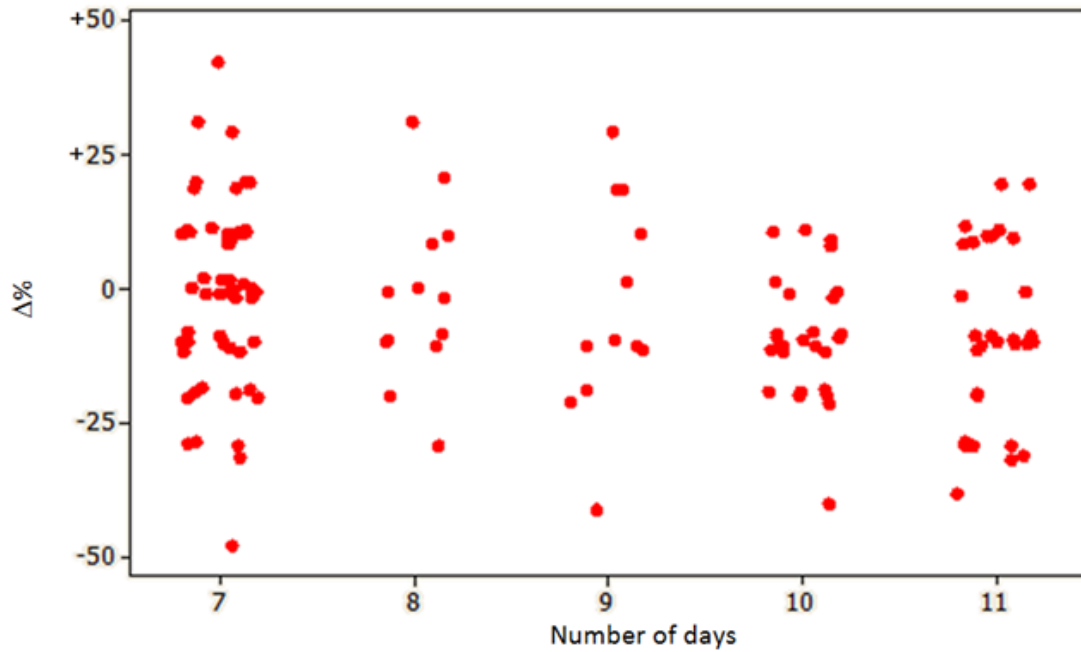


Figure 2. Change in MCQ test score versus number of days between reading feedback and undertaking the follow-up MCQ test. To prevent points overlaying each other, jitter was added to all X and Y values, being a random value between ± 0.3 days and $\pm 2\%$, respectively.