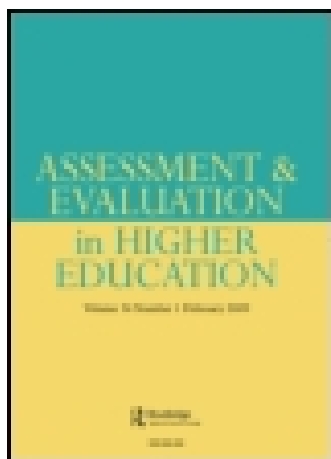


This article was downloaded by: [Swansea University], [Phil Newton]

On: 25 September 2014, At: 05:36

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Assessment & Evaluation in Higher Education

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/caeh20>

Use of the 'Stop, Start, Continue' method is associated with the production of constructive qualitative feedback by students in higher education

Alice Hoon^a, Emily Oliver^b, Kasia Szpakowska^c & Philip Newton^a

^a College of Medicine, Swansea University, Swansea, UK

^b School of Applied Social Sciences, Durham University, Durham, UK

^c College of Arts and Humanities, Swansea University, Swansea, UK

Published online: 17 Sep 2014.

To cite this article: Alice Hoon, Emily Oliver, Kasia Szpakowska & Philip Newton (2014): Use of the 'Stop, Start, Continue' method is associated with the production of constructive qualitative feedback by students in higher education, *Assessment & Evaluation in Higher Education*, DOI: [10.1080/02602938.2014.956282](https://doi.org/10.1080/02602938.2014.956282)

To link to this article: <http://dx.doi.org/10.1080/02602938.2014.956282>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Use of the ‘Stop, Start, Continue’ method is associated with the production of constructive qualitative feedback by students in higher education

Alice Hoon^a, Emily Oliver^b, Kasia Szpakowska^c and Philip Newton^{a*}

^aCollege of Medicine, Swansea University, Swansea, UK; ^bSchool of Applied Social Sciences, Durham University, Durham, UK; ^cCollege of Arts and Humanities, Swansea University, Swansea, UK

Students in higher education are increasingly asked to give feedback on their education experience, reflecting an increase in the importance attached to that feedback. Existing literature demonstrates that qualitative student feedback is valued and important, yet there has been limited evaluation of the means by which qualitative student feedback is collected, and how collection method influences the depth of feedback from students. We evaluated the depth of feedback written by students on programmes at three different universities in the United Kingdom, using an established evaluation instrument. We found that the use of a structured pro forma (Stop, Start, Continue) was associated with feedback of greater depth than that produced using free text entry. We then evaluated the effect of switching from a free text to a structured pro forma at one institution and found, again, that the structured pro forma was associated with feedback of greater depth. In addition, students indicated a preference for the structured pro forma.

Keywords: student feedback; Stop Start Continue; pro forma; student satisfaction

Introduction

Feedback is a fundamentally important component of education. It is a two-way process, with feedback provided both to and from educators and students. The feedback provided by educators to students is normally provided with a single, simple aim – to help students learn. This feedback has been extensively studied (see Shute 2008 for a recent review), and the accumulated literature has informed best practice guidelines for the production and use of feedback written to students (e.g. see Marriott and Teoh 2014; Williams and Brennan 2004).

In contrast, the feedback provided by students to educators is collected for multiple reasons and for multiple audiences. Educators use student-generated feedback to help them improve and develop their teaching as well as to understand the current educational attainment of their students. University managers and administrators use student feedback for quality assurance and human resources purposes, while prospective students use feedback from previous students to help them make decisions about their education (Huxham et al. 2008; Richardson 2005). These three pathways for the collection and use of student feedback clearly have different needs and there is currently a lack of clear guidance on policy and best practice for the collection of

*Corresponding author. Email: p.newton@swansea.ac.uk

the most useful feedback from students to assist these different agendas. In addition, despite the existence of relatively well-established methods, there is often a need for the development of new tools which address local contexts and needs (e.g. Gruber et al. 2010).

The impact of student-generated feedback has gradually escalated, despite this lack of evidence to inform policy and practice. Recent developments in the United Kingdom include the National Student Survey (NSS), used to collect data (largely quantitative) about the student experience and used in the evaluation of higher education institutions (Buckley 2012). Similarly, student evaluation of teaching (SET) forms a significant part of overall performance evaluations of teachers in the United States of America. Although this feedback is used largely for the quality assurance purposes described above, there is good evidence that one effect of SET is to improve future teaching (Marsh and Roche 1997), although many controversies still exist about the influence of confounding variables, including predicted/expected grades and the perceived difficulty of the course (Basow and Martin 2012).

The NSS and SET use largely quantitative measures of recording student feedback, and, although these can provide important information to all stakeholder groups, for course development and teaching improvement, there is a need to collect and analyse qualitative feedback from students (Grebennikov and Shah 2013). Qualitative feedback adds context and detail to issues which arise in the quantitative data (Symons 2006), while evaluative studies have demonstrated that students value providing free text comments in feedback (Kindred and Mohammed 2005), and the majority will leave qualitative comments when given the opportunity to add them to an otherwise quantitative SET form (Brockx, Van Roy, and Mortelmans 2012). In order to focus on course development, qualitative feedback needs to be focused on that end, rather than, for example, the rating of individual teachers (Edström 2008).

However, the resources required to conduct a meaningful analysis of qualitative data are much greater than for quantitative feedback, and these resource issues are often a deterrent to the collection of open-ended qualitative feedback (Richardson 2005). An attempt to automate the process of coding student qualitative feedback concluded that manual coding was 'arguably superior' (Palmer and Campbell 2013). The heavy resource requirement is exacerbated by a well-established principle that, in general, face-to-face administered paper-based surveys to capture student feedback achieve a higher response rate than online surveys (Nulty 2008). Hence, there is a need to ensure that the collection of qualitative data for course development is maximally efficient, not directly replicating qualitative feedback while retaining opportunities for full expression and allowing for the production of feedback that is deep and specifically aimed at course development. Addressing this issue was the central aim of the current study: to provide practical, evidence-based guidance for those seeking to optimally collect and use student feedback.

We analysed the qualitative feedback written by students on their learning experience at three programmes at universities in the UK with pre-existing variation in their feedback collection methodologies. Two programmes used a structured format, similar to the 'Stop, Start, Continue' model. This is a well-established method of collecting feedback in many disciplines, including education (e.g. see George and Cowan 1999, 28). It asks structured questions that collect student views on what educators should no longer do (stop), should consider adding (start) and features that should be retained (continue). The third programme collected feedback using a blank, free text only pro forma.

To analyse feedback, we adapted a tool that has been used in similar studies on instructor-generated feedback, and is used to simply code individual feedback statements according to their depth, regardless of their content (Brown and Glover 2006; Newton, Wallace, and McKimm 2012).

In the current paper, we present findings that, when comparing between programmes, the use of a structured, questioning pro forma was associated with the production of deeper feedback by students when compared with a blank pro forma. Similarly, a switch from a blank to a structured pro forma by one programme was associated with an increase in the depth of feedback produced, and a positive evaluation of the new form by students.

Method

We conducted two studies. The first compared feedback from three different courses, each of which utilises a distinct pro forma to collect feedback from students. The second study then evaluated the impact of changing the structure of the feedback pro forma for one of those programmes. The structure of the new form was informed by and based upon the results of the first study.

Study 1

Written student feedback from three different courses at two different institutions in the United Kingdom was evaluated. This module-level feedback had been collected as part of the routine evaluation of each programme.

Course A is a graduate-entry medical degree, an undergraduate course undertaken by students who already have an undergraduate degree. Each intake has about 70 students. Teaching in the first two years of the course (on which the current study is focused) is arranged around discrete teaching weeks. Students provide qualitative feedback weekly on each university-based teaching week, using a pro forma that, at the beginning of Study 1, had a free text entry box for every teaching session occurring that week.

Course B is an undergraduate Bachelor of Science degree, with an annual intake of between 80 and 100 students. Teaching occurs in modules taught over 12 or 24 weeks. Students provide qualitative feedback on their course at the end of each module, and at the end of each academic year. Feedback analysed from Course B was from 12 different modules, collected at the end of the module using the 'Stop, Start, Continue' format. This format consists of presenting three text boxes containing the words Stop, Start and Continue. In the Stop box, students are given the opportunity to make suggestions about what they do not find useful or helpful, for example, 'stop speaking so quickly'. In the Start box, students have the opportunity to make suggestions or give ideas about what they think could be started which may aid their learning, for example, 'start giving out class handouts'. The Continue box allows students to comment on best practice and aspects of teaching that they find useful, for example, 'continue the use of class quizzes'.

Course C is a level 2 undergraduate module in the Arts and Humanities, taken by students who have already completed a module in the same subject area. Each intake has 25–40 students. Teaching occurs in the module over 11 weeks in the first semester. Students provide qualitative feedback at the end of each module. Feedback for Course C is collected at the end of each module. Students are given a feedback

form containing text box for each of the following questions: (1) What was the best thing about the module? (2) What would you change about the module? (3) Any other comments? Feedback was analysed from four modules.

Coding of feedback

Qualitative feedback was coded using a feedback system similar to those developed by Newton, Wallace, and McKimm (2012) and Brown and Glover (2006). Every statement in the feedback was coded according to two criteria – the first being whether its overall theme was that the issue being commented on was ‘positive’, ‘negative’ or ‘adequate’ (coded as P, N or A, respectively), and the second being the depth of the feedback. For example, if a statement merely indicated at the most basic level whether something was good or not good, it was coded as 1 (descriptive); if a statement indicated why something was or was not good, it was coded as 2 (qualified); and if a statement included constructive suggestions for change or development, or if there was a clear and simple inference for how this could be achieved (e.g. ‘the lecturer spoke too fast’), then it was categorised as 3 (constructive). Statements made by a student that did not relate to the content of the lecture/module in question, for example, reporting that they did not attend a lecture, were classed as ‘other’. See Table 1 for a summary and some examples from the study. Although this instrument was originally developed for the evaluation of feedback written to students by instructors, the feedback under study here was that collected to inform course development on the basis of student views on current teaching effectiveness (Berk 2012). The differences between category 1, 2 and 3 statements were considered highly relevant to achieving that end; a constructive category 3 statement (e.g. ‘the teaching was poor because the lecturer used too many abbreviations’) is

Table 1. Feedback categories used for coding with representative examples found in the study.

| Feedback category | Basic theme of statement | Examples from study |
|-------------------|---|---|
| P1 | This was good | ‘Excellent’, ‘Very good’, ‘Good’, ‘Great as always’, ‘Great thanks’, ‘Amazing’ |
| P2 | This was good because ... | ‘Like how Dr ... repeats key points throughout the lecture’, ‘Good at signposting’, ‘Well paced’ |
| P3 | This was good because ... but would be even better if ... | ‘a lot of detail perhaps a bit too much’, ‘good but could have been split into two lectures’ |
| N1 | This was bad ... | ‘awful’, ‘Hard’ |
| N2 | This was bad because ... | ‘... was very intimidating’, ‘seemed nervous’, ‘boring topic’ |
| N3 | This was bad because ... and would be better if ... | ‘The essay subject was too broad, ... I’d suggest to give a number of possibilities’ ‘The lecture was not pitched to our current level of understanding’, ‘spoke too quickly’ |
| A1 | This was adequate ... | ‘OK’, ‘Adequate’, ‘Reasonable’ |
| A2 | This was adequate because ... | ‘OK, not very detailed’, ‘Adequate ... found it difficult to understand what was being said’, ‘OK, not sure what the aim ... was’ |
| A3 | This was adequate because ... and would be better if ... | ‘OK, it would have been good to be in smaller groups’ |

clearly more useful for course development than the relevant descriptive category 1 statement (the teaching was poor).

Although all feedback was labelled as positive, negative or adequate, the authors were most interested in the depth of the feedback, that is, how useful or constructive each statement was, rather than how well the lecture or module had been received by the students. To ensure consistency, all coding was performed by one author (initials removed for peer review) after training by another (initials removed for peer review).

Study 1 – results and discussion

Students in Course A made a total of 1072 statements, students in Course B made a total of 264 statements and students in Course C made 131 statements. The detailed coding results for each depth category are shown by percentage in Table 2, with a representation of the feedback collapsed by category shown in Figure 1. Percentages were used to factor out any effect of cohort size.

The data represented in Table 2 shows that the feedback collected from Course A, using a blank pro forma, were largely depth category 1 and 2 responses, whereas very few category 1 responses were made by respondents in Course B and C which used more structured approaches. In contrast, the feedback collected from Course B was largely in the most constructive category 3. Thus, although more feedback had been collected from Course A, the depth of that feedback was much lower than in the two other courses, and the descriptive nature of the comments were of a nature that would be of limited use for course development, and could have been simply captured using a quantitative method.

As there was much variation across the three courses in the percentage of statements attributed to each feedback category, a further analysis was conducted in which the percentage of statements for each question type was analysed. This was conducted on the feedback collected for Course B and Course C as the feedback format asked specific questions. This analysis could not be conducted for Course A as students were only asked to comment on each teaching session and were not asked specific questions. Table 3 shows that asking students what they would like to stop occurring and what they would like to start occurring both tended to result in N3-type responses. Asking students what they think should be continued typically resulted in P3-type responses. For Course C, it was found that asking students ‘what did you like about this module’ typically resulted in P2 responses, whereas ‘what would you change about this module’ generally resulted in N3 responses.

Table 2. Percentage of statements in each feedback category by course.

| | Depth | | | | | | | | | Other |
|--------------------------------|-----------------|-----|-----|---------------|-----|-----|------------------|------|-----|-------|
| | 1 (Descriptive) | | | 2 (Qualified) | | | 3 (Constructive) | | | |
| | P | N | A | P | N | A | P | N | A | |
| Course A (Blank) | 42.2 | 0.4 | 2.2 | 33.5 | 4.2 | 2.1 | 4.3 | 9.1 | 0.8 | 1.4 |
| Course B (Stop/Start/Continue) | 2.7 | 2.3 | – | 3.4 | 0.1 | 0.8 | 35.6 | 54.6 | – | – |
| Course C (Best/Change/Other) | 2.3 | – | – | 61.8 | 1.5 | – | 1.5 | 29.8 | – | 3.1 |
| Mean | 15.7 | 0.9 | 0.7 | 32.9 | 1.9 | 0.9 | 13.8 | 31.1 | 0.3 | 1.5 |
| SD | 22.9 | 1.2 | 1.2 | 29.2 | 2.1 | 1.0 | 18.9 | 22.8 | 0.5 | 1.5 |

Downloaded by [Swansea University], [Phil Newton] at 05:36 25 September 2014

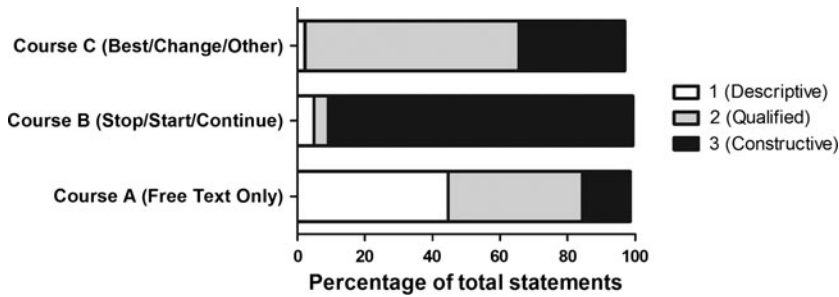


Figure 1. Percentage of statements in each feedback category by course and the type of pro forma used to collect the feedback.

Table 3. Percentage of statements in each feedback category by question type.

| | | Depth | | | | | | | | | Other |
|----------|-------------------------|-----------------|----|----|---------------|-----|-----|------------------|------|----|-------|
| | | 1 (Descriptive) | | | 2 (Qualified) | | | 3 (Constructive) | | | |
| | | P1 | N1 | A1 | P2 | N2 | A2 | P3 | N3 | A3 | |
| Course B | Stop | 0.7 | – | – | – | 0.7 | – | – | 20.9 | – | – |
| | Start | – | – | – | – | – | – | – | 36.2 | – | – |
| | Continue | 1.5 | – | – | 3.7 | – | 0.8 | 34.7 | 1.5 | – | – |
| Course C | Best thing ¹ | 1.6 | – | – | 48.8 | – | – | – | – | – | – |
| | Change ² | 0.8 | – | – | 1.6 | 1.6 | – | 1.6 | 29.5 | – | – |
| | Comments ³ | – | – | – | 11.6 | – | – | – | 0.8 | – | 2.3 |
| | Mean | 0.8 | – | – | 11.0 | 0.4 | 0.8 | 7.3 | 14.8 | – | 0.4 |
| | SD | 0.7 | – | – | 19.1 | 0.6 | – | 15.4 | 16.1 | – | 1.0 |

¹What was the best thing about this module?

²What would you change about this module?

³Any other comments?

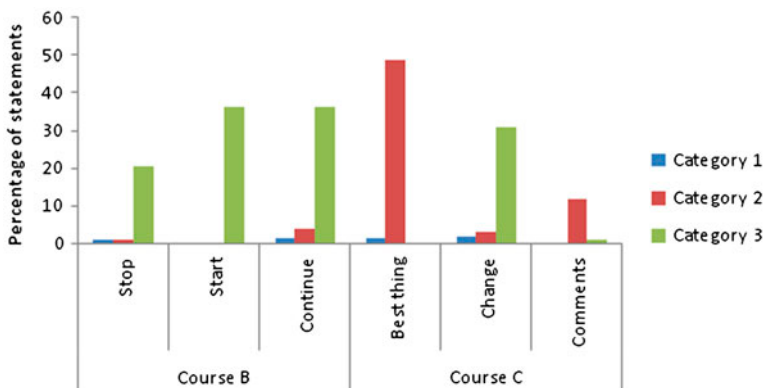


Figure 2. Percentage of statements in each feedback category by question type.

Figure 2 indicates that for Course B, the Stop/Start/Continue format resulted in students giving feedback which made constructive suggestions for change. For Course C, asking students what they liked best about the course typically resulted in

responses regarding what was good about the course and why, whereas specifically asking what students would change about the course resulted in students making category 3 suggestions for improvements. When students were asked if they had any other comments, the responses given tended to be category 2 statements which, although useful and qualified, did not offer constructive suggestions for change or development.

Taken together, these findings suggest that the specificity of questions asked in evaluation forms may influence the constructiveness of the feedback obtained. Using only open-ended questions, such as the format used on Course A, was associated with feedback which was mostly descriptive, and may, therefore, be limited in terms of enabling course development. However, when questions were phrased to specifically ask what aspects of teaching should be changed or continued as per Course B, much more of the feedback collected was of a type which made suggestions that could potentially inform constructive change in teaching practice.

One interpretation of these findings may be that more constructive qualitative feedback can be obtained when using a structured form. However, this interpretation is subject to many confounds – the feedback was collected from different students on three different courses, in different disciplines, at two different institutions, with students at different stages (undergraduate vs. postgraduate) and at different frequencies (weekly vs. at the end of a semester). Thus, it is difficult to say with any certainty whether the differences in the quality of the feedback were a direct function of the feedback format or due to another variable. To understand this further, we designed a second study.

Study 2 – introduction

In this study, the qualitative feedback form for Course A was changed from the open-ended format used previously to a modified Stop/Start/Continue format. The curriculum for Course A is arranged around discrete teaching weeks. We tested the modified form over four teaching weeks across two cohorts of Course A, by simply replacing the old format (as used in Study 1, and which had been used for every teaching week completed by those cohorts previously) with the new format. Cohort size was the same as in Study 1 (70), and feedback collected was then evaluated as in Study 1. Given the evidence from Study 1 that the open-ended form was associated with poorer quality feedback, concurrent use of the open-ended form in ‘control’ courses or institutions was not appropriate. We, therefore, compared the feedback generated on the new form with the feedback that had been completed for the same teaching weeks in the previous academic year, to assess whether changing the format of feedback was associated with better quality feedback.

Study 2 – method

At the end of each teaching week, students were presented with a feedback form which asked the following questions:

- (1) Stop: What would you like us to stop doing?
- (2) Start: What suggestions do you have?
- (3) Continue: What is being done well that you would like to continue?
- (4) Good practice/highlights?

- (5) Any other comments?
- (6) Any comments about the new feedback form.

Question 4 was included as much of the teaching on Course A was given by external speakers who may value course feedback as evidence of their own professional development and engagement with education. Question 5 was included to allow students to make more general comments relating to the course that may not have fitted within the Stop/Start/Continue structure, in keeping with literature suggesting that open-ended questions are valued by students and may result in important feedback (Grebennikov and Shah 2013; Kabanoff, Richardson, and Brown 2003).

The Stop/Start/Continue form was distributed at the end of each of four teaching weeks to the first and second year cohorts; therefore, representing eight teaching weeks in total. All feedback was analysed and coded as in Study 1. Feedback from the same teaching weeks in the previous academic year of Course A was obtained and coded in the same way.

Study 2 – results and discussion

In the original free text format, the Year 1 students made a total of 872 statements and the Year 2 students made a total of 611 statements. When presented with the Start/Stop/Continue format, the Year 1 students made a total of 313 statements, whereas Year 2 made a total of 204 statements. Results from the coding of these statements are shown in Table 4, which shows that, in agreement with the findings from Study 1, the original unstructured feedback format yielded a high number of category 1 responses, with 44.5% of statements falling within the P1 category.

We again collapsed our analysis into depth categories only. For the Stop/Start/Continue format, however, 50% of statements were classed as category 3 (see Figure 3). With regard to the number of statements classed as category 2, there was little change between the original format and the Start/Stop/Continue format.

Analysis by question

To identify whether a particular feedback question was more likely to result in a particular type of statement, an analysis by question type was conducted. Table 5 shows that ‘Stop’ almost always resulted in N3-type statements, ‘Start’ typically resulted in N3 or P3 statements and ‘Continue’ tended to result in P3 statements. The majority of P2 responses were written in the good practice section. There were no N1 or category A statements.

Table 4. Percentage of statements in each feedback category in the original free text format and the Stop/Start/Continue format.

| | Depth | | | | | | | | | Other |
|---------------------|-----------------|------|------|---------------|-------|------|------------------|-------|---|-------|
| | 1 (Descriptive) | | | 2 (Qualified) | | | 3 (Constructive) | | | |
| | P | N | A | P | N | A | P | N | A | |
| Free text only | 44.52 | 1.31 | 2.62 | 23.73 | 15.55 | 0.98 | 3.27 | 6.06 | 0 | 1.96 |
| Stop/Start/Continue | 3.92 | 0 | 0 | 28.43 | 12.75 | 0 | 25.98 | 24.02 | 0 | 4.9 |

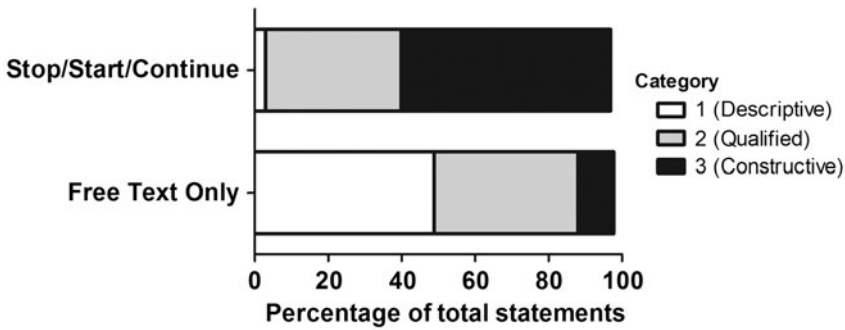


Figure 3. Percentage of statements in each feedback category in the original free text feedback format and the Stop/Start/Continue format.

Table 5. Percentage of statements for each feedback category per question. No ‘A’ category statements were found.

| | P1 | P2 | P3 | N1 | N2 | N3 | Other | Total |
|---------------|-----|------|------|----|-----|------|-------|-------|
| Stop | – | 0.2 | 0.2 | – | 1.9 | 15.4 | 0.9 | 18.6 |
| Start | 0.2 | 3.2 | 6.2 | – | 3.4 | 10.9 | – | 23.9 |
| Continue | 0.9 | 6.8 | 15.2 | – | 0.2 | 0.4 | – | 23.5 |
| Good practice | 1.3 | 15.4 | 7.7 | – | 1.1 | 0.6 | 0.2 | 26.3 |
| Comments | 0.9 | 2.6 | 0.2 | – | 1.1 | 2.1 | 0.6 | 7.5 |
| Total | 3.2 | 28.1 | 29.4 | – | 7.7 | 29.4 | 1.7 | |

Student evaluation of the new form design

Students were asked to comment on the new Stop/Start/Continue design, using an open-ended question. The depth and volume of feedback obtained in response to this question were limited, and the only theme emerging was a preference (or not) for the new form. Of the 31 comments, received, 20 were positive and 11 were negative, meaning that 65% of student comments indicated a preference for the new form. Within the positive theme, individual comments indicated that students found the new form ‘better’, ‘more constructive’ and observed that it allowed them to give feedback on the week as a whole.

General discussion

Study 1 reported that the format of a feedback form may have implications for the depth of qualitative feedback that is obtained. Specifically, the use of a form which gave the opportunity to provide open, free text responses was associated with large quantities of descriptive feedback that did not necessarily allow for course development, or even for educators to identify specific elements of teaching provision that were well received by students. The Stop/Start/Continue format used in Course B, which specifically asked the responder to make suggestions for change, was associated with less feedback in terms of quantity of statements, but with highly constructive feedback in those statements given. A similar pattern was found for Course C, whereby asking students what they would change about a course was more likely to

result in constructive comments, whereas asking what they liked most about a course tended to result in more descriptive statements about the content of the course.

The findings of Study 1 suggest that the structure of a feedback form might have a significant impact upon the type of qualitative feedback that is collected, and that the specific questions asked may influence the responses given. However, a key limitation of Study 1 was that it compared the feedback of three different groups of students on three very different courses, at two different institutions, with feedback collected at different times and with different frequencies.

Study 2 addressed this limitation using a cohort design with one module, demonstrating that the Stop/Start/Continue format increased use of constructive statements relative to the open-ended format. Additionally, Study 2 found that, whilst students appeared to make fewer statements using the Start/Stop/Continue format, the quality of feedback improved. This is an important finding as there is an ever-increasing expectation on students to provide regular course feedback, and qualitative data are a rich source of important information that may aid course development (Grebennikov and Shah 2013; Symons 2006). We must express caution, however, that obtained data are likely to be useful only to the extent to which users have appropriate time and reflective capacity to consider them (Richardson 2005). Thus, it is important to identify ways in which the richness of qualitative feedback can be retained (and even enhanced) while reducing the resource load for analysis. Lastly, it was apparent that students, as a group, preferred the Stop/Start/Continue format. Having students engaged with providing feedback is a fundamentally important part of obtaining successful feedback (Buckley 2012; Richardson 2005), and thus the preferences expressed here for the Stop/Start/Continue format further support its use.

There are some additional limitations which should be considered in the interpretation of the results presented here. The tool we have used to code the feedback assigned categories based upon a simple depth criterion. Obviously, such an approach will not account for all the complexity of qualitative feedback collected from students. For example, one N2-type statement which points out a significant limitation (e.g. 'the content of lecture X was factually inaccurate) may be of much greater importance than an N3 statement on a much less significant topic (e.g. 'the handouts for lecture Z were useful but having page numbers would have made them easier to use'). However, having analysed feedback from across multiple teaching episodes for each course type, the impact of this limitation should be reduced.

Another limitation is that the tool used (Brown and Glover 2006; Newton, Wallace, and McKimm 2012) was developed for the analysis of feedback written to, rather than by, students, with a focus on constructively improving student achievement. As described in the methods section, the focus of this instrument is aligned with the main aims of the study, making it suitable for use. Nevertheless, future developments of this work might involve the development of specific instruments for the coding of student-generated feedback that takes account of all the motivations for collecting student feedback, such as teacher appraisal, benchmarking, quality assurance and review (Buckley 2012).

Finally, it is important to consider what happens once feedback is collected. Although the current study indicates that a structured feedback form is associated with the generation of feedback that is deeper and more constructive, this does not necessarily mean that these constructive suggestions are then acted upon. These are important areas for future study, for example, is the feedback generated via the 'Stop,

Start, Continue' format actually used to change educational provision? Is it more useful for that purpose than feedback generated by pure free text methods? Further, do educators prefer feedback in the Stop-Start-Continue vs. the free text format?

In summary, much of the literature on assessment and feedback has focused on the generation of feedback for students, and on quantitative feedback by students, with less attention apparently paid to the means by which qualitative feedback written by students can be optimised for course development and other outcomes. Although there have been developments in this area (e.g. see Buckley 2012), and the use of student feedback for appraisal of teachers is well established in the USA and other countries (Basow and Martin 2012; Berk 2012), there was a need for further study to ensure that the maximal benefit is harvested from the effort put into the generation of feedback by students, and the subsequent analysis of that feedback by educators intending to improve teaching provision. The present findings highlight the importance of using feedback forms which are specifically tailored towards the type of evaluation that is required, as the specific questions asked on a feedback form impact upon the type of feedback that is collected, and therefore the extent to which that feedback may be useful.

These findings suggest that the purpose of an evaluation may be key to determining the most appropriate format by which feedback is collected. If the purpose of an evaluation is to identify how a programme can be improved, for example, it is important to include questions that explicitly ask what improvements can be made. If, however, the purpose of evaluation is to identify more general views and opinions, then open questions may be appropriate. Whilst the use of open, free text questions such as those used by Course A in Study 1 may be a source of useful information, such formats do not lend themselves to collecting feedback which could inform development, and the volume of feedback generated may provide an impediment to its use. It is likely that a combination of both structured questions and free text may be preferable, as this allows students to make suggestions for change, whilst also ensuring they have the opportunity to express any other views that may not strictly fit within the Stop, Start or Continue headings.

The findings reported here support the use of a modified Stop/Start/Continue format which allows for the efficient collection of qualitative student feedback while retaining opportunities for full expression.

Notes on contributors

Alice Hoon is a lecturer in Assessment and Evaluation at the College of Medicine at Swansea University. She contributed to study design and manuscript preparation. She also coded and analysed all the feedback described in the study.

Emily Oliver is a lecturer in Sport and Exercise Psychology in the School of Applied Social Sciences at Durham University. Her research interests include motivation and engagement. She contributed by facilitating data collection, editing the initial manuscript and revising post-peer review.

Kasia Szpakowska is an associate professor in the College of Arts and Humanities at Swansea University. She contributed by facilitating data collection.

Philip Newton is an associate professor in the College of Medicine at Swansea University. His research interests include evidence-based education. He contributed to study design, supervision, data analysis manuscript preparation and revision post-peer review.

References

- Basow, S. A., and J. L. Martin. 2012. "Bias in Student Evaluations." In *Effective Evaluation of Teaching*, 40–49. Society for the Teaching of Psychology. <http://teachpsych.org/ebooks/evals2012/index.php>.
- Berk, R. A. 2012. "Top Five Flashpoints in the Assessment of Teaching Effectiveness." *Medical Teacher* 35: 1–12. doi:10.3109/0142159X.2012.732247.
- Brockx, B., K. Van Roy, and D. Mortelmans. 2012. "The Student as a Commentator: Students' Comments in Student Evaluations of Teaching." *Procedia – Social and Behavioral Sciences*, International Conference on Education & Educational Psychology (ICEEPSY 2012) 69 (Dec.): 1122–1133. doi:10.1016/j.sbspro.2012.12.042.
- Brown, E., and C. Glover. 2006. *Research Article: Written Feedback for Students: Too Much, Too Detailed or Too Incomprehensible to Be Effective?* <http://www.bioscience.heacademy.ac.uk/journal/vol7/beej-7-3.aspx>.
- Buckley, A. 2012. *Making It Count*. UK Higher Education Academy Resources. http://www.heacademy.ac.uk/resources/detail/nss/Making_it_count.
- Edström, K. 2008. "Doing Course Evaluation as if Learning Matters Most." *Higher Education Research and Development* 27 (2): 95–106. doi:10.1080/07294360701805234.
- George, J. W., and J. Cowan. 1999. *A Handbook of Techniques for Formative Evaluation: Mapping the Student's Learning Experience*. London: RoutledgeFalmer.
- Grebennikov, L., and M. Shah. 2013. "Student Voice: Using Qualitative Feedback from Students to Enhance Their University Experience." *Teaching in Higher Education* 18 (6): 606–618. doi:10.1080/13562517.2013.774353.
- Gruber, T., S. Fuß, R. Voss, and M. Gläser-Zikuda. 2010. "Examining Student Satisfaction with Higher Education Services: Using a New Measurement Tool." *International Journal of Public Sector Management* 23 (2): 105–123. doi:10.1108/09513551011022474.
- Huxham, M., P. Laybourn, S. Cairncross, M. Gray, N. Brown, J. Goldfinch, and S. Earl. 2008. "Collecting Student Feedback: A Comparison of Questionnaire and Other Methods." *Assessment & Evaluation in Higher Education* 33 (6): 675–686. doi:10.1080/02602930701773000.
- Kabanoff, B., A. Richardson, and S. Brown. 2003. "Business Graduates' Perceptions of the Quality of Their Course: A View from Their Workplace." *Journal of Institutional Research* 12 (2). <http://www.aair.org.au/articles/volume-12-no-2/12-2-business-graduates-perceptions-of-the-quality-of-their-course-a-view-from-the-workplace>.
- Kindred, J., and S. N. Mohammed. 2005. "'He will Crush You Like an Academic Ninja!': Exploring Teacher Ratings on Ratemyprofessors.com." *Journal of Computer-Mediated Communication* 10 (3). doi:10.1111/j.1083-6101.2005.tb00257.x. <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2005.tb00257.x/full>.
- Marriott, P., and L. K. Teoh. 2014. *Computer-based Assessment and Feedback: Best Practice Guidelines*. Higher Education Academy UK. Accessed June 27. <http://www.heacademy.ac.uk/resources/detail/disciplines/finance-and-accounting/guidelines-using-ICT-in-assessment-feedback>
- Marsh, H. W., and L. A. Roche. 1997. "Making Students' Evaluations of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias, and Utility." *American Psychologist* 52 (11): 1187–1197. doi:10.1037/0003-066X.52.11.1187.
- Newton, P. M., M. J. Wallace, and J. McKimm. 2012. "Improved Quality and Quantity of Written Feedback is Associated with a Structured Feedback Proforma." *Journal of Educational Evaluation for Health Professions* 9: 10. <http://www.jeehp.org/DOIx.php?id=10.3352/jeehp.2012.9.10>.
- Nulty, D. D. 2008. "The Adequacy of Response Rates to Online and Paper Surveys: What can be Done?" *Assessment & Evaluation in Higher Education* 33 (3): 301–314. doi:10.1080/02602930701293231.
- Palmer, S., and M. Campbell. 2013. "Practically and Productively Analysing Course Experience Questionnaire Student Comment Data." In *AAEE 2013: Proceedings of the 24th 2013 Australasian Association for Engineering Education Conference*, January. <http://dro.deakin.edu.au/view/DU:30060776>.
- Richardson, J. 2005. "Instruments for Obtaining Student Feedback: A Review of the Literature." *Assessment & Evaluation in Higher Education* 30 (4): 387–415. doi:10.1080/02602930500099193.

- Shute, V. J. 2008. "Focus on Formative Feedback." *Review of Educational Research* 78 (1): 153–189. doi:[10.3102/0034654307313795](https://doi.org/10.3102/0034654307313795).
- Symons, R. 2006. *In Their Own Words: Finding out What Students Think about Their University Learning Experience*. University of Sydney Institute for Teaching and Learning. <http://www.itl.usyd.edu.au/synergy/article.cfm?articleID=281>.
- Williams, R., and J. Brennan. 2004. *Collecting and Using Student Feedback Data: A Guide to Good Practice*. Higher Education Academy UK. http://www.heacademy.ac.uk/resources/detail/Collecting_and_using_student_feedback.