Using Microsoft Excel to code and thematically analyse qualitative data: a simple, cost-effective approach.*

Ronan Bree

Gerry Gallagher

Dundalk Institute of Technology

Abstract

As the number of learning and teaching continuing professional development (CPD) courses increases in Higher Education Institutions (HEIs), so too does the accompanying number of learning innovations being implemented and evaluated. The evaluation process requires valid and reliable data collection and analysis procedures to be established. In many cases, qualitative methods such as interviews, focus groups and free-text responses are employed for this purpose. These methods generate large volumes of data, which must be coded and analysed in a thorough and professional manner. While commercial software packages can assist in this analysis, in a difficult economic climate, the cost of campus-wide licenses for such can be quite prohibitive. In a recent publication aimed at enhancing the learning environment in practical sessions, Bree et al. (2014) implemented a simple, cost-effective technology-based analysis of captured focus group data with a widely used software suite. This report describes in detail the qualitative data analysis process designed and implemented by the two authors (a link to a screencast outlining the method is also provided). Ensuring data analysis processes are performed correctly will generate valid data, leading to an increase in the number of peer-reviewed publications describing learning and teaching innovations; with each of these ultimately enhancing the learning environment for students and developing higher-quality graduates. It is hoped this report will assist educators enrolling on learning and teaching courses, or those performing research projects in the area who are considering employing qualitative evaluation methods.

KEYWORDS: Learning, Teaching, Qualitative Methods, Evaluation, Data analysis, Thematic Analysis.

* URL: http://ojs.aishe.org/index.php/aishe-j/article/view/[281]

1 INTRODUCTION

In 2011, the Hunt report was published by the Irish Government outlining a National Strategy for Education in Ireland to 2030 (Department of Education and Skills, 2011). One significant recommendation stated academics must become conscious of, and engage with, learning theories and their application, and implement suitable teaching and learning approaches based on the student and the learning environment (no longer remaining experts in just their particular discipline). In recent years, the learning and teaching arena has evolved into its own, emerging research area. According to a report from The National Forum for the Enhancement of Teaching and Learning in Higher Education on professional development, in Ireland there are currently sixty-eight accredited learning and teaching programmes available in twenty-two Higher Education Institutions (HEIs) (National Forum for the Enhancement of Teaching and Learning in Higher Education, 2015). These programmes often involve a research strand, in which a learning and teaching innovation aimed at improving the learning environment for students is designed, implemented and evaluated in an ethically approved manner.

1.1 QUALITATIVE EVALUATION APPROACHES

Many evaluations of learning and teaching interventions and research projects will involve qualitative methods originally used in social science, including for example, questionnaires/surveys, interviews and focus groups are regularly employed, either individually or combined (Evans & Mathur, 2005; Adams & Cox, 2008; Cohen et al., 2011). It is imperative the data collected from these approaches is triangulated and analysed in a professional, unbiased and thorough manner, contributing to ensuring the results and findings are valid and reliable (Cohen et al., 2011). In some institutions, expensive software can be purchased to assist in the analysis process.

In many cases, once qualitative data is collected, the researcher must transcribe all audio or written data in to an electronic text format to facilitate analysis. While there are excellent software options available to purchase that can assist with the analysis process itself (e.g. NVivo by QSR International or Atlas.ti), in the current economic climate it may not be feasible for many Higher Education institutions to

purchase campus-wide licenses accommodating every investigator in an ever-increasing area of research. During previous research (Bree et al., 2014), the authors designed and implemented a relatively easy and cost-effective method to analyse focus group transcripts/data. In the cited study, a large quantity of raw, handwritten data from two focus groups needed to be thematically analysed and triangulated. This short report describes the approach used for this analysis, which can be performed using the Microsoft Office suite. The approach is transferable to any field or data set involving large amounts of textual data such as interview or focus group data. Therefore, it is imperative that if a suitable and cost-effective data analysis approach such as this is designed and implemented successfully, it should be shared with the wider community.

2 ANALYSING QUALITATIVE DATA

According to Braun and Clarke (2006), the term 'thematic analysis' refers to the "*method of identifying, analysing and reporting patterns (or themes) within data*" and represents a foundational, conceptually demanding method for qualitative analysis that researchers should learn, due to the skill set it can provide. In the study by Bree et al. (2014), several elements of the frameworks presented by Braun and Clarke (2006) are evident, ensuring rigour in the data analysis process. For example, a rich thematic description of all the data collected was performed ensuring the reader was presented with a valid reflection of the entire data set. An inductive approach was implemented by the authors, meaning the data was coded, or categorised for analysis, without fitting it to a pre-determined coding frame. This ensured the analysis process was driven by the data collected during the evaluation process rather than any analytic preconceptions. The level at which themes are identified, whether it be semantic (where surface meanings of the data are used) or latent (where underlying ideas and concepts are identified in data) must also be considered. In the process outlined in this report, the inductive approach implemented by the authors first identified themes which were analysed initially in a descriptive form (with data organised to show patterns in a semantic content) before progressing to an interpretative form which attempted to look beyond the surface of the data (where the broader meaning and ultimate implications of the themes/patterns are deduced, often via engagement with literature) (Frith and Gleeson, 2004; Braun and Clarke, 2006).

With regard to the physical process of managing the data analysis, the authors were initially guided to a traditional paper-based approach whereby comments from each moderator's transcript are cut out and thematically arranged by sticking them to whiteboard under certain headings, essentially coding the data. It is at this stage that emerging themes and key points are identified. During this identification process, it is important to maintain an open mind, one welcoming and accepting of alternative views and sustain this during all stages of the data analysis. The analysis process continues with the theming of the data into categories, reflecting and synthesising the data, condensing data to assist reporting, interpretation of the data and the findings – all prior to commencing the final report (Brenner, 1985). In addition to an overview of the key data and emerging findings, the report can often contain partial, anonymous quotations to complement the findings identified. These final data transcripts, data analysis reports and findings need to be shared with all participants for verification and cross-checking. The ultimate step, which involves rethinking, must be carried out before the report can be finalised. With this structure in mind, Table 1 outlines the data analysis process performed by the authors in their earlier study (Bree et al., 2014) and aligns the processes used to the phases of thematic analysis described by Braun and Clarke (2006) (See Appendix 1; Table 4).

Data Analysis Process Tasks Performed	Thematic Analysis Phase (see Braun & Clarke, 2006; Appendix 1; Table 4)
Focus groups performed as per Krueger (2002)	
Sessions co-moderated and hand-written notes recorded	1
Hand written notes transcribed in to electronic format (Microsoft Word initially and subsequently transferred to Microsoft Excel; one comment per Excel cell)	1
Points from each moderator cross-referenced and compared ensuring all points were recorded accurately while duplicate entries were removed	1 & 2
Final list of focus group comments/points compiled for analysis	1 & 2
First pass over data centred on the identification of themes	2 & 3
Themes identified in the data with each appropriate cell colour coded, i.e. each data point was colour coded to match thematic area	3,4 & 5
Microsoft Excel's filter applied to data. The filter was established to sort by cell colour (and hence thematic area)	3,4 & 5
Second pass over data identified overlapping points and consolidated data points; consolidated data copied to new Microsoft Excel worksheet (see screencast link provided at end of section 2.0)	4 & 5
Numerous further passes over data, condensing data at each stage before moving to new worksheet (this is the most time-consuming step of the process)	5 & 6
Generation of data overview summary and top 5 key points under each theme	6

Table 1: An overview of the focus group data collection and analysis approach. Where appropriate, tasks are presented aligned with thematic analysis phases described by Braun & Clarke (2006). Please refer to Table 4 in Appendix 1 for a description of each of the phases.

Essentially, to analyse the data generated in Bree et al. (2014), handwritten transcripts from both focus groups and from both moderators were transcribed in to electronic format via Microsoft Word (the majority of HEIs have licenses with Microsoft for both faculty and students to use the Microsoft Office 365 suite). The transcripts were organised per question and the comments from each moderator were put in regular font or *italic* font to distinguish between them. A 'b' was added to the second focus group's data to distinguish it from the first group's data and to facilitate triangulation at a later stage (i.e. Q1 data compared to Q1b data – see Appendix 1, Figures 2a, 2b).

Following this, all the data was migrated to a Microsoft Excel worksheet, generating a single column consisting of all comments from both moderators (see Appendix 1, Figures 2a and 2b for example of transcript data entered in to Microsoft Excel for one focus group question). The data was initially reviewed to remove any duplicate entries, arising from moderators recording the same points during the focus group. The data was then analysed with a view to assigning thematic areas. Each individual cell was reviewed and assigned to a thematic area, to which a cell colour code was applied. As mentioned previously, this was performed using an inductive approach; there were no pre-determined thematic areas in place before the coding process was initiated. This data-driven, inductive approach ultimately led to the identification of ten thematic areas, each labelled with a different colour (see Table 2 for an example of the themes analysed in the particular study). If some points fell under two themes, the comment/worksheet cell was duplicated and each cell colour coded appropriately to ensure everything was recorded.

Feedback
incremental marking
process/undertaking approach
lab manual/exercises/skill tests
group work
emotion / reflection
Lecture-lab-knowledge
exemplar
pre-submission lists/checklists/self-assessment/format
other recommendations

Table 2: An overview of the ten themes identified for which each data point was assigned a corresponding colour code.

Microsoft Excel possesses a tool regularly used to sort a column's data in ascending or descending order. However, this tool can also sort the data based on the colour of the cell (or even font colour if necessary). Using this tool, the focus group data was sorted by the cell colour assigned to each cell, and hence sorted by thematic area. This sorting and collation approach brought together all the key points on each theme, facilitating further analysis (all taking place in a worksheet of the Excel workbook) (See Appendix 1, Figure 3 to see an example of the result of the colour coded sorting process for the 'feedback' themed data – i.e. all cells coded with a green colour now sorted using the Excel data sorting facility). However, at this stage, there remained a large volume of data to be analysed (337 comments/worksheet cells).

Initially, the data was copied to a second Excel worksheet. This step involved a 'first pass' over the data, making attempts to combine or group similar points or comments, ultimately a first attempt to consolidate the data. Every time this data pass task was performed for each thematic area, the data was moved to a new worksheet. The reason for the multiple worksheets, was that one could always go back a step to the previous unconsolidated data set if needed. Also, this approach assisted in the validation of the data analysis by an independent party as they could follow each step performed during the process via viewing the various worksheets. In each worksheet, another data pass was performed, further identifying sub-themes in the data, linking appropriate quotes to points, further condensing the data. Ultimately, this step was repeated numerous times to finally generate the top five key points per theme in addition to the key points related to that theme, i.e. the data overview (see Figure 1 for an overview of the study's data analysis process and Table 3 for an example of the output for one thematic area).

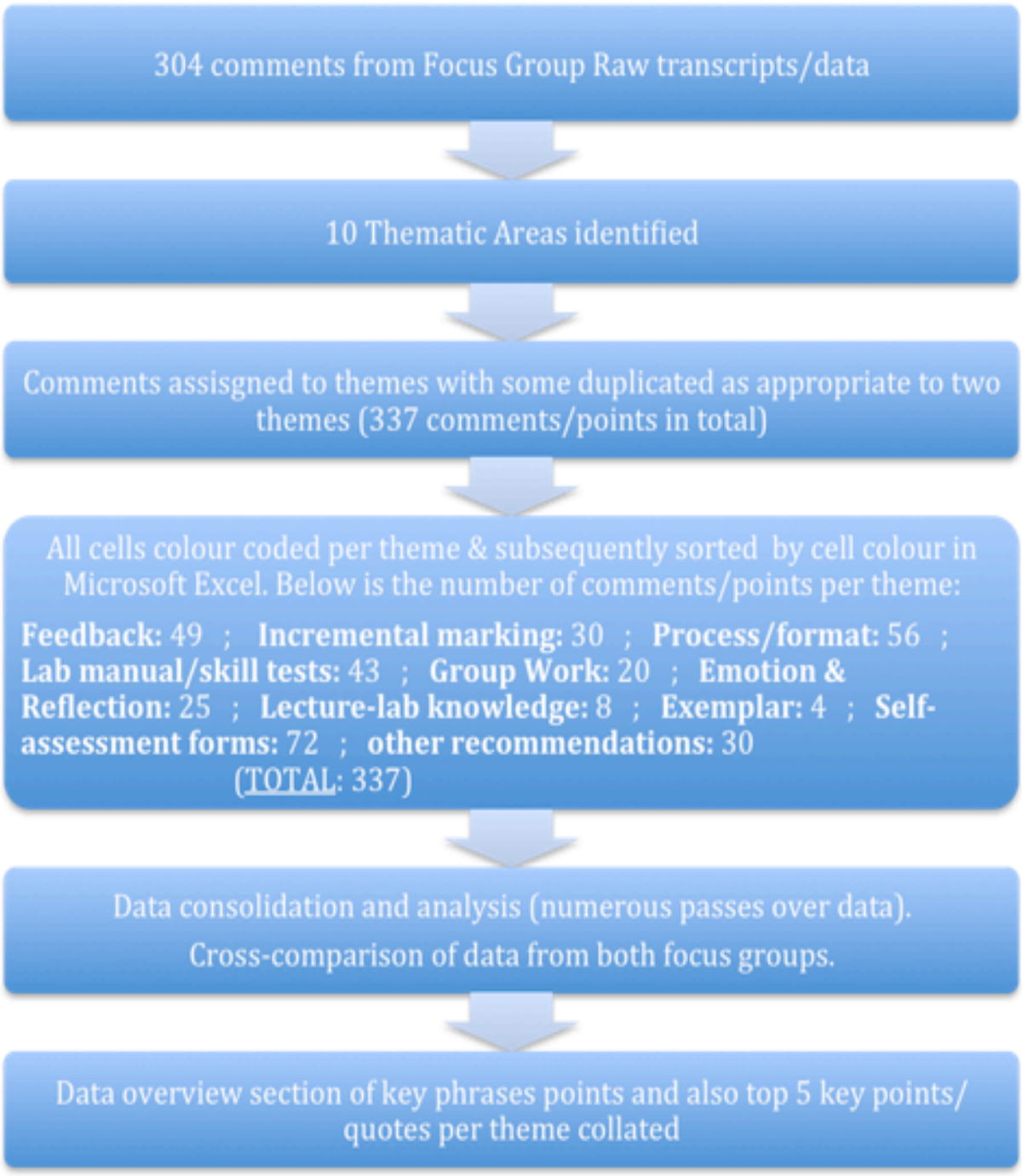

Figure 1: Schematic overview of the data consolidation process used in Bree et al. (2014) by the authors.

	Data overview & top 5 key points
THEME	FEEDBACK
Data overview	*Correct your mistakes, improve as you go (with sheet), feedback sheet lets you correct your work, apply the feedback, adapt to lecturer style, better than ticks/mark, clarity, leave feedback with student, essential, quick return of feedback appreciated and very important, got feedback while you still remembered, tips, track progress, "can't improve without feedback", feedback works forward - you knew what to do to get better, student made own notes from feedback - guide, detailed feedback, sheet sections a good idea, specific, breakdown lets you see what parts to work on, personalised better than general feedback, more effort-more reward, effort from lecturer visible and appreciated. feedback is the most important factor to do with the labs. questions helped the understanding.*
1	Feedback essential for the process of feed-forward improvement; "can't improve without feedback" vs. 'ticks'
2	Quickly returned, Personalised, section by section feedback sheets is the way to give feedback
3	Empowering student will assist implementing feedback, making their own feedback notes
4	Realised the better the effort, the better the feedback received, better learning
5	Demanding on staff member (although acknowledged and appreciated by students)

Table 3: Consolidation of Focus Group Data. An example of the 5 key points and data overview extracted from the focus group raw data (example thematic area presented here: Feedback)

Overall, this process assisted the integration and triangulation of data from two individual focus groups. In order to view this process in action, the authors recommend watching a brief screencast of the analysis being performed with mock data which can be viewed at the following link, which can be accessed directly or by scanning the QR code shown.

https://youtu.be/VIYBEE-1GbA

3 CONCLUSION

With increasing numbers of staff in Irish HEIs undertaking CPD courses in learning and teaching, it is critical to have in place, and share, resources that can assist the research of these and other educators globally. This report focuses on a cost-effective approach used to assist the thematic analysis and triangulation of qualitative data collected in learning and teaching research based focus groups – one that is easily transferable to interview and survey data. The approach integrates a widely used application software suite for organising, coding and classifying data. In particular, it uses colour and sorting features of Microsoft Excel and crucially, does not require an advanced knowledge of the software. The report emphasises the importance of the data analysis step in increasing the validity and reliability of findings, and outlines a method of analysis which is accessible to researchers. In this way, it is hoped to contribute to the rigorous analysis of data, thus enhancing evaluative processes and ultimately leading to more peer-reviewed publications for researchers and an enhancement of the learning environment for students.

4 REFERENCES

Adams, A. and Cox, A. L. (2008). Questionnaires, in-depth interviews and focus groups. In: P. Cairns, A.L. Cox, eds. *Research Methods for Human Computer Interaction*. Cambridge, UK: Cambridge University Press, pp. 17–34.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), pp. 77-101

Bree, R.T., Dunne, K., Brereton, B., Gallagher, G. and Dallat, J. (2014). 'Engaging learning and addressing over-assessment in the Science laboratory: solving a pervasive problem.', *The All Ireland Journal of Teaching and Learning in Higher Education (AISHE-J)*, 6(3), pp. 206.1-206.36 Available at: http://ojs.aishe.org/index.php/aishe-j/article/viewFile/206/290 [Accessed 18 Apr. 2016].

Brenner M. (1985). *The Research Interview: Uses and Approaches*. London: Academic Press.

Cohen, L., Manion, L. and Morrison K. (2011). *Research Methods in Education*. 7th edition. London and New York: Routledge Press.

Department of Education and Skills (2011). National Strategy for Higher Education to 2030 - Report of the Strategy Group. [online]. Available at: http://www.hea.ie/sites/default/files/national_strategy_for_higher_education_2030.pdf [Accessed 17 Mar. 2016].

Evans, J.R. and Mathur, A. (2005). The Value of Online Surveys. *Internet Research*, 15 (2), pp. 195-219.

Frith, H. and Gleeson, K. (2004). Clothing and embodiment: men managing body image and appearance. *Psychology of Men and Masculinity*, 5, pp. 40-48.

Krueger R.A. (2002). Designing and Conducting Focus Group Interviews. *Eastern Illinois University*, [online]. Available at http://www.eiu.edu/~ihec/Krueger-FocusGroupInterviews.pdf [Accessed 12 Dec. 2015].

National Forum for the Enhancement of Teaching and Learning in Higher Education (2015). *Professional Development Reports: A Snapshot of Accredited Professional Development Provision in Irish Higher Education.* [online] Available at: http://www.teachingandlearning.ie/wp-content/uploads/2014/03/Accredited-Professional-Development-Final.pdf [Accessed 18 Apr. 2016].

5 APPENDIX 1:

Phase Number	Phase Name	Description of Phase
1	**Familiarising yourself with your data**	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2	**Generating initial codes**	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3	**Searching for themes**	Collating codes into potential themes, gathering all data relevant to each potential theme.
4	**Reviewing themes**	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
5	**Defining and naming themes**	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme.
6	**Producing the report**	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

Appendix 1; Table 4: Thematic Analysis Phases as presented by Braun and Clarke (2006).

	normal font = primary moderator's comments, italics = co-moderator's comments 1,2,3,4,5,6 = group 1; 1b, 2b, 3b, 4b, 5b, 6b = group 2
	Focus Group
Q1	
1	**Feedback Sheets** – allow you to turn wrongs in to rights
1	**Incremental marking system** (with feedback sheets)
1	Styles of other lecturers can have impact – different styles. Allows adaptation.
1	More marks when you improve
1	**Lab Manuals** – questions during the lab, exercises, principle covered well
1	Not trying to think back
1	"understood as you went"
1	Other labs, finished in 1 hour, material "not clear in head"
1	**Pre-lab talk important** – "stick in your head", "know what you are doing in an experiment" "not done for the sake of it"
1	...leads you to leaving the lab knowing what happened.
1	No write up weeks – did not effect learning (as exercises helped learn/understand).
1	**Lecturer sign off** of lab work before leaving the lab: Gave a good overview of the lab, key points again being summarised.
1	One student felt they didn't need this as they check themselves.
1	Methods in lectures coinciding with labs – "all related", "linked"
1	**Lab Manual** – calculations exercise – good to have them spread out over the term rather than do all 20 in one session. Good to fit them in to free time/gaps in lab sessions. Empowered them – do it yourself and then cross reference when going through them in class.
1	Different **discussions** – lead to viewpoints
1	**Group project**: coincided with reading week, would have preferred it during term time. Too free; still a learning curve; group all appreciated why it is there.
1	1 drop it; 5 keep it.
1	Might not like it but felt it was still good.
1	*Feedback sheets were clear and helpful*
1	*Incremental marks*
1	*Knew how to correct your mistakes*
1	*Each lecturer has his or her own system – one can learn from Ronan's Feedback, build on knowledge*
1	*Could build on it further in semester 2 - lets you kick on in semester 2 Feedback sheets let you ask for clarification*
1	*As you improved, your marks went up*
1	*Lab manual layout – questions make you think in the lab – understanding the principles as you went In other labs, which are fast, you don't understand, hard to write up*
1	*Here, student walked out of lab confident*
1	*Labs made it stick in your head "I understood as I walked out" Questions exercises – "you weren't doing them for the sake of it" Reduced workload, even with no write-ups, you still understood it. Not having lab write ups didn't stop the learning*
1	*Check list at the end was helpful – 1 student didn't like this as she said she checked herself anyway. Checklists avoids forgetting aspects*
1	*Signing off lab at the end – was that helpful? Yes, good overview, re-enforced key points.*

Appendix 1; Figure 2a: A screenshot of the data coding process for part of one focus group question whereby each cell is colour coded based on the content using an inductive approach. Please refer to Table 2 in the main text for colour code:identified theme correlation.

1b	Feedback – show how you went wrong
1b	Writing up – see problems, pre submission checklist, taking notes in lab
1b	This approach helps you generate professional and wanted lab reports , feel happy with submissions Using the biochemistry approach to write-ups as a basis for other modules
1b	Exemplar - a good thing, but best not to have it too early. Best to do at least the first one on your own, find your own way, make your own mistakes, build on that. Do it first to know what is right/wrong. Develop your own way rather than copy/mirror others.
1b	Lab manual – exercises – very helpful
1b	Helped remember material after the lab had finished
1b	Questions in manual became a go-to tool
1b	manual/exercises - Kept rough idea in your head (questions), helped in exams.
1b	4 lab reports – focus – need to get marks, want to achieve more, improve through incremental system
1b	Worth more = more work
1b	Good to have 1st one at 4%
1b	At end = lab quality much better, big difference between first and last submission
1b	For last – talked and planned out report before starting
1b	Feedback - important for system to help you.
1b	High % lab reports – could be costly if you didn't understand. Concern from one student that labs became more difficult as term progressed – and were also worth more.
1b	*At the beginning, you didn't know what the lecturer wanted, but the feedback sheets explained where you went wrong.*
1b	*Feedback worked as it helped you to get better marks.*
1b	*I didn't know how to actually structure a report. So the pre-submission list helped me a lot. It clarified what was needed, made it more professional.*
1b	*I'm doing better in labs generally now. I am using the basis I learned (in biochem) now in other labs. The example of really good work helped. It was more helpful at that stage (half way through term).*
1b	*The 1st write up is experimental. You have to make your own mistakes to learn from it. Build on them.*
1b	*We wouldn't have paid enough attention if we saw it (good exemplar) first, people would have just copied it.*
1b	*Questions in the lab manual helped with understanding. They re-enforced it before you left the lab. It helped you when you were going back over it.*
1b	*manual - It meant you knew what the lab was about as you did it, before you wrote it up.*
1b	*Helped prepare for the end of semester exam as well as write-up.*
1b	*Fewer labs made you work harder. The fact that the marks went up made you work harder.*
1b	*When something is worth more, you try harder.*
1b	*It was god having the lower ones at the start – as you were learning how to do it, like a trial run. By the end, the quality was mush higher, even though you only wrote 4 write-ups.*
1b	*By the end, we planned and discussed it properly before we wrote it.*
1b	*The feedback made you able to do a good job on the final lab, which was worth the most.*
1b	*As the labs get harder, the marks go up, the penalty for not understanding is more costly. You need more background understanding from earlier labs. (one student)*

Appendix 1: Figure 2b: A screenshot of the data coding process for part of one focus group question whereby each cell is colour coded based on the content using an inductive approach. Please refer to Table 2 in the main text for colour code:identified theme correlation.

	normal font = primary moderator's comments, italics = co-moderator's comments 1,2,3,4,5,6 = group 1; 1b, 2b, 3b, 4b, 5b, 6b = group 2
	Focus Group
Questio	Open Coded data
1	**Feedback Sheets** – allow you to turn wrongs in to rights
1	*Feedback sheets were clear and helpful*
1	*Knew how to correct your mistakes*
1b	Feedback – show how you went wrong
1b	Feedback - important for system to help you.
1b	*At the beginning, you didn't know what the lecturer wanted, but the feedback sheets explained where you went wrong.*
1b	*Feedback worked as it helped you to get better marks.*
1b	*The feedback made you able to do a good job on the final lab, which was worth the most.*
2	Improve as you go (with use of feedback sheet)
2	On student's idea – keep the marking system quiet? But when discussed further, group said no. Feedback again playing a big role.
2	*Feedback was helpful – lets you learn what the lecturer wants*
2	*Feedback sheet lets you correct your work.*
2	*Hide the idea of using an incremental system? Using feedback to refine your approach each week.*
2	*People who do try in the lab benefited, you got more feedback (vs. comment above)*
2b	Feedback – very specific
2b	Feedback helps you improve over the term Detail in feedback – still needed.
2b	*The feedback was necessary to help you improve for the late labs.*
3	This system "should be in all of them" (courses).....should be the standard. Normally ticks/marks – not enough information or clarity in this method Like the personalised approach
3	General feedback = apply it
3	Good to leave feedback with the student
3	Each section on the feedback form was a good idea (given feedback on each section of report) Knew what did/didn't work for each section.
3	"essential" in process
3	"can't improve without feedback"
3	Quick return of feedback = appreciated and v important
3	*Should be in all of them (modules).*
3	*Brilliant.*
3	*Ticks and marks are no help to improving (in other labs)*
3	*Feedback explained what to do, how to fix it. Improvement was visible, you knew what to di to get better.*
3	*Helpful, even though the next one will be on a different topic.*
3	*Keeping the feedback sheets was helpful, leave feedback with student.*
3	*Sheet layout used was helpful. You knew exactly where the feedback applied. Breakdown let you see what parts to work on.*

Appendix 1: Figure 3: An example of the Microsoft Excel colour coding sorting process detailed in the text. Here the cells were sorted based on being coded with a green colour (which corresponded to the feedback theme identified – see Table 2 in the main text).