Making internal feedback explicit: harnessing the comparisons students make during two-stage exams

David Nicol & Geethanjali Selvaretnam

Published online: 28 Jun 2021.

Submit your article to this journal ☐

Article views: 1846

View related articles ☐

View Crossmark data ☐

Citing articles: 1 View citing articles ☐

Routledge
Taylor & Francis Group

Making internal feedback explicit: harnessing the comparisons students make during two-stage exams

David Nicol ![ORCID] and Geethanjali Selvaretnam ![ORCID]

Adam Smith Business School, University of Glasgow, Glasgow, UK

ABSTRACT

In two-stage exams, students complete some academic work individually then immediately complete the same work again in small groups. Prior research shows that students perceive the dialogue they engage in at the group stage aids their individual learning. Outcome studies also suggest that engaging in dialogue benefits individual student learning. This investigation takes a different perspective. Learning during two-stage exams is conceptualised as deriving from internal feedback processes, with comparison seen as the mechanism that fuels these processes. To investigate this interpretation, students were asked to write a feedback commentary on their own learning during a two-stage exam. The intention was to make explicit the internal feedback that would naturally occur during the group stage, as students compare their earlier individual performance against the group dialogue and the unfolding group output. Analysis showed that students not only generated content and process feedback but also self-regulatory feedback, and that this feedback was of a higher quality than, and went beyond, the comments the teacher provided. This investigation yields new insights into the nature, scope and power of internal feedback processes, and into how to improve students' learning during such two-stage sequences.

Introduction

In two-stage exams, students complete some academic work individually and then, after handing in this work for grading, do the same work again in small groups of three or four. (Gilley and Clarkston 2014; Zipp 2007). All students in the group must agree the group output which also contributes to their overall individual grade. The academic work might be a written assignment such as a problem task or a set of multiple-choice questions. In most implementations, the individual work accounts for a larger proportion of the grade than the group submission (Zipp 2007). The two-stage exam can take place during a course or at the end of the course. This two-stage sequence is also easily implemented for formative purposes without grading. It is important to note that two-stage exams are also referred to as two-phase exams, two-stage collaborative exams, pyramid exams and cooperative exams (Efu 2019). Two-stage exams, rather than two-stage examinations, is the most frequent designation so that is the term used in this article.

Many positive learning benefits have been claimed with regards to two-stage exams, which have been researched across a range of disciplines, especially engineering and science. Students engage in active and collaborative learning which gives them the opportunity to practice and develop valuable group work skills (Lusk and Conklin 2003; Muir and Tracy 1999; Rieger and

Heiner 2014). During the dialogue that occurs at the group stage, students give and receive verbal feedback, which helps elaborate their understanding of the work they produced at the earlier individual stage (Drouin 2010; Gilley and Clarkston 2014; Levy, Svoronos, and Klinger 2018). The immediacy of this feedback is unusual in normal teaching practice and it is even more uncommon in an examination situation (Rieger and Heiner 2014).

Even the best students learn from this dialogue, as giving explanations improves one's own understanding (Zipp 2007). Also, the group output is usually superior to any individual output (e.g. Levy, Svoronos, and Klinger 2018), hence the best students as well as weaker students have a high-quality work against which to benchmark their own. Students report that participating in two-stage exams is motivating and research suggests that this is correlated with improved learning (Rieger and Heiner 2014; Shindler 2004). Research also shows that students retain what they learn through this two-stage process more than they do through individual working. The latter is reflected in students' individual performance in subsequent tasks (Cortright et al. 2003; Gilley and Clarkston 2014)

In published research, the feedback dialogue resulting from collaboration at the group stage is posited as the main factor contributing to improved learning in two-stage exams (Gilley and Clarkston 2014; Levy, Svoronos, and Klinger 2018; Rieger and Heiner 2014; Zipp 2007). Dialogue fosters 'debate and exchange of ideas… which allows students to learn from their peers, filling each other's gaps in knowledge' (Levy, Svoronos, and Klinger 2018, p2). Evidence for this interpretation comes from observing students as they engage in two-stage exams, from students' self-reports and indirectly from outcomes studies. What is absent is any research about the internal feedback processes that students engage in during two-stage exams. That is the focus of this article.

Internal feedback

Defining internal feedback

Feedback is normally conceived as the information that someone provides to students about their work and performance. The provider is usually a teacher but in two-stage exams it is a number of peers. However, from a cognitive perspective this external information does not really constitute feedback or operate as feedback until students internalise it and create new knowledge and understanding out of it. In other words, it is the internal feedback processes deriving from this external information that lead to improvements in understanding and performance.

There is a body of research that takes this perspective that students generate internal feedback based on their processing of external information (Butler and Winne 1995; Nicol and Macfarlane-Dick 2006; Panadero et al. 2019). In this research, internal feedback generation is viewed as an ongoing and pervasive process. It is the catalyst for all learning and for students' self-regulation of learning. Researchers adopting this position usually maintain that students generate internal feedback by *monitoring* or *self-assessing* their own performance as they regulate their learning (Butler and Winne 1995, Andrade 2018; Yan 2020).

Recently, Nicol and his colleagues have reframed internal feedback as the result of *comparison* processes (Nicol 2019; 2020; Nicol, Thomson, and Breslin 2014; Nicol and McCallum 2021). In this reframing, students generate internal feedback by *comparing* their current work against their learning goals and against any other information they attend to in the learning environment that will help them achieve those goals. That information might come from teachers as comments, from sharing work with peers, from dialogue with them, or from textbooks, online resources, videos or performance observations. The following is the definition of internal feedback that underpins this article.

> Internal feedback is the new knowledge that students generate when they compare their current knowledge and competence against some reference information. (Nicol 2020, p2)

Note that internal feedback is not a product in this definition: rather it is a process of change in knowledge - conceptual, procedural or metacognitive.

Evidence of internal feedback generation

There is a paucity of research on internal feedback generation and its effects on students' learning (Nicol 2020). However, there are some published studies where the operation of internal feedback can be inferred. Peer review is a good example. Recent research shows that students learn as much or more from reviewing the work of peers than from receiving comments from peers (Huisman et al. 2018; Li, Lui and Steckelberg, 2010: Cho and Cho 2011; Lundstrom and Baker 2009). When researchers ask students how they learn from reviewing, they invariably report that they compare their peer's work against their own, and generate new ideas about how to improve their work (i.e. internal feedback) out of that comparison (Li and Grion 2019; McConlogue 2015; Nicol, Thomson, and Breslin 2014). Comparison is natural and spontaneous during peer reviewing because before students review the work of their peers, they will have produced similar work themselves in the same topic domain (Nicol 2014). Students also report learning different things from comparing their work with the work of peers as opposed to comparing their work against comments they receive from peers (Li and Grion 2019; Nicol, Thomson, and Breslin 2014).

Another example is a study by Lipnevich et al. (2014) where undergraduate psychology students wrote a draft research report (2-3 pages) and then were given some resources which they could use to update their drafts, so as to increase their grades. The intention was to give students 'standardized feedback that was the same for all students' (p540) and so reduce teacher workload in providing comments. One group was given a rubric, another group were given three exemplars varying in quality (i.e. high, medium and low) and a third group, both the rubric and the exemplars. All three groups showed improvements from draft to redraft without any teacher comments, with the rubric having most effect on grade improvements. Although Lipnevich et al. (2014) do not use the word *comparison*, this research clearly demonstrates that different combinations of information elicit different internal feedback effects.

Difficulty in researching internal feedback

While some researchers have made a case for internal feedback (Butler and Winne; 1995; Nicol and Macfarlane-Dick 2006; Andrade 2018) and others have identified comparison as its core and generative mechanism (Nicol, Thomson, and Breslin 2014: Nicol 2019; Nicol and McCallum, 2021), there is an absence of direct research showing what students actually generate from making comparisons. Current studies either rely on self-reports by students long after the comparisons have taken place (McConlogue 2015; Nicol, Thomson, and Breslin 2014) or on evidence of learning improvements in test-retest situations (Huisman et al. 2018; Lipnevich et al. 2014). This gap in research is unsurprising given that internal feedback is not a mainstream way of framing feedback in educational research much less in practice, and that internal feedback processes are by definition covert and hence not open to scrutiny.

The research reported here addresses the implicit nature of internal feedback by using a methodology through which its outputs are made explicit. Students were prompted to make deliberate feedback comparisons and to write down what they learned from them. The methodology is similar to that in Nicol and McCallum (2021), where students produced a short essay then compared their essays with those of their peers and wrote an account of their learning from those comparisons. These researchers found that students generated a self-feedback commentary that was more elaborate than that which the teacher provided. They also found that when students made comparisons across multiple essays, they generated feedback of a type

that the teacher would have had difficulty providing, for example, about the quality of their own essay relative to other essays.

Making the results of comparison processes explicit is not just a means of researching inner feedback. Natural comparison processes arguably always occur in two-stage exams. Making them explicit is a way of intensifying the learning from them. Research on metacognition shows that when students make explicit their own cognitive processes, they take more executive control over them (Pintrich 2002) and they learn more (Bisra et al. 2018). For example, when students externalise their own thinking as they are reading a complex text this actually improves their learning, as they identify gaps and from this, and attempt to close them by drawing on latent knowledge (Chiu and Chi 2014; Fonseca and Chi 2011). Indeed, the extensive research on learning from self-explanation could be re-interpreted from an internal feedback perspective.

Two-stage exams and internal feedback generation

Figure 1 depicts the different feedback generating opportunities available to students during a two-stage exam. The assumption is that students start by setting goals for the task. These goals will be shaped by student's motivation, the task instructions, the nature of the task, their perceptions of their abilities, and by other situational demands (e.g. timeline, working environment). Most students will have prepared for the examination so they will have elaborated and primed their relevant knowledge.

Two-stage exams provide two implicit opportunities for internal feedback generation by students: at the individual stage during the execution of the task itself and then at the group discussion stage where the students do the task again together as a group. Students first write an answer to the question on their own, making use of their prior knowledge and their
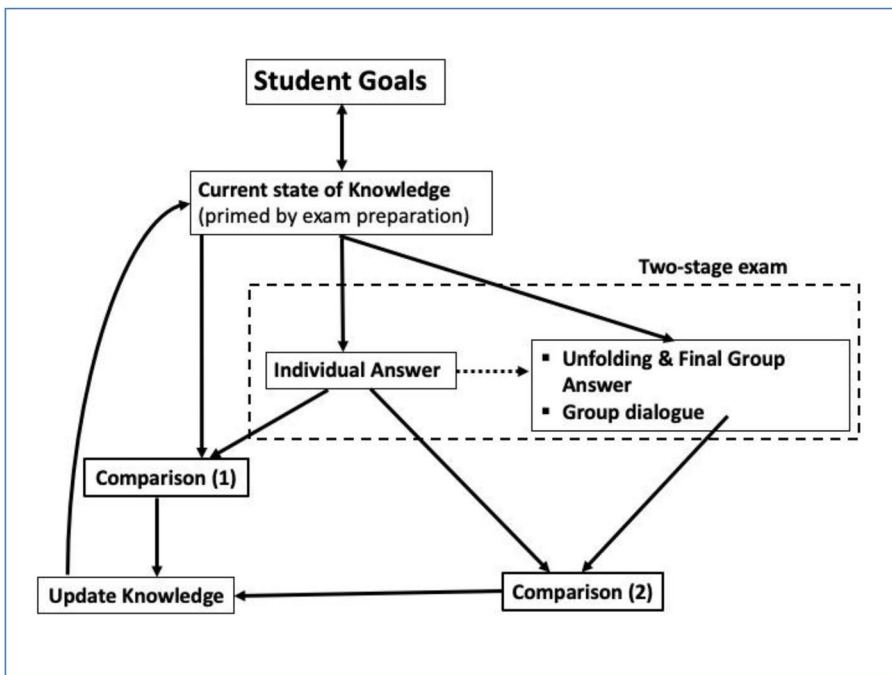

Figure 1. Internal feedback mechanism of two-stage exams.

out of class preparation. In this situation, students can compare what they are producing individually against their own goals for the task and against whatever internal reference information they can draw on to help them achieve those goals (e.g. memories of their study notes).

At the group stage, opportunities for internal feedback generation occur again. Students can compare what they have just produced at the individual stage against: (i) the unfolding and final written answer and against (ii) the ongoing group dialogue, which is about how to produce that answer. Hence, there are two external sources of information for comparison. Since the group output is produced in parallel with the group dialogue, students have an opportunity to generate feedback not only about the quality of what they produced individually beforehand (content feedback) but also on how the work is produced (process feedback).

This view of peer dialogue as a source of comparison information is somewhat different from, although not incompatible with, the way that dialogue is conceived in feedback research (Ajjawi and Boud 2017; Nicol 2010; Zhu and Carless 2018). Feedback researchers conceive peer dialogue as the means by which students fill gaps in each other's knowledge, co-create meaning by discussing, debating and articulating their thinking. Askell-Williams and Lawson (2005) found, however, that one perception that students have of the benefits of peer dialogue is that it creates comparison opportunities through which they acquire new perspectives on their own work, new ways of doing things and through which they generate ideas they had not thought of before. This aligns with other literature where it has been argued that dialogue occurs at two levels - inner dialogue (talking with oneself) and external dialogue (talking with others) - and that it is their co-existence that leads to new possibilities for thought and action (Hetherington and Wegerif 2018; Tsang 2007).

Research questions

Based on the study by Nicol and McCallum (2021) and the arguments above, the overall prediction was that by making explicit comparisons at both stages of the two-stage exam, individual and group, students would generate significant and valuable self-feedback. Also, we anticipated that students would connect the feedback they generate at the individual stage with that which they generate at the group stage. For example, if they identified gaps in their understanding at the individual stage, we would expect them to use that information for comparison at the group stage to address those gaps. In addition, given the nature of the comparators at the group stage - the ongoing dialogue and the emerging and final group output – another expectation was that students would generate both process and subject content feedback at the group stage.

Of specific interest in this study was how the feedback generated by students would compare against the comments given by the teacher. Answers to this question are important in terms of helping teachers identify what feedback they can best provide and helping develop students' self-regulatory feedback capability. In this study, the teacher's comments were only available to students a few weeks after the two-stage exam. Prior research on peer review by Nicol and McCallum (2021) led us to believe that students' own feedback commentary would differ from the teacher's feedback comments, yet what those differences might comprise in two-stage exams was difficult to predict. A further area of interest was in how students' written accounts of their internal feedback would compare with what they talked about in the group discussion. This is an un-researched area as it taps into the differences between internal and external dialogue, so this aspect was exploratory. In sum, the following are the three broad research questions.

RQ1: What kinds of feedback do students self-generate at the individual and group stage of two-stage exams and how do they interconnect the feedback across these two-stages?

RQ2: How does student-generated feedback during two stage exams compare with the feedback provided by the teacher?

RQ3: How does the feedback that students generate compare with what they actually say in the communicative exchanges they engage in during their group discussions?

Method

Participants

The students in this study were enrolled in an honours level optional course - Economics of Poverty, Discrimination and Development - at a Scottish university. There were 24 students and the course comprised 10 lectures of two hours each over 10 weeks. Only 23 students were included in the data analysis because one student did not adhere to the task requirements. Ethical approval to carry out this study was provided by the College of Social Sciences Ethics Committee for Non-Clinical Research Involving Human Subjects at the University where this study took place [Reference number 400160193].

Preparation for the two-stage exam

To prepare students for the two-stage exam format, each two-hour lecture in the first 5 weeks of the course involved around 15 minutes dedicated to discussion of questions using a two-stage format. Students were presented with a question during the lecture and given time to consider the answer on their own, and to write down some notes about it. Then they were asked to discuss the same question again in small groups. After this, the teacher led a whole class discussion of the answer with students being asked to share their thoughts.

In week 6 of the semester students participated in the two-stage exam that forms the basis of this article. They were asked to prepare for a class test covering the material discussed in the first five weeks of lectures. This was a summative assessment making up 15% of the total grade for the whole course over two hours of dedicated time. The question was: *Evaluate the weaknesses of the multidimensional poverty index and suggest how it might be improved, noting also any limitations in the suggestions you make*. The students first wrote answers to the question on their own, they then answered the same question again in groups of three. The students were allowed to form their own groups.

Making internal feedback explicit: reflective questions

After producing their individual answer as well as after the group answer, students were asked to write answers to some reflective questions. The three submissions – the individual submission, the group submission and answers to the reflective question - were equally weighed in the grading. The reflective questions shown in Table 1 were the prompts used to make internal feedback explicit. Essentially, students were asked to consciously make the comparisons they would naturally have made anyway, at some level, and to explicate the learning resulting from them in writing.

Table 1. Reflective questions designed to make internal feedback explicit.

Reflective questions after individual stage
- If you are aware of any weaknesses in your answer, or if there are aspects of your answer that were still troubling you when you handed it in, could you explain what the issues were in a few sentences.
- If you could have talked to an expert while writing your answer what feedback would you have asked for (i.e. on what aspects of your work would you have asked for feedback)? Explain.

Reflective questions after group stage
- How did the group answer differ from your own answer? Note two differences and clearly explain what they were.
- What did you learn about your own answer from the differences mentioned above? Please explain.
- Overall, which was better the group answer or your own answer? Why do you think this? If you think they were both the same, explain why you think this.
- Did the group discussion make you aware of any weaknesses or strengths in your answer that you were not aware of at the time you wrote it? If yes, explain what you became aware of?
- Can you explain exactly how you would improve your answer now, if you had the opportunity to do so?

Timeline and research questions

The upper part of Figure 2 shows the timeline of the students' activities and the lower part shows how the data from the answers to the reflective questions was used to answer the three research questions.

Coding of students' feedback commentaries

The coding of the students' feedback commentaries was both deductive and inductive. A thematic analysis was first conducted using the teacher-feedback categorisation of Hattie and Timperley (2007). These researchers categorise teacher feedback comments as targeting four different levels: task, process, self-regulation and self-level. Task-level refers to comments about how well a task is performed. Process-level refers to comments about the processes and strategies students use to perform the task. Self-regulatory comments are about students' skills in monitoring, self-assessing and regulating their performance during the task. Self-level feedback refers to teacher's comments on the student as a person.

In doing this thematic analysis it became clear that the Hattie and Timperley (2007) categorisation only had a loose fit to our data. Hence, inductive analysis led us to settle on three categories of student-generated feedback comments: (i) content feedback (ii) feedback on the writing process and (iii) self-regulatory feedback. Table 2 provides definitions and scope for these three feedback categories. What distinguishes self-regulatory feedback from content and writing process feedback is that it is not specific to the current work that students produce but is generalisable to future work. Further analysis of the self-regulatory commentaries following an inductive process (Braun and Clarke 2006) to generate themes and codes led to the sub-categories: planning, writing process, different perspectives/approaches, wider reading, language skills, group skills.

In all coding, the unit of analysis was a phrase or sentence in the feedback commentary with a distinct meaning (Maguire and Delahunt 2017). Both authors independently coded the same samples, checked their coding interpretations, and then continued with another round
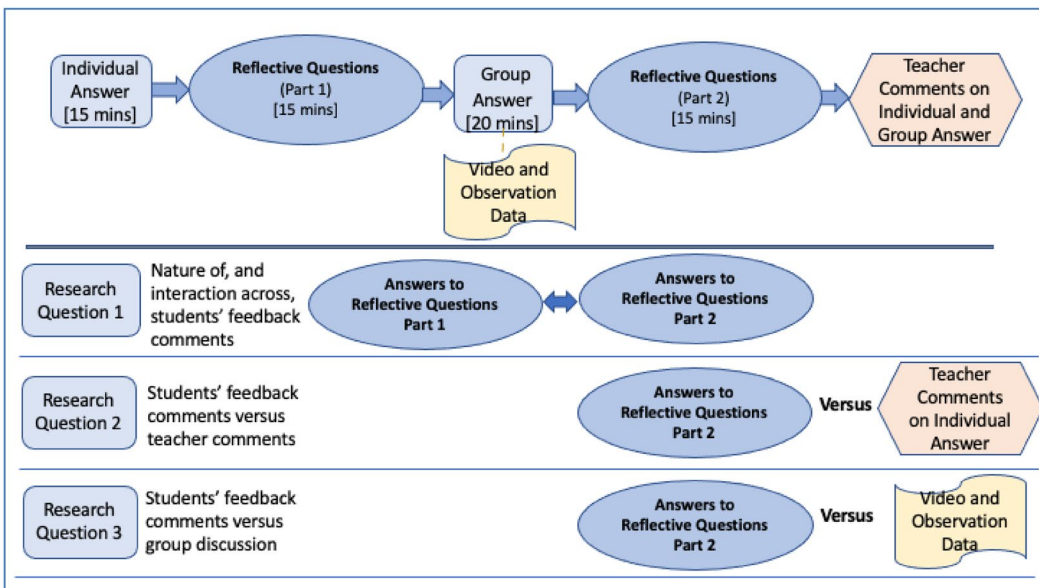

Figure 2. Timeline of two-stage exam how the research questions were addressed.

Table 2. Coding categories for student-generated feedback comments.

Category of internal feedback	Definition and scope
Content	Generates feedback about subject content, the correctness or accuracy, about strengths, weaknesses, errors, incorrect responses etc.
Writing Process	Generates feedback about quality of writing, its logic, structure, convincingness, balance or quality of argument
Self-regulatory	Generates feedback about own thinking, abilities, learning strategies (effectiveness or not), skills especially group working and writing insofar as this went beyond the immediate task and had a future orientation.

Table 3. Number, categorisation and examples of students' feedback comments after individual submission.

	No. comments	Mean (SD)	Examples
Content	31	1.35 (0.71)	I would have asked [the expert] the way the MPI was made to see whether the way the measure has been created may imply some issues. I made a suggestion about having more indicators for calculating MPI but did not know whether this would work.
Writing	32	1.39 (0.72)	…how to put my argument across so he understands what I am trying to say. I would have asked how to write my answer – should it be written by importance of weaknesses, should it include less or more points with less or more explanation. …not sure how convincing my argument was and how to present it well.
Self-regulation	4	0.17 (0.38)	I couldn't gather my thoughts before I started writing.
Total	**67**	**2.91 (1.25)**	

of coding. Once consistency and agreement had been established, the second author coded the remainder of the students' commentaries.

Results

Scope and inter-connectedness of student generated feedback (RQ1)

Feedback after individual stage
When students were asked to comment on the weaknesses they perceived in their own individual work after producing it, or to write down what they would ask an expert, 67 comments were produced (i.e. concerns were raised) by the 23 students (Mean = 2.91 and SD = 1.25). Table 3 presents the data with an example for each feedback category – content, writing and self-regulatory feedback. Notably, only a few students generated feedback on their strategies for self-regulation at the individual stage.

Feedback after group stage
When students compared their individual work against the group work, they generated feedback about how their answer could have been better in terms of its content and writing process. At this stage, unlike the individual stage, all students also generated self-regulatory feedback. Table 4 presents these findings for all three feedback categories, along with an example for the content and writing categories. Table 5 shows the thematic sub-division for the 53 self-regulatory feedback comments with examples.

Individual to group feedback mapping
The internal feedback conception predicts that students will use the group stage to address feedback issues they have identified at the individual stage. To examine this, we compared the feedback comments students generated after the individual stage with the comments they

Table 4. Number, categorisation and examples of students' content and writing feedback after the group submission.

	No. comments	Mean (SD)	Examples
Content	45	1.96 (0.95)	I learned that there was more to add to the improvements I had already come up with and that there are many more flaws with the MPI than I originally thought.
Writing	46	1.96 (0.81)	I learned that my own answer was fairly scruffy and badly structured.
			[I learned] a better way to structure my answer and present my ideas because I was really impressed by the way we did it together… [in the group].
			Completeness: I didn't go to that depth as they did.
			To get the best answer, be specific on one point rather than coming across many points but not in depth.
Self- regulation	53	2.35 (1.31)	See Table 5 for further analysis and examples.
Totals	**144**	**6.26 (1.77)**	

Table 5. Number, categorisation and examples of students' self-regulatory feedback comments after group submission.

Self-regulatory feedback categories	No.	Examples
Writing Process	14	We discussed each point before it was written we clarified and fully answered each point, whereas when I am writing my own answers I may speed through without completely explaining what I mean.
Different Perspectives/ approaches	13	I learned to always explore other perspectives and evaluate every aspect of an answer or problem.
Group skills/ benefits and importance	11	I think I learnt how valuable it is to talk about your ideas with others. I could apply this idea to essay writing too.
Language Skills	7	…weakness of my writing maybe the language, perhaps my vocabulary is not as diverse as others my age.
Planning	5	In exams it is difficult to have the time to plan the answer. So, I often write what comes to my mind as I am writing. This can make the point I want to get across difficult to do so, even though I may know the material well.
Wider reading	3	I would probably include some alternative outside hypothesis other than Amartya Sen. Therefore, engage more outside research and book reading.
Total	**53**	

generated after the group stage for the two categories content and writing. Table 6 shows a clear mapping of students' self-identified concerns and their insights at the group stage regarding how to resolve those concerns (i.e. their own feedback solutions). Of the 63 concerns raised after the individual stage, 48 were revisited and addressed at some level at the group stage.

Individual grade versus group grade

An analysis to ascertain how the quality of students' individual work compared against what their group produced confirmed that no individual student performed better than their group, although for a small percentage the individual grade did equal the group grade. The implication is that all students were comparing their work against work of a quality at least as good if not better than their own.

Feedback at group stage versus teacher comments (RQ2)

Feedback generated by students and teacher feedback in different feedback categories

All students produced significantly more feedback, in terms of number of words, after the group stage than the teacher produced (t-test, $p \leq 0.01$). More importantly, Table 7 shows students' and teacher feedback comments (i.e. units of meaning), in each of the three feedback categories. A

Table 6. Mapping of feedback concerns raised at the individual stage with students' elaborations at the group stage.

	Individual stage concern			Group stage solution
Category	Example	No.	No.	Examples
Content	…slightly frustrated by my lack of concrete information about other alternative measures to the health/ education/standard of living deprivation measures.	31	23	the group answer was more comprehensive and well-informed as we were able to fact-check with each other… considerations about how a generic poverty measure may not be as useful when applied to a broad spectrum of societies as social context and social norms should not be undervalued in how they impact the experience of and perspective on poverty, as I came to realise during the discussion part of the question.
Writing	I was not sure how in detail the problems and improvements should be described. I would ask if the explanation is in enough detail and if he [the expert] would write more critically or just like I did.	32	25	I need to be more critical, explain my answers in more detail and also try to build in relevant examples.
Totals		**63**	**48**	

Table 7. Feedback Comments: Teacher versus Students.

	Teacher Comments		Student Comments	
	No.	Mean (SD)	No.	Mean (SD)
Content	41	1.78 (0.51)	45	1.96 (0.95)
Writing	30	1.31 (0.46)	46	2.00 (0.83)
Self-regulation	0	0	53	2.30 (1. 30)
Total	**71**	**3.09 (0.65)**	**144**	**6.26 (1.75)**

two-way analysis of variance (ANOVA) was carried out on this data with source of feedback (teacher or student) and category of feedback (content, writing process and self-regulation) as the independent variables and units of meaning as the dependent variable. This revealed a significant difference between the mean number of comments given by the teacher and students ($p \leq 0.01$), as well as an interaction between source and category of feedback comments ($p \leq 0.01$).

Further analysis confirmed that students produced significantly more writing process and self-regulatory feedback than the teacher (t-test, $p \leq 0.01$), but there was no significant difference in content feedback between students and teacher. The most noteworthy finding was that all students generated self-regulatory feedback comments while the teacher generated no comments of this kind.

Qualitative differences between teacher and students' self-generated feedback

Further analysis of teacher comments for the content and writing categories against students' self-generated comments shows that invariably students' comments were more elaborate and specific than the teacher's comments. Table 8 provides examples of what 'more elaborate' might entail for different categories of comments. While the teacher gave general comments about the strengths and areas in need of improvement, the students were more likely to state exactly how the improvements could be made.

Students' feedback commentary versus group discussion (RQ3)

Due to video malfunction, there was no permanent record of what students talked about in the group discussion. Hence, we were not able to directly compare the students' commentary

Table 8. Teacher Feedback compared with Students' Feedback commentary.

Category	Examples of comments	
	Teacher	Student
Detailed and specific understanding of how to improve on the content.	Few very good points, scope for many more.	I would add the importance of social needs like the access to clean water, proper food, electricity and so on. I think that those have to have more weight in the measure, rather than having a bike or a motorcycle. Also, I would like to explain better the differences between the countries.
	Scope for some more thinking and more effective discussion.	The group answer included even more weaknesses. MPI only focuses on health, education and objects when it comes to the standard of living and fails to include aspects of democracy, freedom of speech, freedom to do what you wish, opportunities and social exclusion. These are all aspects that play a more and more important role when it comes to determining how poor people are.
Realisation of misunderstandings and flaws.	The discussion should have been more about the make-up of the index itself. More scope for discussion.	I would remove my comments about people being off the grid - they would make up a really small percentage. The mentally ill with eating disorders should be counted as unhealthy, no matter how wealthy they are. I would add points about liberty, freedom, mental health and happiness.
Learnt how arguments could be stronger, explanations and examples.	Reasonably clear and effective (i.e. room to be even better).	It was also more eloquently written as we collaborated to come up with the best wording and way to describe, was much more in depth. I learned how to express my ideas more concisely. I would try to explain myself better.
	Should explain/justify more effectively.	The main difference was the in-depth answers with clear explanations and justification whereas I wrote important points and short descriptions. I need to be more critical, explain my answers more and also try to build in relevant examples.
Learnt a better way to structure and present the answer	You have introduced the discussion effectively. But you have to explain the points you make in more detail.	I did not come up with the structure my teammates did. I didn't go to that depth as they did. To get the best answer, be specific on one point rather than coming across many points but not in depth. Be specific and get into a point comprehensively rather than coming across many points in shallow.
	Since this is a poverty measure trying to measure depravation of education, what you write first seems more appropriate.	Structure could have been better. It also felt like a more mature answer. I wish I had included correlation, outdatedness of assets and mortality rate of curable diseases. These were all things I had thought about but maybe couldn't think how to write them in the most effective way.

on their internal feedback against the external feedback commentary. However, one of the authors did listen to the discussion of one group and took notes. The most notable and incontrovertible finding from this observation was that although all students wrote self-regulatory feedback comments this was not a subject of discussion within the observed group. Specifically, no student in the observed group discussed their own thinking processes, about the value of group working or about what they would do to improve their own exam preparation in future.

Discussion

Most researchers maintain that the feedback dialogue that occurs during the group stage in two-stage exams is the main mechanism by which students learn (Gilley and Clarkston 2014; Levy, Svoronos, and Klinger 2018). Nicol (2019; 2020) instead claims that feedback is primarily an internal process even though its generation is invariably fuelled by the comparisons that students make of their own work against external information. To explore this latter interpretation, a study was conducted wherein students were required to make their internal feedback explicit during a two-stage exam. This study provides new insights into learning from this two-stage sequence,

about the feedback that students are able to self-generate, about how that feedback compares with teacher comments, and about how it relates to the external dialogue that helps trigger it.

Scope and inter-connectedness of student-generated feedback [RQ1]

All students generated content, writing process and self-regulatory feedback during this two-stage exam. Of special note was the detailed process feedback and the large number of self-regulatory feedback comments. That students can generate such feedback by themselves has not been evidenced in this way before. The internal feedback students generated was also shown to be inter-connected across the two-stage sequence. Students generated feedback about weaknesses in their work at the individual stage and then they identified how to address those weaknesses at the group stage. This cyclic and interconnected nature of internal feedback processes from the student perspective has also not been the subject of research. Yet it has some important practice implications beyond two-stage exams. For example, by having students explicitly identify gaps in their work after producing it, teachers would be better able to select and make available suitable comparison information to help students close those gaps. Such information might come as comments, but it could also come from a textbook, a lecture presentation, a video or the work of peers.

Student-generated feedback versus teacher comments [RQ2]

Another finding in this implementation of two-stage exams was that the feedback students generated was more elaborate than the comments they received from the teacher. Even more important, whereas students produced a large number of self-regulatory feedback comments, the teacher produced almost no comments of this kind. Hattie and Timperley (2007) maintain that self-regulatory feedback is the gold standard for teacher feedback comments. Yet Hattie (2009), based on a meta-analysis of empirical research, found that 90% of comments from teachers were at the task level and that teachers rarely provide self-regulatory comments. Hence this finding that students can generate significant self-regulatory feedback unaided should be built on in other contexts.

With regard to teacher comments some caution is required lest the results above give the impression that these are sub-standard in comparison with student-generated comments. This cannot be inferred as we did not compare like with like. We did not ask students to compare and write down what they learned from teacher comments, so we do not actually know what they would generate from that comparison. It could also be argued that the results of this study would have been different if the teacher merely spent more time commenting. While this is undoubtedly true, it is not the whole story. As noted earlier, the self-regulatory feedback that students generated is rarely provided by teachers and will always be difficult for them to provide as only students have full access to their own thinking and planning processes. Hence, students are very likely always to write different comments, and more self-regulatory comments than their teachers.

The main message from this investigation is that there is significant merit in having students generate their own feedback using information other than teacher comments. This will increase students' engagement in feedback processes, attenuate teacher dominance associated with comments, and result in students generating more varied internal feedback. It is also more in line with the long-term goal of higher education - that students develop the capacity to self-regulate their own learning using information from multiple sources.

Student-generated feedback versus external dialogue [RQ3]

Despite technical issues with the video recording, the observational data indicate that the feedback students generate from the dialogue they engage in goes beyond the verbal exchanges

they have with each other. This can be inferred from the students producing a significant amount of self-regulatory written feedback comments but the members of the group that was observed did not make any such self-regulatory verbal comments. This shows that what feedback students derive from external dialogue goes beyond what is evident in the dialogue itself.

This view, that dialogue is a source of information for comparison and for internal feedback generation, goes beyond current conceptions of feedback dialogue (e.g. Nicol 2010; Zhu and Carless 2018). As such it opens up new possibilities for research and practice. One implication is that inner feedback will invariably be more powerful if students produce some work individually before engaging in a dialogue because this will prime their knowledge and generate initial feedback which will magnify the comparative effects of the subsequent dialogue. In current practice, students are often put in groups to discuss something without producing individual work beforehand. Group dialogue could be further improved by having students not only discuss and compare their work with that of peers but also at the same time compare these works against some other works or information provided by the teacher. For example, if the teacher also gave them a rubric or a pair of high-quality exemplars as comparators, this would broaden the scope of the dialogue and enhance the quality of the feedback that students generate about their own work from that dialogue. Dialogue is a universal comparator that can be used to amplify students' learning whenever they engage in other resource-based comparisons.

Making the results of the comparison explicit

The implementation of two stage-exams reported here differs from conventional implementations of this method. Students were required to externalise their internal feedback through written commentaries. Also, dedicated time was allowed for this and the students' feedback commentaries contributed to the grade. Hence, one might argue that the interpretation of feedback generation in this two-stage exam might not apply to its traditional counterpart.

In response, we contend that this method merely makes explicit the natural comparison processes that students already engage in during two-stage exams, but that by doing so it increases the power of the internal feedback that students generate. In other words, this is a method for improving learning from two-stage exams. Other researchers suggest that the two-stage method could be improved by preparing students better for group working and by paying attention to the group composition as this will affect the quality of participation and interaction (Efu 2019). This research suggests that a more direct method of improving learning is to make the feedback comparisons students implicitly engage in a conscious and deliberate act. This is easily implemented as it only requires that students go back and re-consider their individual answer in the light of the group answer and make explicit the learning that results.

Explicitness has other important benefits. When students see their own feedback productions this raises their awareness of how feedback is generated and how they might go about better generating it themselves without calling on a teacher. Explicitness also results in students' giving concrete expression to their thinking about abstract concepts. This both fosters learning and transfer and develops metacognitive knowledge and skills (Tanner 2012). From a teacher perspective, having students make their internal feedback explicit provides tangible data about where students have difficulties in their learning. Hence, teachers are better able to target their comments to students' needs or better still to select the next comparator to address those needs.

Further research

There is considerable scope for further research on this two-stage method. One limitation of this study was that technical problems meant that we were not able to compare directly what students said in their group with the written feedback commentary they generated. This avenue

of inquiry is worth pursuing in future research. A further limitation in this study was the small sample size and the application of this revised two-stage method in a single discipline, economics. While arguably this method would transfer easily to other disciplines, research in different disciplines and in other contexts is required to confirm this. Another consideration is that this revised two stage-method be trialled and researched in formative contexts where there is no grading. Lastly, as noted earlier, there is also a pressing need to investigate what feedback students generate from teacher comments and how that compares with what they generate from other information sources.

Conclusion

In recent research, Nicol (2019; 2020) takes the view that students are making feedback comparisons in all teaching and learning settings and that teachers should capitalise on this natural capacity. This investigation shows how this comparison process applies in two-stage exams and how this method can be made more effective by having students explicitly generate their own feedback. Importantly, this research helps move our thinking away from teacher comments as the main comparator to a scenario where the teacher's role is to identify and select a range of suitable comparators and to plan for their use by students. Possible comparators are numerous and might include videos, information in journal articles or textbooks, peer works, rubrics and observations of others' performance. The goal of the feedback endeavour is to help students develop their ability to make such comparisons by themselves, and to build their capacity to generate their own feedback from many information sources not just comments, so they are better able to plan, evaluate, develop and regulate their own learning.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on Contributors

Professor David Nicol leads the Teaching Excellence Initiative in the Adam Smith Business School, University of Glasgow. His current research makes a critical contribution to the study of how learners generate internal feedback in everyday activities, and how to help them harness this natural capacity for feedback agency in educational and professional settings.

Dr Geethanjali (Geetha) Selvaretnam is a Senior Lecturer at the Adam Smith Business School, University of Glasgow. She is currently carrying out research in development economics, on students' learning during inner feedback generation, peer support activities and multicultural interactions.

ORCID

David Nicol 🆔 http://orcid.org/0000-0001-7327-8028
Geethanjali Selvaretnam 🆔 http://orcid.org/0000-0002-7609-8388

References

Ajjawi, R., and D. Boud. 2017. "Researching Feedback Dialogue: An Interactional Analysis Approach." *Assessment & Evaluation in Higher Education* 42 (2): 252–265. doi:10.1080/02602938.2015.1102863.

Andrade, H. L. 2018. "Feedback in the Context of Self-Assessment." In *The Cambridge Handbook of Instructional Feedback*, edited by A.A. Lipnevich and J.K. Smith, 376–408. Cambridge: Cambridge University Press.

Askell-Williams, H., and M. J. Lawson. 2005. "Students' Knowledge about the Value of Discussions for Teaching and Learning." *Social Psychology of Education* 8 (1): 83–115. doi:10.1007/s11218-004-5489-2.

Bisra, Kiran, Qing Liu, John C. Nesbit, Farimah Salimi, and Philip H. Winne. 2018. "Inducing Self-Explanation: A Meta-Analysis." *Educational Psychology Review* 30 (3): 703–725. doi:10.1007/s10648-018-9434-x.

Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. doi:10.1191/1478088706qp063oa.

Butler, D. L., and P. H. Winne. 1995. "Feedback and Self-Regulated Learning: A Theoretical Synthesis." *Review of Educational Research* 65 (3): 245–281. doi:10.3102/00346543065003245.

Chiu, J. L., and M. T. H. Chi. 2014. "Supporting Self-Explanation in the Classroom." In *Applying Science of Learning in Education: Infusing Psychological Science into the Curriculum*, edited by V. A. Benassi, C. E. Overson, and C. M. Hakala, 91–103. Society for the Teaching of Psychology. http://teachpsych.org/ebooks/asle2014/index.php

Cho, Y. H., and K. Cho. 2011. "Peer Reviewers Learn from Giving Comments." *Instructional Science* 39 (5): 629–643. doi:10.1007/s11251-010-9146-1.

Cortright, R. N., H. L. Collins, D. W. Rodenbaugh, and S. E. Dicarlo. 2003. "Student Retention of Course Content is Improved by Collaborative-Group Testing." *Advances in Physiology Education* 27 (3): 102–108. doi:10.1152/advan.00041.2002.

Drouin, M. A. 2010. "Group-Based Formative Summative Assessment Relates to Improved Performance and Satisfaction." *Teaching of Psychology* 37 (2): 114–118. doi:10.1080/00986281003626706.

Efu, S. I. 2019. "Exams as Learning Tools: A Comparison of Traditional and Collaborative Assessment in Higher Education." *College Teaching* 67 (1): 73–82. doi:10.1080/87567555.2018.1531282.

Fonseca, B., and M. T. H. Chi. 2011. "The Self-Explanation Effect: A Constructive Learning Activity." In *Handbook of Research on Learning and Instruction* edited by R. E. Mayer, and P. A. Alexander, 296–321. New York, NY: Routledge Taylor and Francis Group.

Gilley, B. H., and B. Clarkston. 2014. "Collaborative Testing: Evidence of Learning in a Controlled in-Class Study of Undergraduate Students." *Journal of College Science Teaching* 043 (03): 83–91. https://www.jstor.org/stable/43632038. doi:10.2505/4/jcst14_043_03_83.

Hattie, J. 2009. *Visible Learning: A Synthesis of 800+ Meta-Analyses on Achievement*. London: Routledge.

Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. doi:10.3102/003465430298487.

Hetherington, L., and R. Wegerif. 2018. "Developing a Material-Dialogic Approach to Pedagogy to Guide Science Teacher Education." *Journal of Education for Teaching* 44 (1): 27–43. doi:10.1080/02607476.2018.1422611.

Huisman, B., N. Saab, J. van Driel, and P. van den Broek. 2018. "Peer Feedback on Academic Writing: Undergraduate Students' Peer Feedback Role, Peer Feedback Perceptions and Essay Performance." *Assessment & Evaluation in Higher Education* 43 (6): 955–968. doi:10.1080/02602938.2018.1424318.

Levy, D., T. Svoronos, and M. Klinger. 2018. "Two-Stage Examinations: Can Examinations be More Formative Experiences." *Active Learning in Higher Education*, 1–16. https://doi.org/10.1177/1469787418801668

Li, L., and V. Grion. 2019. "The Power of Giving Feedback Ad Receiving Feedback in. peer Assessment." *All Ireland Journal of Teaching and Learning in Higher Education* 11 (2): 1–17. https://ojs.aishe.org/index.php/aishe-j/article/view/413/671.

Li, Lan, Xiongyi Liu, and Allen L. Steckelberg. 2010. "Assessor or Assessee: How Student Learning Improves by Giving and Receiving Peer Feedback." *British Journal of Educational Technology* 41 (3): 525–536. doi:10.1111/j.1467-8535.2009.00968.x.

Lipnevich, A. A., L. N. McCallen, K. P. Miles, and J. K. Smith. 2014. "Mind the Gap! Students' Use of Exemplars and Detailed Rubrics as Formative Assessment." *Instructional Science* 42 (4): 539–559. doi:10.1007/s11251-013-9299-9.

Lundstrom, K., and W. Baker. 2009. "To Give is Better than to Receive: The Benefits of Peer Review to the Reviewer's Own Writing." *Journal of Second Language Writing* 18 (1): 30–43. doi:10.1016/j.jslw.2008.06.002.

Lusk, M., and L. Conklin. 2003. "Collaborative Testing to Promote Learning." *Journal of Nursing Education* 42 (3): 121–124. doi:10.3928/0148-4834-20030301-07.

Maguire, M., and B. Delahunt. 2017. "Doing a Thematic Analysis: A Practical, Step-by-Step Guide for Learning and Teaching Scholars." *All Ireland Journal of Higher Education* 9 (3): 3351–33514. http://ojs.aishe.org/index.php/aishe-j/article/view/335.

McConlogue, T. 2015. "Making Judgements: Investigating the Process of Composing and Receiving Peer Feedback." *Studies in Higher Education* 40 (9): 1495–1506. doi:10.1080/03075079.2013.868878.

Muir, S. P., and D. M. Tracy. 1999. "Collaborative Essay Testing: Just Try It!." *College Teaching* 47 (1): 33–35. doi:10.1080/87567559909596077.

Nicol, D. 2010. "From Monologue to Dialogue: Improving Written Feedback Processes in Mass Higher Education." *Assessment & Evaluation in Higher Education* 35 (5): 501–517. doi:10.1080/02602931003786559.

Nicol, D. 2014. "Guiding Principles of Peer Review: unlocking Learners' Evaluative Skills." In *Advances and Innovations in University Assessment and Feedback* edited by C. Kreber, C. Anderson, N. Entwistle and J. McArthur. Edinburgh: Edinburgh University Press doi:10.3366/edinburgh/9780748694549.003.0011.

Nicol, D. 2019. "Reconceptualising Feedback as an Internal Not an External Process." *Italian Journal of Educational Research Special Issue*, : 71–83. https://ojs.pensamultimedia.it/index.php/sird/article/view/3270.

Nicol, D. 2020. "The Power of Internal Feedback: Exploiting Natural Comparison Processes." *Assessment and Evaluation in Higher Education*. doi:10.1080/02602938.2020.1823314.

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. doi:10.1080/03075070600572090.

Nicol, D., A. Thomson, and C. Breslin. 2014. "Rethinking Feedback Practices in Higher Education: A Peer Review Perspective." *Assessment & Evaluation in Higher Education* 39 (1): 102–122. doi:10.1080/02602938.2013.795518.

Nicol, D., and S. McCallum. 2021. "Making Internal Feedback Explicit: Exploiting the Multiple Comparisons that Occur during Peer Review." Online first. doi:10.1080/02602938.2021.1924620.

Panadero, E., Broadbent, J.D. Boud, and J. M. Lodge. 2019. "Using Formative Assessment to Influence Self- and Co-Regulated Learning: The Role of Evaluative Judgement." *European Journal of Psychology of Education* 34 (3): 535–557. doi:10.1007/s10212-018-0407-8.

Pintrich, P. R. 2002. "The Role of Metacognitive Knowledge in Learning, Teaching, and Assessing." *Theory into Practice* 41 (4): 219–225. doi:10.1207/s15430421tip4104_3.

Rieger, G. W., and C. E. Heiner. 2014. "Examinations That Support Collaborative Learning: The Students' Perspective." *Journal of College Science Teaching* 043 (04): 41–47. doi:10.2505/4/jcst14_043_04_41.

Shindler, J. V. 2004. "Greater than the Sum of the Parts? Examining the Soundness of Collaborative Exams in Teacher Education Courses." *Innovative Higher Education* 28 (4): 273–293. doi:10.1023/B:IHIE.0000018910.08228.39.

Tanner, K. D. 2012. "Promoting Student Metacognition." *CBE Life Sciences Education* 11 (2): 113–120. doi:10.1187/cbe.12-03-0033.

Tsang, N. M. 2007. "Reflection as Dialogue." *British Journal of Social Work* 37 (4): 681–694. http://www.jstor.org/stable/23722589. doi:10.1093/bjsw/bch304.

Yan, Z. 2020. "Self-Assessment in the Process of Self-Regulated Learning and Its Relationship with Academic Achievement." *Assessment & Evaluation in Higher Education* 45 (2): 224–238. doi:10.1080/02602938.2019.1629390.

Zhu, Q., and D. Carless. 2018. "Dialogue within Peer Feedback Processes: Clarification and Negotiation of Meaning." *Higher Education Research & Development* 37 (4): 883–897. doi:10.1080/07294360.2018.1446417.

Zipp, J. F. 2007. "Learning by Exams: The Impact of Two- Stage Cooperative Tests." *Teaching Sociology* 35 (1): 62–76. doi:10.1177/0092055X0703500105.