“How are we doing?”

The use of learning analytics to support an ergative approach to assessment

Project report: Abridged version

Andrew Whitworth, Manchester Institute of Education
Dina Soliman, Ontario Institute for Studies in Education
Zhuoran You, Manchester Institute of Education

Corresponding author: drew.whitworth@manchester.ac.uk

July 2024

Table of contents

Executive summary.

- This document reports on a Flexible Learning Pilot, conducted in the 2023-24 academic year, that studied and evaluated a new paradigm of assessment, based around the principles of **knowledge-building**.

- This new paradigm is termed **ergative assessment**, being based around the marking of student *work*, as well as (not instead of) the *words* that are the product of this knowledge-building process. Observing the process is akin to students' 'showing their working out', making visible the connections between the process and product, including whether the students are misusing generative AI tools.

- The ergative assessment approach turns a course unit into what Schön has termed a *professional practicum*: a 'safe' space in which students can practice engaging in discourses and making judgments in ways that anticipate those taking place in the professional settings into which they will graduate. These discourses, being recorded on a digital discussion platform, are then available for reflection. Through this dialogue, **students learn to engage not just in professional discourse, but academic too, developing their ability to make 'good moves' in both settings**.

- Ergative assessment is therefore a **dialogue**, one in which students seek — and deserve — answers to the question "how are we doing?" *before* submission of the final output, thus allowing for intervention through feedback.

- A potential risk, and retardant, for the ergative approach is the perception that creating multiple points of observation, with feedback given at each point, will increase staff and student workload to an undesirable extent.

- The Pilot study reviewed this factor and concluded that there was **no significant impact on workload** as long as two additional steps were implemented alongside the revised assessment:
    1. The incorporation of **learning analytics** into the course and assessment design, analytics which are not just deployed in top-down surveillance of the cohort by the course unit lead and tutors, but are also available to students, who can use these data and associated visualisations to reflect on and modify their own practice;
    2. **A more creative approach to the use of 'contact hours' and teaching space**, replacing some lectures in the later part of the syllabus with knowledge-building work, using online tutorials and bringing distance learners into the same spaces as on-campus in a 'hyflex' mode.

- The ergative assessment used on the studied course involved students working in groups of 5-7 on two, two-week discussion activities that served as professional practica. In these they were expected to consult academic literature, make judgments about its relevance to the problem that they were tasked with solving, and work together to produce a series of co-created outputs that illustrated their ability to make informed judgments about this professional setting.

- Data drawn from an extensive analysis of the discussions of 8 groups show how, in activity 1, there was a **cognitive disconnect between process and product; thus, between the 'work' and 'words'**. This was evident in various ways, including a lack of contextualisation in many outputs, an excessive number of citations in outputs generally, and a lack of discussion of their relevance in the preceding discussion. These disconnects could not have been revealed through looking only at the output.

- The multiple points of observation also reveal where some groups did put in good knowledge-building work around certain inputs, but in their outputs, this good work was effaced by generic and bland summaries of the paper.

- Whether these negative effects were the result of the misuse of generative AI or not, the ergative approach allows for **timely feedback to be given in a form which had a measurable and significant impact** on both the process and the products of the second activity. Most groups paid significantly more attention to cited texts; made more informed judgments of relevance; and contextualised outputs more effectively.

- The ergative approach had **no evident negative impact on student satisfaction**. Indeed, feedback was mostly positive, and students reported their own use of analytics to reflect on and improve the quality of their work and their identification as competent actors in this setting.

- Tutors used learning analytics to prepare more effectively for tutorials. A more effective pre-course induction was desired, but this will be addressed in subsequent iterations of the course unit, and lessons learned from the Pilot will help with the integration of the approach into other settings.

- In conclusion, **ergative assessment is effective in assuring assessors that the students have engaged in a knowledge-building process during a course unit (or programme), and that the outputs that are being read and graded are the outcome of this process**. It can reveal where disconnections have emerged between process and product that may be the result of the misuse of generative AI.

- It can also help the course unit leader, the tutors, and the students themselves facilitate this knowledge-building process, and become competent practitioners who can bring to bear a range of digital collaborative environments tools to solve problems: thus, **enhance their digital practice and develop '21st century skills'**.

- Recommendations:

1) The FLP should help **leaders of these courses to self-identify** in order to develop a 'toolbox' of ergative approaches.
2) **Staff development resources** can be developed.
3) **Integrate an understanding of ergative assessment into course unit reviews.** Not *every* course unit need take this approach, but see recommendation 5.
4) **Make room in timetables for the work required**, by substituting online knowledge-building work for lectures in later weeks of courses. Along the way this will help relieve the intense pressure on teaching space on campus.
5) **Plan pedagogy at programme level.** Ensuring that students are engaged with the approach at least once in each semester of study seems sensible.
6) A **scoping review** is needed, looking at available discussion platforms (including, but not limited to, what will be available on Canvas) and assessing how the features and functionality of each might support the ergative approach.

1. Introduction

1.1. Project summary

This report is the principal output from a Flexible Learning Pilot (ID34), *Incorporating Knowledge-building Analytics* (IKeA), that ran in the Manchester Institute of Education in the 2023-24 academic year.

The pilot's primary focus, as its title implies, has been **learning analytics**, and how these might support a move to a different paradigm of assessment, the **ergative** approach. In the ergative approach to assessment, rather than being a "distinct artefact" that appears only at the end of a course, assessment is re-conceived as embedded throughout the learning process (Cope, Kalantzis and Searsmith 2020), and oriented around **the grading of work, not words**. In order for ergative assessment to be possible, students need to be fully present in a **knowledge building** process (Scardamalia 2002) of which the decision to award a particular grade has also taken account. An ergative assessment process can therefore be defined as one in which **the focus is on the knowledge building process, as well as (but not at the expense of) the product that represents that process**, whether that is a written assessment, a presentation or some other output.

To give work this visibility requires **multiple points of observation** within the course environment. This makes ergative assessment processual, and therefore, **dialogic**, and as part of this dialogue students demand, and deserve, answers along the way to the basic question, 'How am I doing?'.

Two issues arise, however. First, that the observations should not just be for the purposes of top-down 'surveillance' from above, meaning, available only to the course unit leader and the institution itself. The observation points exist as much for students to self-scrutinise, and to make judgments about the work of their group: "How are ***we*** doing?" Feedback can consequently be given at the micro-level (the individual student); the meso-level (the small group); and the macro-level (the cohort as a whole), and different analytics may come into play depending on the level.

Second, **workload** must be attended to. Any proposed changes to the assessment paradigm will quickly be sidelined if they were to lead to noticeable increases in the workload of staff or students. Two aspects of the Pilot addressed this issue: first, the course design (see §2.2) and its use of team teaching; second, the integration of digital technologies (and associated teaching and learning processes) into the assessment process.

Therefore, matters of principal interest in this pilot include:

* The overall **design of an ergative assessment and its integration into a course**: what activities it is built around, how the teaching timetable is adjusted to 'fit', the different roles of tutors (TAs) and the course unit lead.

* The **generation of learning analytics** that provide tutor and leader with data on progress and semi-automate the feedback process, without loss of quality.

* Evaluating **the impact of the course design and analytics on the student experience**, including quality of work and reflexivity.

* Evaluating **the impact on the experience and workload of teaching staff (**tutors and the course unit leader).

1.2. Credits and outputs

What you are reading now is the **abridged version** of this report. Mostly, what has been removed from the **full version** are references to the academic and theoretical background of the work and its methodology. Details of the case study course have also been mostly removed.

See the full version of the report for project credits.

As well as these two reports, the project will also produce resources oriented at teachers and learning technologists wishing to develop ergative assessment environments and to draw on learning analytics. These will be developed and released during semester 1 of the 2024-25 academic year.

2. Course design

2.1 General principles: ergative assessment and knowledge building

An ergative assessment approach is oriented around observations of the knowledge-building process as well as of the final product that represents this process. The shift is therefore from assessing and grading *words* to assessing and grading *work* that is staged and carried out over a period of time.

To. give a focus to the knowledge-building work, and a goal towards which students can help each other, setting some kind of group design task is a common approach taken in knowledge-building classrooms (e.g. Lin, Hong and Chai 2020). This design task becomes the 'work' that underpins the ergative assessment. The setting can be one that, in the context of any particular course unit, reflects some aspect of the 'authentic' professional setting for which the course is preparatory. Schön (1983) calls this kind of learning space a *professional practicum*.

Messick (1994, 16) states that we need to ensure that ergative assessments, oriented to this end, have *validity*, and suggests that:

> *…not only should students know what is being assessed, but the criteria and standards of what constitutes good performance should be clear to them in terms of both how the performance is to be scored and what steps might be taken or what directions moved in to improve performance.*

Ergative assessment has to be dialogic, therefore — there would be no point waiting until the end to make the students aware of these criteria; nor to advise them what steps might be taken to improve their performance, if they then had no subsequent opportunity to follow this advice. And learning analytics provide one way of offering the clarity that Messick calls for: ways of illustrating, as clearly as possible, the 'criteria and standards of what constitutes good performance' in this setting, and allowing all involved in the course — leader, tutors and students alike — to have ways of answering the questions: "How am I doing?" and "How are we doing?"

2.2 A specific case

2.2.1. Design of the course and its ergative assessment

The core principles described so far — ergative assessment and knowledge-building — are applicable across subjects and disciplines: this section describes how the principles were interpreted in one specific case.

The *Educational Technology and Communication* (hereafter, ETC) course unit is a 30-credit core course on the MA: Digital Technologies, Communication and Education, one of the portfolio of Master's level programmes run in the Manchester Institute of Education. The *professional setting* into which the course, and programme, are intended to induct students is, most broadly, that of the teacher, or learning technologist, able to make informed judgments about how to incorporate technology (digital or otherwise) into learning environments. The *work* that they are therefore expected to engage in is to make judgments about the appropriateness (or inappropriateness) of particular technologies to address teachers' and/or learners' needs in particular learning environments, whether formal or informal. The overall learning environment is therefore intended as a professional practicum in which students can practice, and thus gain experience with, making these kinds of judgments.

This assessment involves three activities (with four outputs). Two of these are group discussion activities (combined, worth 50% of the final grade), the third is individual (the other 50%). They are expected to reference relevant academic and professional literature in each of these outputs.

1) A role-playing simulation. Groups are presented with a scenario that describes an educational problem faced by a fictitious HE institution. Each group plays the role of a particular stakeholder (managers; innovators; academics; or students) and is asked to produce two outputs over a two-week period, with a total word count of 2,500 words. The first output is a statement about the group's present position on the problem, the second a proposal for future action. In the second output, groups are expected to respond to the statements of other groups. This activity will be referred to as **activity 1** or the **Social Shaping Game (SSG)**.
2) A design task, in which groups are tasked with visiting a museum, evaluating it as a learning environment, and then proposing a design for some kind of digital technology intervention that would enhance this environment in some respect. The output is a poster, which illustrates the museum, the design, and contains around 500 words of text, plus references. This activity will be referred to as **activity 2** or the **Museum** activity.
3) A self-evaluation document (hereafter, **SED**), in which each student individually reflects on some aspect of their use of technology to support their learning.

Activities 1 and 2, by their nature, cannot be marked anonymously, but activity 3 can be, and is. Thus, the course adopts a 'hybrid' approach to anonymous marking, one that remains within university guidelines[1].

2.2.2 Course timeline
See the full version of the report.

2.2.3. Significant features of the Knowledge Forum platform
The discussion environment used in the course in 2023-24 was Knowledge Forum[2] (hereafter, KF), a computer-supported collaborative learning platform developed by Professors Carl Bereiter and Marlene Scardamalia at the University of Toronto. KF is designed around functionalities and features that specifically support the knowledge-building principles outlined in their work (e.g. Scardamalia 2002, Bereiter and Scardamalia 2016).

Although KF was an integral part of this project, the principles and practices we are exploring in this report do not *depend* on this particular technology. However, certain features of KF do make it easier to generate some of the data on which we base our analysis: and on which the learning analytics are therefore based. Please refer to the full version of this report for details of these features.

The recommendations (§8) consider how and whether some of these features, and the analytics they help generate, are available in other discussion platforms, or how easily they might be developed for these platforms.

[1] §4.5 of the University Assessment Policy states: "All formal written assessments must be marked anonymously. However, it is recognised that this may not be possible for all assessments, particularly assessed performances, presentations or Objective Structured Clinical Examinations (OSCEs)."

[2] The course still makes significant use of Blackboard: all teaching materials, set readings, assessment specifications etc are uploaded there. Blackboard is the medium through which the course unit leader talks to the students in a top-down, monologic way. KF is the medium through which the students talk to each other, and the course unit leader, tutors and students talk in a more interactive and dialogic way.

3. Activity 1: evidence of cognitive disconnection

Activity 1 involves the groups creating two outputs. Across both they are expected to demonstrate a basic understanding of how decisions might be taken regarding technology use in a particular setting. They should be justifying their statements based on the academic literature (including some set texts); specific details from the scenario with which they were provided; and, in the second output, the stated positions of other groups in their set. Effective work in this practicum involves a group considering the relevance of their sources vis-a-vis this context (the scenario), establishing priorities, and communicating these to the other stakeholders in the 'game' via their collaboratively written outputs. The different processes involved in this assessment allow for various analytics to be generated and analysed, as this chapter will explore.

3.1. Use of set and found texts

This is a useful analytic at both the meso-level, that is, are groups devoting cognitive work to the set texts and thus the core ideas of the course?; and at the macro-level, that is, is the cohort as a whole engaging with these ideas? Set texts are also a means by which the group discussions can be *contextualised* — that is, placed in a specific course context and the 'information landscape' that students are being expected to navigate in that context.

Groups had been introduced to six set texts by this point. The relevance of the set texts was made clear, and it was expected that all or most of these would feature in each group's *discussion* at some point, even if the group decided, in the end, not to *cite* every one in their outputs. In fact, use of some references[3] was limited within the eight sample groups, as table 3.1.1 shows:

Text	*Discussed* by how many groups (/8)?	*Cited* by how many groups?	Average # notes referencing the text	Average # scaffolds referencing the text
Allen	4	2	1.75	0.75
Coen	3	2	0.66	0.33
Kubrick	5	3	3.4	2.4
Lupino	5	3	2.8	2.2
Pakula	7	5	3	1.57
Scorsese	6	4	5.17	2.5

Table 3.1.1: Use of set texts in activity 1

Students did respond better to the expectation to draw on *found* texts. The discursive moves encouraged here are that students show their ability (individually and within groups) to not only offer information to the group, but to then make *judgments of relevance*, and as a consequence, identify further information needs, thus continually reorganising the group's knowledge-building space and, ideally, optimising it in terms of how these resources feed into the group's final output. Quantitative measures show how in terms of content, most groups were giving themselves *too much* to work with.

[3] To avoid distraction and keep the observations as non-course specific as possible, the names of the authors of these works have been disguised throughout.

The activity 1 data for the 8 sampled groups is as shown in table 3.1.2. The green, amber and red highlights are examples of suggested 'flags': hopefully their use is self-explanatory at this point. Their implications for the ergative assessment approach will be explored further in §3.5.

Table 3.1.2: Use of all citations in discussions and in outputs

A	B	C	D	F	G	H	J	K	L
Group	**No. of members**	**No. of discussed citations**	**Number of used citations (in output)**	**Citations appearing for 1st time in final output**	**OUTPUT citations mean no. of notes**	**UNUSED citations mean no. of notes**	**OUTPUT citations, no. of scaffolds**	**UNUSED citations, mean no. of scaffolds**	**Total of scaffolds used per member**
Amethyst	7	61	25	4	2.36	2.027	1.2	0.389	6.29
Cherry	6	66	26	5	2.148	1.225	1.815	0.825	13.67
Gold	6	51	27	14	1.291	1.296	0.704	0.625	5.67
Jupiter	6	41	15	1	5.133	1.462	1.933	0.731	8.00
Purple	6	42	28	6	1.607	1.571	0.607	0.714	4.50
Red	7	23	11	1	2.727	1.75	3.364	1.167	7.29
Rose	7	29	16	10	3.25	2.154	0.938	1.615	5.14
Sapphire	7	42	27	6	1.926	1.067	0.148	0	0.57

In brief:

* Most of the groups are supplying themselves with a far greater number of sources to work with than is optimal. (For a definition of the 'optimal' number see the full version of the report.)
* Column F shows how many of the works cited in the output were *not discussed at any point* in the discussion which preceded the publication of the output. For Gold and Rose this amounted to over half of their bibliographies. Gold also have a very low number of notes devoted to each *cited* work (in fact, they are the only group to devote more discursive attention to unused citations, though the difference is not a significant one).
* Scaffold usage varies considerably between groups and Sapphire (§3.2) use only four in the whole course of their two-week discussion. Amber figures in the table indicate groups who could certainly have done more in this respect. Cherry try very hard to articulate the discursive moves they are making (see their column L), but with so many papers in the frame, their cognitive work is too thinly spread.

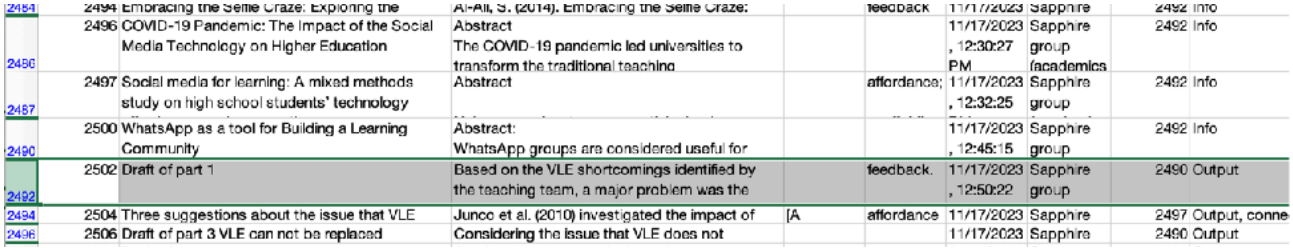
Let us now consider some groups in more detail, looking at the structure of their discussions as well as the content.

3.2. A struggling group

Sapphire received the lowest grades of all 20 groups (not just the 8 in the main sample). The deficiencies are evident in the product, but also in the process too. In the product, no set readings are cited at all in either of the two outputs, nor, at any point, in their discussion (they were the only group of the 8 for which this was the case). And beyond a few words here and there, the scenario is almost completely unreferenced in the outputs. Both these points reflect that **the output is largely unreflective of the course context** (see also §7).

These factors would have resulted in a low mark on their own, but examining the process as well confirms that a fail grade was appropriate for the group as a whole. It also allows us to address the question: *why* did this group produce such a poor output — and what feedback can they be given that highlights issues with both process and product, to encourage them to adopt different practices in activity 2?

A common pattern is for abstracts to cited texts to be posted *after* the reference has appeared in a block of suggested text. Here, a group member posts three abstracts (notes 2496-2500[4]) a few minutes before, in 2502, uploading text for the final output that contains references to those three cited works.



2494	Embracing the Selfie Craze: Exploring the	Al-Ali, S. (2014). Embracing the Selfie Craze:		feedback	11/17/2023	Sapphire	2492	Info
2496	COVID-19 Pandemic: The Impact of the Social Media Technology on Higher Education	Abstract The COVID-19 pandemic led universities to transform the traditional teaching			11/17/2023 , 12:30:27 PM	Sapphire group (academics	2492	Info
2497	Social media for learning: A mixed methods study on high school students' technology	Abstract		affordance;	11/17/2023 , 12:32:25	Sapphire group	2492	Info
2500	WhatsApp as a tool for Building a Learning Community	Abstract: WhatsApp groups are considered useful for			11/17/2023 , 12:45:15	Sapphire group	2492	Info
2502	Draft of part 1	Based on the VLE shortcomings identified by the teaching team, a major problem was the		feedback.	11/17/2023 , 12:50:22	Sapphire group	2490	Output
2504	Three suggestions about the issue that VLE	Junco et al. (2010) investigated the impact of	[A	affordance	11/17/2023	Sapphire	2497	Output, conne
2506	Draft of part 3 VLE can not be replaced	Considering the issue that VLE does not			11/17/2023	Sapphire	2490	Output

Figure 3.2.2: Abstracts immediately before output

The group make virtually no use of scaffolds at all, with only 4 deployed in total across the whole group (0.57/member; the next-lowest figure in the sample was 4.5/member). This was mentioned in the group's feedback:

> *Your KF discussions do not give very much sense that any of you are engaging with the key ideas. At almost no point on KF are connections made by you between two or more notes — suggesting that you are not exploring the possible connections between ideas and theories, you are just focused on producing text. Scaffolds are barely used, by any of you, suggesting you are not thinking about what it is you are saying, and what its relevance is.*

In summary, this group evidently struggled with the work being asked of them in this activity. The advantage of the ergative approach is that the form of this struggle can be discerned not simply from the product, but also, scrutiny of the process. The analytics reveal both the extent of the deficiency as a whole, and the form it has taken. Having records of the discussion available for review has then allowed the course unit leader to offer specific guidance as to where the disconnections have occurred.

3.3 Lack of discussion of used citations

While Sapphire were the group that struggled the most, several other groups received marks and feedback that highlighted occasions where they had cited 'cosmetically' to an excessive extent in their outputs, listed examples, and asked them to reflect on just how much discussion they had given these works (i.e. none). Of the 8 sampled groups, Purple and Cherry also received feedback of this type.

However, the ergative model allows for more nuanced assessments to be made of the process and how it links to group products. A 'red flag' highlighted in table 3.1.2 (column F) is those groups who cited a large number of works *for the first time* in posted drafts of their final output, in other words, without any discussion of the relevance (or otherwise) of these during the process. Gold and Rose were particular offenders here, with more than half (14/27 and 10/16 respectively) of citations in the outputs appearing there for the first time.

More interesting still are cases where there has been good, *visible*, constructive use of authoritative sources, whether set or found texts, that has then been effaced by group outputs in which these sources were cited, but only in highly generic ways. In other words, that useful knowledge-building work evident in the process is not reflected in the product. This feedback was given to the

[4] 'Missing' numbers represent notes that were created by other groups.

Emerald group: not one of the 8 sampled groups but it mentions a clear example of the phenomenon so is worth quoting here:

> *You've worked adequately well to compile the statements, but it is much less obvious that, as a group, you've been discussing key ideas and making judgments about their relevance to the design problem you face. Only at one point do you cite a set reading ([Pakula], briefly, in the proposal) and this must change in activity 2. This despite some ideas from these readings being introduced in the discussion — to potentially useful effect! Take the post from [H] on 9th Nov, citing [Kubrick] and entitled 'key questions from our perspective' — these are good ideas, and useful; [R] validates this with a reply, and both posts show that you do, in fact, have a sense of the innovators' real position in this scenario — but all are simply passed by, and never picked up in the statements.*

Individual feedback then complements that given to the group, allowing differentiation between individual contributions, and permitting students to reflect on their work not just as individuals but with reference to the group they are in.

The feedback is therefore a *move in the overall course dialogue* , made here at the individual, micro-level (student H being prompted to reflect on why her 9th November post was effective, and good knowledge-building practice that she might repeat) and the group, macro-level. Had this point of observation not existed, this kind of reflection-on-action (Schön 1983) could not have been prompted.

3.4. Two better groups

Of the eight sampled groups, Red and Jupiter were the two where the attention they have given to the knowledge-building process is most evident. Some evidence for this was suggested above in §3.1 via the citations count. Generally their discourse is more elegant and efficient than other groups. Across the 8 sampled groups Red gave themselves the fewest number of sources (23) to work with, and cited the fewest number of sources in their outputs, going into each in more detail. In fact, for this analytic their total was the lowest across all 20 groups: yet they were also given the highest group grade for activity 1 (an A+ equivalent).

While not as elegant about it as Red, Jupiter pay a comparatively high level of attention to their cited texts. Of the 8 groups they have the highest value for the 'Notes per output citation' analytic, with 5.1, and the second highest for 'Scaffolds used per member' at 8.00.

It is not just the structure of their dialogue which is effective, however, but the way that these dialogic moves help the discussion evolve. Ideas from various 'outside' sources are incorporated into each group's their own 'discursive map' (see Webster and Whitworth 2017, Whitworth 2020).

Consider this snippet from another Red note: "Since I didn't understand the meaning of 'operational proximity', I read the article cited by the Orange Group…". Little more could be asked for here: they are following up on discussions in another group in their set and building on them with their own reading and then consideration of whether the idea is relevant to the group. The connection between cited works is sealed by a subsequent note from student K, which builds-on but also then redirects and connects with the other concept (knowledge-sharing hub) in the group's discursive map. This idea of a knowledge-sharing hub[5] came from the Purple group. Red do well (as do Jupiter) at using the ideas of other groups in their final output — and citing them properly and in full.

---

[5] Essentially, the proposal that the fictitious university in the SSG's scenario would benefit from setting up a kind of 'Learning Commons' for professional development, bringing together academics, innovators and students alike.

It's interesting that they pick up on the need for more detail, because Purple — one of the groups that struggled in this activity — did not involve themselves in any discussion about what the 'hub' phrase actually meant. What Red have now done (and whether because Purple used generative AI to produce their output or not) is *put in the knowledge-building work that Purple omitted* — they have incorporated the idea into their discursive map, and set off to critique and deepen it.

3.5. Summary: 'red flag moments'

When moderating activity 1, tutors are encouraged to watch out for groups, or individuals, that display low rates of participation. Those who are not engaging can be contacted and reminded that the activity is under way and its complexity demands two weeks' worth of attention (as opposed to all being done with two days to go). Low participation is therefore a traditional 'red flag': that is, an indicator of sub-optimal knowledge-building work. The assumption is that remedial action can be taken to 'kick' the process into a more effective gear, prior to the production of the output.

When students 'show their working' through the ergative approach, this allows other 'red flags' to be generated. A group might be engaged at an acceptable level, making an adequate number of posts, but still struggling when it comes to making certain moves in dialogue (such as, making judgments of relevance, engaging properly with set texts). For example, there was no reason to be concerned about Sapphire's work in terms of the number of posts made. Cherry also displayed a high rate of participation, yet still struggled thanks to having given themselves too much to work with.

Additional 'red flag moments' suggested by the analysis of activity 1 therefore include:

* Too many citations generally, whether in the discussion or the output; or, inadequate discussion of citations introduced early in a discussion, before new ones are thrown in.
* A high number of citations appearing for the first time in final outputs.
* Lack of scaffold usage.

In each case, the opposing tendency could be a 'green flag': good practice that can be encouraged and disseminated.

In summary, as the first significant move in this particular ergative assessment, activity 1 is a key point of observation. Students' constructive use of authoritative sources has been assessed not just through the appearance of citations in the words of the outputs but also their use of these cited sources in a collaborative knowledge-building process, and their ability to bring in ideas from outside the group and make them their own.

A further advantage of the ergative approach, however, is that these first moves in the dialogue have all taken place well before the end of the course. The teaching team are therefore able to respond prior to activity 2, intervening in the evolving assessment dialogue at macro-, meso- and micro-levels. How this intervention may have impacted on students' knowledge-building practiced is analysed in §4.

4. Activity 2: the next move in the dialogue

4.1. The role and content of feedback

Activity 2 plays a different role from activity 1. In it, students and groups are showing not that they can repeat what they did in the SSG but that they can *respond to feedback* by changing their performance — hopefully for the better.

Between activities 1 and 2, macro-level feedback is provided in a 90-minute online session. In this, the course unit leader highlights examples of good products and good processes. The meso- and micro-level feedback, to groups and individuals, is delivered through Blackboard. Provisional grades are included with this feedback. Therefore, following the *point of observation* that was the students' work on activity 1 and its product, feedback is offered that is *timely*, *specific* (by being linked to a provisional grade) and working at all three organisational levels of the dialogue.

In terms of general references to process, groups were reminded to give proper discursive attention in activity 2 to the set texts introduced in weeks 7 and 8; to not include these, or other found texts, as citations in their output without having discussed them first; to draw on scaffolds to articulate the nature of their contributions where they could.

4.2. Changes in the epistemic network in response to feedback

Students on ETC are therefore provided with feedback both intrinsic and extrinsic, working at meso-, macro- and micro-levels, and originating from various sources including the tutors, course unit leader, each other, and learning analytics. Taken as a whole, the various dialogic inputs and interactions in the ETC environment offer feedback with enough detail and persuasive force for there to be a noticeable and significant shift in discursive practices between activities 1 and 2. This shift can be revealed and visualised through measuring, depicting and analysing the *epistemic network*, allowing one to (Shaffer 2017, p. 384): "measure whether a student is talking the way successful students have talked in the past."

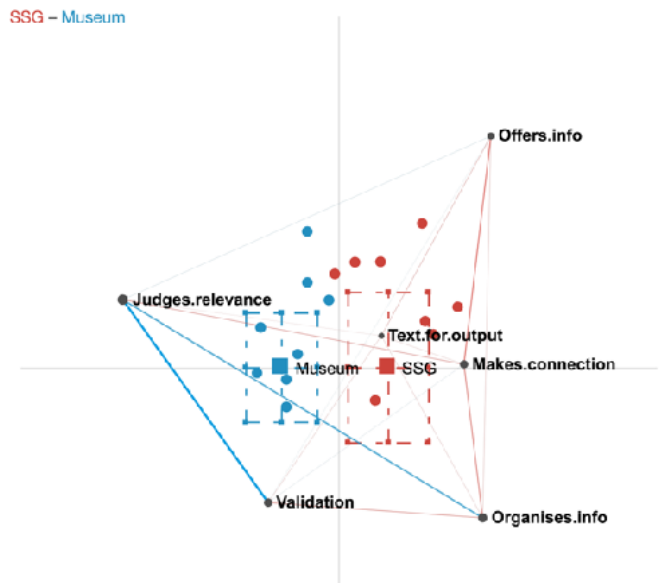


Figure 4.2.1 Epistemic Network Analysis: SSG v Museum for the 8 sampled groups

For details on how to interpret the ENA visualisation in figure 4.2.1, and the coding process which underpins the plot, see the full version of this report.

The ENA plot shows a significant difference between the two activities. In statistical terminology, along the X axis, a two sample *t* test, assuming unequal variance, showed that the centroid of the SSG (mean=0.13, SD=0.19) was statistically significantly different at the alpha=0.05 level from that of the Museum activity (mean=-0.15, SD=0.16, N=13; t(25.92)= 4.28, p=0.00, Cohen's d=1.60). As the relatively thick blue lines connecting the codes indicate, most of this difference is accounted for by all groups making significantly more *judgments about relevance*, and, particularly, connecting these with *validation* and *organising information*.

4.3. Citations and use of set texts

A direct comparison between the numbers of works cited in the Museum activity output compared to the outputs in the SSG is not justified, as the word count is much shorter (500 words compared to 2,500). What the quantitative data shows is that in the second activity, groups gave significantly more discursive attention than they did in the SSG to those citations that made it into the final output. Of the five set texts that students had been working with during block 3 of the course, two of them were heavily used by most groups.

Taking both these data and the ENA into account, the conclusion is that in activity 2 most students are devoting significantly more cognitive work to judging the relevance of cited papers, articulating their thinking using more scaffolds.

4.4 Better organised spaces

When highlighting good *process* in the online feedback session, the course unit lead drew attention to the view of the Bronze group[6], as their KF space for activity 1 was well and attractively organised. He noted that he had found it easy to navigate around[7]. And the 'beautification' of spaces was a practice that spread: compare figures 4.4.2 and 4.4.3, high-level views of the Copper group's spaces for activity 1 and 2. The change is not only one of better organisation but greatly increased aesthetic value, and we suggest this is significant in its own right, as it is a sign this group have decided to take on a sense of *ownership* of their space and devoted attention to it. While this is an aspect of knowledge-building practice that in some ways is specific to KF, it still suggests that the feedback session had a measurable impact. Copper also received the following comment in their group feedback:

> *…your view might have been a bit better organised visually. Take a look at some of the other views in your set — these are much easier to navigate.*

[6] Not one of our sample groups, though had we continued the 'deep reviews' they probably would have been number 9.

[7] Recall that the course unit lead does not moderate the discussion activities while they are taking place. Every group discussion marked in week 9 is come to 'fresh'.

Figure 4.4.2: Copper group's SSG view

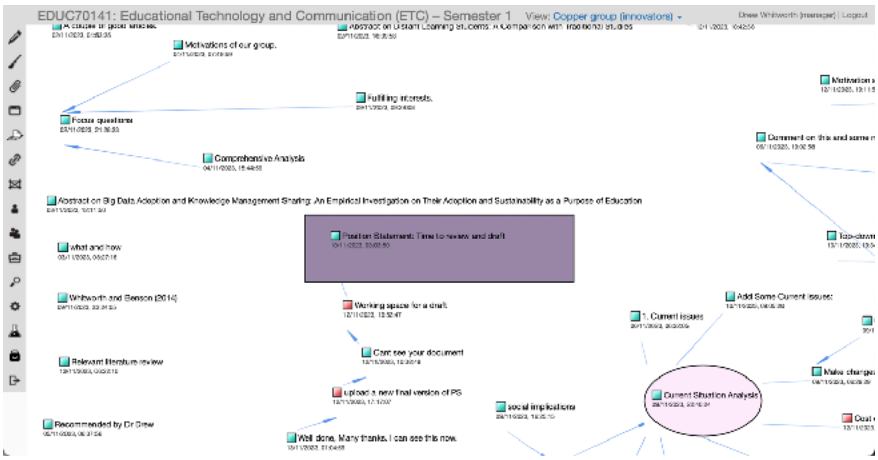
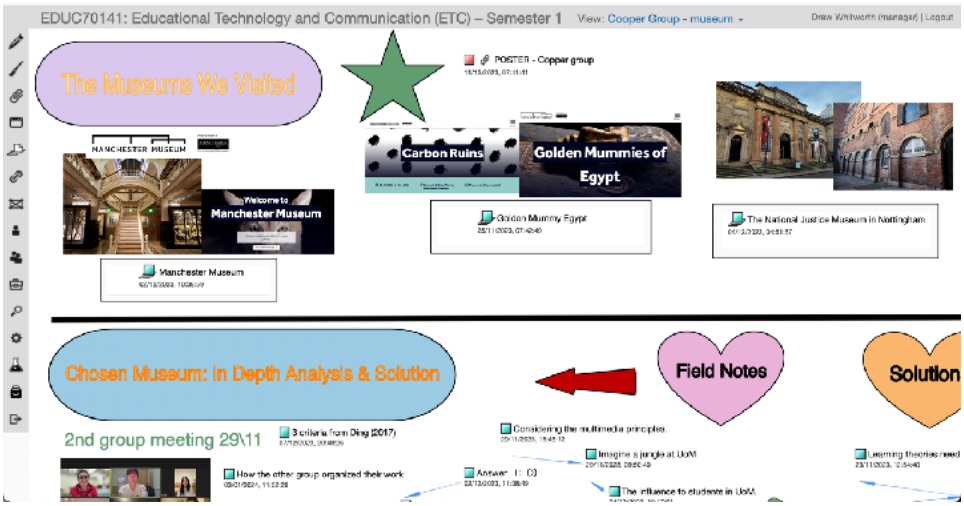


Figure 4.4.3: Copper group's Museum view



5. Student accounts of the process

5.1 Introduction

In the final part of their ETC assessment, the self-evaluation document (SED), students were asked to reflect on two educational technology practices, one drawn from outside their study environment and the other being their experience of KF through the semester. (This was an individual assessment; no group work was involved here.) Thus, around 130,000 words of self-reflective writing about the student use of KF was available for analysis, and all were coded by the Pilot's research assistant (Chris You). In this chapter, the dataset therefore expands to include all the students on the course, and not just those in our eight sampled groups.

Our intention is to view the SEDs as a collected narrative of the student experience, and particularly, to use them as evidence for students' self-reported use of feedback, including analytics. What elements of the different feedback mechanisms did students access, and account for? What analytics did they use, and did they say why? Figure 5.1.1 is a summary of the principal themes that emerged from the coding of these self-evaluation documents (SEDs).

Figure 5.3.1: Key themes from students' self-reflections.

5.2 Feedback

5.2.1. Intrinsic feedback

The SED data reveals that ETC students like the *immediacy* of feedback on the course. They can quickly gauge the reception of their ideas and make prompt adjustments to their understanding or argumentation strategies. Most importantly, though, they recognise that feedback can come from peers as well as tutors and the course unit leaders. This student makes an insightful observation:

> *Students primarily receive feedback through interactions with their peers, observing everyone's behaviours and thought processes, and ultimately forming a community consensus. In this process, the majority of the feedback for a student comes from their fellow classmates.* (Venus)

Though students valued the immediacy of such discourse, this is not the same as its being *instant* in ways it would be in synchronous, classroom dialogue. In fact (here see §6), students find this kind of face-to-face communication more difficult. The asynchronous nature of the online discussion is valued because it allows time for reflection:

> *…taking enough time (almost 1 day…) to consider different perspectives, and formulating a mature response. This process ensures that my response is of high quality and prevents any awkward situations that may arise from not knowing how to respond in real-time communication.* (Orange)

This reflection was perceived by some as adding quality to their contributions. Arguably it also made students more willing to consider and accept peer feedback.

The spread of certain practices, like 'beautification', between groups has already been mentioned in §4.4 and was supported by some observations in the SEDs:

> *I was unaware we could import clipart drawings onto the discussion boards until the first assessment was completed and saw how other groups utilised KF's potential.* (Red)

This kind of inter-group observation was facilitated by the course unit leader through the extrinsic feedback mechanisms, but should also be seen as groups learning good discursive practice from each other. In effect, extrinsic feedback guides groups towards recognising the value of intrinsic feedback: no one is *told* to change practice, but it happens anyway. It is interesting that this student from the Rose group perceives the environment as one 'without supervision':

> *The environment without supervision made me less reliant on teacher's guidance, but more focus on the intrinsic feedback* (Rose)

And this student also notes that they felt a sense of *self-efficacy* when it came to their ongoing development as a competent actor in this setting:

> *I figure out how to improve in the future without teacher intervention.* (Sapphire)

5.2.2. Extrinsic feedback

The SEDs offer evidence of where students recognised, and responded to, extrinsic feedback from the course unit leader and tutors. There were points when this was stated unequivocally, in what is, admittedly, a kind of 'leader says jump, we jump' reaction:

> *As [course unit leader's] feedback from the first discussion highlighted the importance of key concepts in our discussion, we focused more on incorporating key concepts into our museum discussion.* (Cherry)

This student from the Sapphire group offers a little more detail on how they were able to change practice to recover from their struggle with discussion activity 1:

> *I should have endeavoured to understand the core ideas behind authoritative sources and how they related to the topic I was exploring. Thankfully, with the guidance of my teacher, I recognised this and made improvements in my museum activities.* (Sapphire)

It is not just the course unit leader who offers feedback, however. This student acknowledges the role of their tutor, and notes that their utterances were made in the course of the tutor's moderation of their discussion activities:

> *[Tutor] doesn't comment on every post all the time. Instead, he gives feedback on more significant points. [Tutor] usually doesn't provide direct suggestions on what to do. Instead, he prefers to give us some prompts to thought, such as build-on our post and asking questions like "What would happen if more thought was given to this direction?" or "What would happen if something was added to the consideration?" This type of guidance not only avoids relying on teacher feedback but also enhances the thinking processes of learners.* (Venus)

More constructively still, tutors were able to use their contributions to *model* good knowledge-building practice, and suggest specific 'good moves' in the dialogue to groups and individuals:

> *After the communication with [tutor], he suggested me to divide the whole passage into different topics. I realised I presented too much information in one note and my group mates may feel pressure and difficult to respond. Therefore, I posted three notes about the set reading in the second*

assignment in different topics, which involved the name design, communication model and game mechanism, then I obtained many replies from my group mates and refined these thoughts with them. (Ruby)

5.3. Learning analytics and scaffolds

5.3.1 Use of learning analytics

Students made a number of references in their SEDs to the learning analytics that the discussion platform made available to them, discussing which ones they used, but also what for, and reflecting on the benefits these analytics brought them. Tools where usage was reported within the SEDs include: Ideas building; Key concepts; Time machine; Word cloud; and scaffold diversity and scaffolds (see §5.3.2).

For details, see the full version of this report.

5.3.2. Scaffolds

Scaffolds "gave me a sense of direction when building a note" (Saturn). This direction applied not only to the content of the contribution, but also its tone and phrasing. As this student noted:

> *As group members express various views by using professional terms in discussions and various writing styles, and because we are all non-native learners, there will be some ambiguity in English using and reading. Thus, this scaffolding functions as an emphasis.* (Cherry)

This guidance was considered particularly useful when it came to expressing disagreement. This is something many students find challenging, a reticence that helps explain why it can prove difficult for students to engage in critical dialogue in a classroom, or even in an asynchronous discussion environment. However:

> *One scaffold support in particular, "This idea might be less relevant for our work because", permitted the rejection of certain ideas in a less awkward manner.* (Red)

> *…they guided me to refute and challenge others' opinions. At first, I was afraid of making mistakes in KF due to the inertial thinking that "speak wrongly" and "contribute wrongly" will be humiliating and negative, so I did not dare to speak or refute others. However, the scaffolds guided me to say "disagree" and give other people suggestions to improve, which developed my confidence to express myself.* (Gold)

Therefore, scaffolds were perceived as guidance on how to phrase what might otherwise have been considered ambiguous and possibly awkward utterances. They helped students reduce their epistemological and procedural uncertainty in this setting.

5.4. Conclusion

A common theme that emerges is that the availability of analytics is *motivating*; along with the scaffolds, they provided both illustration, and guidance, as to what kinds of moves were being expected of students in the overall course dialogue. Not all students reflected on their use of analytics by any means, and so for at least a portion of the cohort, these data were just not perceived as relevant. Yet at the same time, no one reported that having access to these kinds of data and visualisations retarded their participation: an argument sometimes used in the debate about analytics is that they prompt a feeling of competition between students, and that it can be demotivating when a student sees themselves as 'bottom of a leader board'. But the KF analytics referred to by students do not present as 'rankings'.

6. Tutors and course leader: their role and perceptions

6.1. The tutors' role(s)

The ETC tutor role involves facilitating and encouraging discussion, and monitoring engagement, chasing up groups or individuals who are not reaching an acceptable level of participation. As §5 noted, tutors play a role in *modelling* good knowledge-building practice. This involves (Bereiter and Scardamalia 2016, p. 23) "engaging students directly with good moves in knowledge-creating dialogue". Students might be told that they are expected to make moves such as judging the relevance of found papers, for instance, but that does not mean they are all aware, in advance, of what might constitute such a judgment.

An obvious instance of the impact of a particular tutor on knowledge-building practice came with the groups in the Metal set, guided by one of the two tutors who had previously worked with the course unit lead, and with KF. This whole set displayed far higher levels of cross-group interaction than in the other four sets. This quote (from a KF note) suggests how this tutor prompted the whole set to consider other groups' work, and he reported that this was also a specific activity undertaken in the tutorials.

> *Don't forget to consider the simulation in the context of all of the stakeholders. Try to pre-empt and consider what the objections might be, and how the proposal impacts upon each of the stakeholder groups. Don't confine your research simply to this Group. Do have a look at what the other Groups are saying. Perhaps you could allocate one of the Group as a researcher who looks at what your group are saying and then looks for similar or opposing views within other groups. Understanding the requirements and thoughts of the other Groups will help to better inform your position.*

6.2. Tutors' accounts of the process

Semi-structured interviews were conducted with the five tutors. These took place on Zoom and lasted for around 1 hour each. The interview questions focused on how tutors used KF to prepare for and conduct tutorials, and how this affected their workload and assessment. We additionally asked them about their experience with distance learning students. All interviews were transcribed and coded thematically using NVivo. Themes are described below.

6.2.1. Discussion practices around tutorials

Tutors reported that they were able to use KF to familiarise themselves with students' preparatory work for tutorials, and that this enhanced their own tutorial preparation. It allowed tutors to identify common themes that were already emerging in discussions, as well as gaps that could be addressed in the forthcoming session. One tutor noted that:

> *I looked at some of the notes, just to get an idea about key themes and what I could address in the tutorial...I wanted to get a sense of who was talking already – not to challenge them immediately but just to ask them about the idea; just wanted to know who to pick on.*

Thus, having visible records of the students' ongoing cognitive work helped tutors to establish, in advance, levels of student engagement and participation, both in terms of individuals, for groups, and for the set as a whole.

However, there were challenges in encouraging live interaction and immediate post-tutorial engagement, which was disappointing for the tutors who expected more active participation. Tutors observed a reluctance among students to volunteer ideas in public discussions, preferring to post on KF where they felt more comfortable, yet this did not always translate into active tutorial discussions. Some attributed this to language issues or fear of being put on the spot. The course unit lead noted that:

I was… aware of the dichotomy between students' propensity to talk in public, and their intelligence on KF, and indeed when we did the poster session…

An online discussion environment can seem less 'exposed' to the students. Its asynchronous nature means that responses can be considered before being made; a rehearsal of the lines of the performance, in effect. Using KF *during* tutorials — synchronous sessions — did not result in the same level of engagement as when used *before and after* tutorials, perhaps because this sense of exposure returns in synchronous dialogue regardless of the medium.

6.2.2 Use of analytics

Tutors used analytics to determine student engagement, identifying patterns in who was initiating discussions versus merely responding. Analytics were used to ensure key concepts were being addressed in discussions. Tutors used this data to identify gaps in content coverage and to guide students towards areas not yet explored.

I looked at the engagement, who was engaging with who, the patterns, who was only responding to posts as opposed to generating posts, just to see the originality of posts. Responding is a different set of skills. I was looking for movers and shakers.

I also looked at key concepts, what was being talked about and most importantly what was being missed.

Specific analytics that tutors mentioned using were: the Ideas building tool, the activity dashboard, key concepts, scaffold diversity, and the time machine. These are much the same as those reported by the students, with the exception of the activity dashboard: this offers a relatively crude quantitative measure of 'total activity' that, for tutors, had some relevance in this early stage of the course.

Some tutors looked at other groups or tutor profiles as a benchmark. There is an aspect of 'learning by example' here, picking up practice from the more experienced (or effective) tutors.

I looked at analytics in the other groups, just to see if there are any patterns…

In a sense, this also enfolds the tutors in their own self-reflective, ergative assessment approach: not only can the students make changes to practice in the light of analytics and other feedback, but so might the tutors.

6.2.3 Missed opportunities and challenges for tutors

Some tutors felt that they hadn't fully explored the capabilities of KF. They desired to explore the tool's analytics more deeply to enhance tutorial preparation and student engagement. They recognised the potential for analytics to reveal patterns in student interactions and highlight key concepts that were being overlooked.

A recurring theme was insufficient training or awareness about specific tool functionalities. Tutors felt they missed out on maximising KF's potential due to limited knowledge. There was also a lack of confidence in using analytics effectively. Tutors were unsure how to interpret and apply analytics data to enhance tutorial preparation and student engagement.

6.2.4. Suggested enhancements

See the full version of this report.

6.2.5. Distance learning

Distance learning students provided a greater diversity of perspectives, enriching the group interaction and discussions both in KF and during tutorials (a quality of discourse termed *alterity*:

see Webster and Whitworth 2017). These findings suggest that DL students play an important role in the dynamics of group interaction and the overall learning experience.

6.2.6 Workload

Becoming familiar with a new way of working, accommodating a new technology, itself has workload implications. Tutors noted that they would have liked to have used the KF platform over a more extended period of time to leverage its full benefits.

The application for teaching assistance specified that to run one set was defined as 45 hours' work, a load that varied across the semester, peaking in weeks 7 and 8 — that is, the moderation of activity 1 (see the breakdown in Annex 3). In our interviews, though, no tutor offered any specific observations about whether this allocation matched their actual hours of work on the course.

Tutors experienced varied effects on their workload due to KF. Some found it added slightly to their planning time, while others did not notice a significant change. The tool was seen as helpful for tutorial preparation by some, indicating a shift rather than an increase in workload.

6.3 Course unit leader reflections

6.3.1 Group allocation process

See the full version of this report.

6.3.2. Marking workload

See the full version of this report.

7. Discussion

7.1 Ergative assessment: a dialogic foundation for design

In their new book *The Theory of Educational Technology*, Wegerif and Major (2024) call for a *dialogic* foundation for design: and they mean the design not just of technology, but of learning more broadly. The ergative assessment model is one such foundation, being inherently dialogic. These dialogues are taking place at micro- and meso-levels, as students discuss and work out answers to the problems they have been set, individually and as groups. There is also a dialogue taking place at a macro-level, structured through the design of the course. Each activity in ETC is a move in this broader dialogue, which is in turn comprised of sub-moves, through which feedback can be provided in various ways. With the assessment guidelines, marking rubric and scenario the course unit lead makes an initial move in the dialogue. Students respond — working at first with their tutor in the tutorial setting — then on KF, in their groups. Feedback is the next counter-move. Activity 2 a response in turn.

The emergent nature of ergative assessment explains why it is important, even essential, that there are *multiple* points of observation within a course. Later points are not simply repeats of earlier ones; students are not being asked to do the same things again, albeit perhaps with a different problem. Rather, they are being asked to *respond* to a dialogic move, namely the feedback they have received — a move made even more forceful for students when it comes accompanied by a grade (particularly a grade that is lower than they hoped for). The different levels of dialogue interact around a range of different knowledge representations that include notes, set texts, scaffolds, group statements, feedback and more. It is the accumulated record, and structuring, of these various representations that constitutes the *observable work* within a course unit, and the basis of the ergative assessment approach.

As students navigate their way through these knowledges representations, producing more of them as they do so, "How are we doing" becomes a key question. This is not just a plea for reassurance ('Are we doing OK?'), though it can be that. It is also the central element of the self-reflective approach (Schön 1983). At the start of the course, and as it proceeds, students are epistemologically and positionally *uncertain* — and they, understandably, seek to reduce this uncertainty. They seek clues from the course unit leader and from the environment s/he has created as to what kinds of practice will be rewarded. This is why the giving of feedback during, or very soon after, each point of observation is so important.

7.2 The role of analytics and scaffolds

With each iteration in the course unit dialogue, then, the ergative approach requires teachers to, firstly, *conduct an observation* of some kind, then *assess the observed contributions* of both individuals and groups, and finally to *give feedback*. These requirements may, on the surface, seem to add significantly to the workload required to run a typical course, whether at undergraduate or postgraduate levels[8]. Yet with the support of analytics, the workload is feasible[9]. Some of the analytics referred to in this report were only generated post facto, but their value has been

[8] We believe that the principles discussed in this report are fully applicable at both levels of study.

[9] This is a guesstimate based on anecdotal evidence and a gut feeling, but around 150 students is probably the upper limit for ETC in its current form: meaning, the maximum number of students that one person, with five tutors, can comfortably run. With a cohort any bigger than this it would start to become difficult to get certain key dialogic moves produced in time, most significantly the post-activity feedback in week 9. However, with the addition of a second authority figure (also marking in week 9 and after Xmas) and correspondingly more TA hours, the model could scale up to 250 students or more.

ascertained, and in future iterations of the course, might potentially be generated 'on the spot'. Particularly this is true for:

- the auto-recognition and counting of citations, and, as a result, the generation of 'red flags' where there are evident disconnections between the knowledge-building process and the product of this process, as suggested in §3;
- the coding of each note to allow the generation of 'real-time' visualisations of the epistemic network.

7.3 Why generative AI is a challenge: and what we can do about it

We have done our best to minimise direct references to generative AI tools, such as ChatGPT or Claude, in this report. All the same, we do not believe it is going too far to say that if the type of disconnection between the cognitive process and the products of that process that we saw in activity 1 of ETC is (or will shortly become) widespread, this is an existential challenge to the whole integrity of HE's assessment model.

There were several occasions in which there was ample evidence that students had used some kind of AI tool to summarise the content of authoritative sources, and in some cases, the work of other groups. But drawing on AI to offer *inputs* into a knowledge-building process should be unproblematic, as long as students then *put in the necessary knowledge-building work* to integrate these inputs into their discursive maps: making judgments of relevance, critiquing the input and possibly deciding that some other source might be *more relevant and appropriate to the context*. It was not the simple use of AI that created problems for some groups, but their subsequent failure to engage in dialogue about, and therefore build knowledge around, its input.

Bigger problems arose with those groups who — probably — delegated the production of their *outputs* to AI. Many activity 1 outputs offered little more than streams of text in which the content of a piece of literature would be summarised in one sentence and its relevance to the *specific* context of the professional practicum (in this case, the scenario) ignored. This kind of writing was common enough across group outputs in activity 1, with none of the five different citations given within it discussed during the knowledge-building practice in anything more than a cursory way:

> *To meet these challenges, the quality and diversity of online courses need to be improved ([Fincher], 2001). Secondly, school management team should provide us with sufficient resources such as manpower, infrastructure resources, training, policies, and incentives ({Fosse], 2011). Thirdly, some structural changes to the VLE are supposed to be implemented by the technology department, such as a presentation of new technologies and features. At the same time, all online resources are advised to be permanently available to all faculty and students ([Fosse], 2011; [Anderson], 2003; [Fincher], 2001). Finally, flexible learning technology is helpful to enhance the personalized student experience ([Bigelow], 2015).*

The results were invariably the same as noted above for the Emerald group: whatever insights the group, or some individual members, had gained through their discussion were wiped away.

Yet using AI is, of course, attractive to students. For a start, its use is being intensively hyped, and students are as susceptible to hyperbolic marketing as anyone else. They are being told that this technology will change how we think, how we work, how we communicate — so it becomes an option. We have already noted that students are epistemologically and professionally uncertain when they enter this setting, and seek ways to *manage uncertainty* during the knowledge-building process. Add to this the linguistic uncertainty common to many students for whom English is not their first language (and even of many for whom it is, but still feel they cannot write in an academic idiom) and drawing on the help of generative AI to *help* produce an output is a very attractive option. This is not 'cheating', nor is it work avoidance.

This is why a critically important feature of ergative assessment, in this regard, is that it is *contextualised*. The students work on exploring a particular and *specific* context that is relevant to the professional work they are practicing. The context might be wholly simulated (as with the SSG on ETC) or be partly or mostly 'real' (like the Museum), but either way it forms the background against which students' judgments are made (Whitworth 2020): as a result, references to the context should appear frequently in all ETC student outputs.

One of the clear markers of generative AI use in several of the activity 1 outputs was, therefore, the very generic nature of their discussions. Many references were made to technology use in HE as a general phenomenon, but never related to specific characteristics of the simulated setting as defined in the scenario. However, just as with the lack of reference to set texts — and this is another way of defining an assessment context more specifically — *feedback can be given to groups and individuals* about this lack of contextualisation. And in activity 2, of all 20 groups there was only one which presented a 'generic' poster: the other 19 all included specific details from their chosen design setting.

In the end this is the problem with the misuse of generative AI tools. Use one badly and a student, or a group, will be delegating their cognitive work to an agent that doesn't understand the context, because it has *not been part of the knowledge-building process that has led to this understanding*. Thus, the agent is not an *authoritative* source, when judged against the specific context that is the background to the knowledge-building process.

The writing required at the end of an ergative assessment process is therefore not the kind of writing that can be *wholly* produced by generative AI, although it is accepted that these tools might help with tidying up grammar and phrasing. Making a judgment of relevance about a specific context, and reorganising a group's discursive map as a *response* to this judgment, are discursive moves that AI finds difficult or impossible to make. These conclusions suggest the importance of embedding ergative approaches into assessment practices across the university.

8. Recommendations and development plans

Ergative assessment works. It works in the following ways:

* The approach works as a way of assuring ourselves that the students have engaged in a knowledge-building process during a course unit (or programme), and that the outputs we are reading, and grading, are the outcome of this process. It can reveal where disconnections have emerged between process and product that may be the result of the misuse of generative AI.
* It has a positive impact on the student experience, and on the quality of student work. Students appreciate the regular answers to the question, 'How am I/are we doing?'. The feedback provided in ETC, particularly after activity 1, had a measurable and statistically significant impact on their knowledge-building practice, for the better.
* It works to develop both academic and professional skills in a safe setting (a 'professional practicum'), and recording the dialogues which take place in this setting allow for reflection on and in action when it comes to students developing their identities as competent learners and, eventually, competent professionals.

As is true of ETC's dialogues, the conversation about assessment at UoM, and throughout higher education worldwide, needs to take place at three levels. Hence, this report has been written with these different audiences in mind:
* the micro-level of the individual practitioner (who might be a course unit leader, but also a learning technologist, learning designer);
* the meso-level of programme design and management;
* the macro-level of teaching and learning committees, strategists and policy makers.

Our recommendations for moving forward are as follows:

1. What good work with ergative assessment is already being done at UoM?
No claim is being made that ETC is unique. Some other courses out there are already designed in these ways: this much was clear from a 'wicked problems' session run by the Pilot Owner (Drew Whitworth) in the UoM Teaching & Learning conference on 26/6/24.

The first recommendation is therefore that the Flexible Learning Programme helps **leaders of these courses self-identify**: features from their particular approach to ergative assessment and/or their use of learning analytics could accumulate into a 'toolbox'.

2. Training and staff development
ETC tutors wanted more initial training in this new approach, and understandably. In the context of the 2023-24 offering, everyone (the project team, course unit leader, tutors) was finding their way to some extent, just as much as the students. But better and more focused work can be done in future offerings, based around the insights gained from the detailed evaluation conducted in this Pilot.

And for the broader context, from the 'toolbox' mentioned above, **staff development resources** can be developed. These do not have to be complex: they should start with a list of self-reflective questions…"What constitutes student work on my course?", and continue in much the same vein, guiding the course and assessment designers to reflect on, for example, whether they may already have points of observation set up, how to specify a clear and valid marking rubric, and more. These resources will be developed during semester 1 2024-25.

3. Integrate an understanding of ergative assessment into course unit reviews
Examples of good practice are useful, but they are not sufficient to achieve the kind of widespread shifts in practice that are required as a response to the challenge posed by generative AI.

We therefore recommend **all course unit specifications need to indicate points of observation prior to the final coursework submission**: and it works best when these are accompanied by some kind of summative assessment — a grade, in other words (or more precisely, a clear indication of what contribution this particular part of the activity will make to the final grade). Each part of the assessment should include a **clear marking rubric**, **including reference to learning analytics** if these are to be directly drawn on in grading. (Not *every* course unit requires this kind of scrutiny, but see recommendation 5 below.)

4. Better planned timetables
Ergative assessment is not 'over assessment' — ETC students only rarely feed back that they feel they have too much to do — not least because alongside it, we can and should **make room in timetables for the work required**. Course unit approval committees should be more willing to **challenge unimaginative teaching timetables** — are those two hours of lecturing in week 11 *really* necessary? What will they actually add to a knowledge-building process that should already be well under way by that point in the semester? (Along the way this recommendation may also help relieve the intense pressure on teaching space on campus, which was another broadly intended outcome of the Flexible Learning Programme.)

Let us look more closely at 'contact hours' on course unit specifications, too. Those '100 hours of private study' that were blithely written into it: what exactly are students being asked to spend this time doing? We do not need to start micro-managing our students' time — but we should offer more detail, for the sake of clarity, as to the kinds of things we expect them to be working on, and when.

5. Plan pedagogy at programme level
We should recognise that course units are, themselves, moves in a broader dialogue. A participant in the workshop on 26/6 raised the valid point that ETC's approach sounded great, but what was its worth if it was used on a foundational course unit and then never followed up in later units that returned to a traditional, closed-cognition approach?

It is not necessary for every single course unit to 'go ergative': but **ensuring that students are engaged with the approach at least once in each semester of study** seems sensible. This would allow for a sustained check on the progress of individuals, and also to set up an intentional, programme-wide dialogue when it came to assessment. As is the case within ETC, the later activities are not just 'repeat performances' but later moves in the emergent dialogue. Similarly, ergative assessments in year 2 of a UG programme could be designed around the knowledge that all students had already participated in this kind of knowledge-building work in year 1, with reflection on what was learned that first time round being integrated into the preparatory work for the later unit — and so on through the programme.

6. Conduct a scoping review of possible learning analytics
The Pilot study shows that, **with the use of learning analytics** and the support of teaching assistance (tutors), as well as the careful design of the timetable, the approach is viable in workload terms. Some specific analytics have been referred to frequently throughout this report, but it needs to be remembered that these are not being suggested as 'one size fits all' solutions, appropriate across all course contexts and disciplines. For example, other group assignments in which students are being asked to work together to solve a design problem — the core of the ergative approach — may not have the same requirements when it comes to citations (or, the 'constructive

use of authoritative sources'): the emphasis may instead be on, say, how well students can interact with a client, or with members of the public. In these types of professional practicum, the 'good moves' in dialogue will take a different form, and be identifiable through different textual markers.

Yet the basic principle remains the same: through the use of analytics of some kind, and clarity to students and markers alike about how these are being, and can be, used to bring 'many eyes' to bear on the "How are we doing?" question, we can bring the kind of objectivity and validity to the grading of ergative assessment that Messick (1994) rightly demands. Ergative assessment depends on at least some of the knowledge-building dialogue within a group or cohort being recorded; this means it must be based around some kind of digital discussion platform. Knowledge Forum was used on ETC as one example of such a platform, but it is not the only game in town: in any case, it is inevitable that UoM's management would be reluctant to advocate the widespread use of a platform with which there was no business relationship[10].

We therefore recommend that a **scoping review** is needed, looking at available discussion platforms (including, but not limited, what will be available on Canvas) and assessing how the features and functionality of each might support the ergative approach. What analytics are built-in to the platform, what might be generated in an indirect way, and how about other features (like idea A below) that are not related to analytics?

For more specific details of features to focus on, see the full version of this report.

Bibliography

Bereiter, C., & Scardamalia, M. (2016). " Good Moves" in knowledge-creating dialogue. *QWERTY-Interdisciplinary Journal of Technology, Culture and Education*, 11(2), 12-26.

Cope, B, M. Kalantzis, & D. Searsmith. 2020. "Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies". *Educational Philosophy and Theory* Feb 2020: 1-17.

Hay, D. B. (2007) Using concept maps to measure deep, surface and non-learning outcomes, *Studies in Higher Education*, 32 (1), 39–57.

Lin, P. Y., Hong, H. Y., & Chai, C. S. (2020). Fostering college students' design thinking in a knowledge-building environment. *Educational Technology Research and Development*, 68, 949-974.

Lloyd, A. (2012). Information literacy as a socially enacted practice: sensitising themes for an emerging perspective of people-in-practice, *Journal of Documentation*, 68 (6), 772–83.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational researcher*, 23(2), 13-23.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. *Liberal education in a knowledge society* 97 (2002): 67-98.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Routledge.

Shaffer, D. W. (2017). *Quantitative Ethnography*, Cathcart Press.

[10] Should any readers of this report be interested in investigating KF with an eye on possibly using it in their own teaching, please contact the corresponding author on drew.whitworth@manchester.ac.uk .

Smith, P. V., & Whitworth, D. (2024). Anonymous assessment: is it still worth it?. *Teaching in Higher Education*, 1-9.

Webster, L., and Whitworth, A. (2017). Distance learning as alterity: facilitating the experience of variation and professional information practice, *Journal of Information Literacy*, 11 (2).

Wegerif, R. and L. Major (2024). *The Theory of Educational Technology: Towards a Dialogic Foundation for Design*, Routledge.

Whitworth, A. (2020). *Mapping Information Landscapes*: *New methods for exploring the development and teaching of information literacy*, Facet.