Learning, Teaching and Assessment in Higher Education (LTAHE)
Critical Review of Teaching Portfolio

I was appointed on an ART contract as a Lecturer in Cancer Informatics in November 2022 (Department of Biology). This academic year I have supervised stage 4 project students, designed and delivered stage 2 tutorials (2 x 6 student groups), taught and marked on the stage 2 "Big Data Biology" course (BIO00047I) and taken over 9 undergraduate supervisees. I supervise a final-year PhD student and serve on 2 TAP committees. With semesterisation, I am designing the new stage 3 "Genomics" module, and one strand of the stage 2 "Becoming a Bioscientist 4" module. Adjacent to my University role, I am an Elixir-UK Data Stewardship Training Fellow, working with this pan-European organisation to improve data acumen in the life sciences, including generating online and in-person postgraduate training.

I have always had a keen interest in teaching and have actively sought opportunities to develop. I supervised project students and delivered tutorials at York as a Postdoctoral Researcher and Research Fellow. During my PhD at the University of Edinburgh I worked as an undergraduate tutor, demonstrator and marker, and completed over 200 hours of private tuition to high school students. I completed the HEA's "Introduction to Academic Practice" course in 2016, gaining my AFHEA. I am an active STEM ambassador.

I am passionate about helping students to understand, interrogate and present biological data, embedding and applying this knowledge within exciting current research and cutting-edge techniques.

LO1 - Design coherent blocks of teaching and learning activity, modules or programmes of study that are appropriate to the students and subject being taught

Stage 2 tutorials in the Biology BSc and MBiol programmes are 50 minute blocks of small-group teaching (6 students) timetabled fortnightly. There are assessment requirements, but content is decided by the tutor. This provides an excellent opportunity to develop and test new teaching activities for use in other settings. In the Spring/Summer set (8 sessions) this year I had two groups and designed a single series delivered twice, covering the "Uses and abuses of sequencing technologies". By this stage, Biology undergraduates have only been superficially introduced to these topics. In previous years my tutorials were content-focused and predominantly an exercise in knowledge transmission. This was an active decision given their limited exposure to these topics, but it created a group divide: students who "got it" and thrived, and those who didn't. This limited engagement when discussing topics as a group, and reduced the success of practical activities. It also resulted in me talking a lot!

Given this past experience, I instead adopted a learning-focused approach so that students could apply what they had learned [1] to problem-solving activities, paper presentations, and topical discussions. Unlike previously, I formally defined learning objectives which were visible on my whiteboard (and referred to) throughout. Each session was divided into: 1) an introductory exercise using post-it notes and key words where students could highlight to me areas they understood or didn't; 2) a discussion facilitated by me which built on the post-it note exercise; 3) a student-led problem-solving, presentation or discussion task; and 4) a recap led by me with the opportunity to ask questions, revisiting the key words from the introduction. I also dedicated the final session to 1:1 feedback to cover the tutorials as a whole, the assessed essay, and areas where students thought the series could improve.

Overall I found this approach to be a great success. I felt more organised, all sessions had active engagement from the students (LO2), and feedback was positive. Students were particularly engaged by the challenging ethical discussions in our penultimate session, with 7/9 attendees mentioning it unprompted as a highlight in their feedback. This was particularly promising as I designed it as a trial for a full workshop on genomic ethical considerations as part of the new stage 3 "Genomics" module. Most of these students will take that

module, so I am hopeful they will feel like active participants in informing their own upcoming teaching, rather than their feedback simply helping cohorts in subsequent years. A drawback of this tutorial series was that each session needed a critical mass of attendees for the student activities to really flow, and one of my groups only had >50% attendance in 2 of the sessions. In such small groups I was agile enough to adapt planned activities, usually by expanding the roundtable discussion. However, this must be a key consideration when designing activities for larger groups.

Before this year, most of my solo teaching was content-focused. But with the University preference for flipped classrooms, I was keen to practise workshop-style teaching. My tutorial redesign was one, as was my microteaching activity, where I delivered the central dogma of biology to a mixed peer group from Law, Computer Science and Health Sciences. I received good feedback on my slide design, how engaging the activity which I had given them the skills to complete was, and the overall structure of the session, which started with the big picture, gave them skills for completing a simplified version, and then returned to the big picture with the application of their new skills.

Both these positive experiences provided affirmation that I was getting the structure and balance of my workshop-style activities right, at least for small groups. Another session of teaching I designed this year was a brand new 1-hour workshop delivered to a mixed group of 40 MBiol and PhD students, postgraduate researchers and group leaders. This session introduced the cancer bioportal (cBioPortal), a browser-based online resource for cancer genomics, through an introductory presentation (10 minutes), live website demonstration (10 minutes), a set of problem solving tasks (25 minutes), then a final 15 minutes of recap and presentation on advanced applications. This was the subject of my teaching observation and is discussed further in the following section (LO2).

LO2 - Employ teaching and learning methods that are inclusive, participatory and will enable effective student learning

During the design of the "Introduction to cBioPortal" workshop, I was most concerned about the balance between content delivery and task-based, learner-led exploration of the dataset. I wanted the session not just to be a summary of what cBioPortal was, but rather a tutorial on how to use it. I was concerned that in a 1-hour session we would not be able to cover all the material, leaving students dissatisfied compared to the learning objectives, which I had shared with the advert for the session. At sign up I collected data on the registering students, including on the specific cancers they worked on. This enabled me to ensure that at least one of the three prepared problem solving tasks covered cancers of interest. I wanted to ensure that all teaching materials were accessible for students after the session, including a demonstrative screen capture of the website walkthrough, so I hosted a website for the teaching material. This was good for the students, but also makes it straightforward to repeat the session again.

For the observation, I requested feedback on 1) organisation of the session including timekeeping, and 2) whether students were engaged throughout.

Helen's feedback on my organisation was very positive: "*Clearly an extremely well-prepared session*", "*Well-planned with carefully thought-out structure and timing*", "*A very comprehensive, learned-oriented and helpful overview*". The students seemed to agree in their post-session feedback, with all responders suggesting the balance of theoretical and practical components was "about right", and that they would recommend the course to others. The course was given an average rating of 4.6/5.

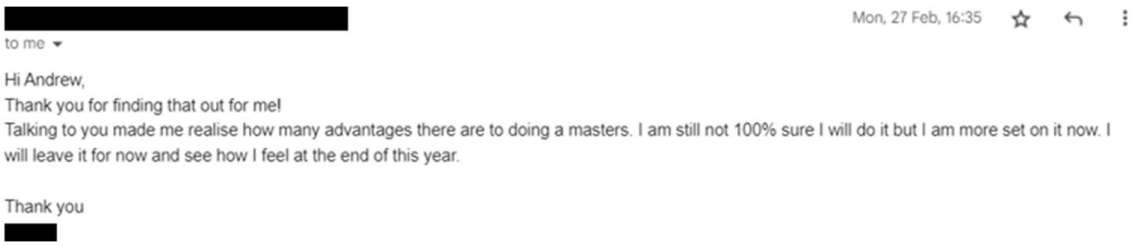
Most students thought the best part of the session was working with the data in the problem-solving activity, and that the worst was not having enough time to explore all data and problem-solving tasks fully. Helen thought the planned 25 minutes was "*...a good length of time for such tasks, and appropriate to give a choice*

of tasks…". However, Helen raised some very good points not about balance, but about regularity of interaction with students. Until the task started at 22 minutes in, I hadn't asked the students a question, or checked engagement and understanding. Helen suggested that I *"...consider increasing opportunities for student interaction throughout the session…*", with set points for questions and/or a digital resource, such as a Google doc or Padlet which could be reviewed while students were engaged in the tasks. I have taken on these points fully. I admit I didn't want to get into a broad Q&A when session timings were tight ("*...a brisk but manageable pace.*"), but there wasn't time for checking answers to the problem solving tasks either (and these were also not on the session website). Helen was very positive about how I supported students during the problem-solving tasks, "*...circulated round the room regularly…monitoring students attentively without being intrusive, and answering queries*", although did suggest that I incorporate other ways of monitoring learning into my teaching plan. There is long-standing evidence for the benefit of breaks and activity shifting for concentration [2,3], and this is likely to be more pronounced in a group of undergraduate students within a long teaching block, rather than postgraduate researchers who specifically signed up for a one-off session.

As this observation happened in January, I was able to take on feedback directly into my tutorial sessions, as mentioned in LO1. I regularly checked understanding with questions and new activities which required students to apply material they had learned, rather than simply repeating it. 2 of my students commented on this specifically in their feedback sessions, "*...very useful to revisit topics...after long gaps between tutorials and within the same tutorial*". Whilst I am yet to apply the observation feedback in a large workshop setting, it is already part of my plan for workshops in my taught modules next year. I need to review various online tools such as Padlet and Mentimeter to see how these can help with student engagement.

LO3 - Develop learning environments and approaches to student support and guidance

Becoming an Academic Supervisor has been new for me this year, but my experience in small group teaching and previous reflections on adapting teaching styles to be inclusive for heterogeneous groups (using the literature, e.g. [4,5]) has been beneficial. My general approach to all students is to be friendly, open and empathetic, but in recent years I've had to develop more of my challenging/directive role when supervising year-long student projects. In supervision meetings my students at all stages have asked for opinions and advice on module choices, placements and careers, and I've been careful to give constructive but honest responses, rather than "safe vagueries". This has been well received (e.g. Figure 1), with students now seeking me out for advice, rather than simply waiting for the next timetabled meeting.


[redacted] Mon, 27 Feb, 16:35

to me

Hi Andrew,
Thank you for finding that out for me!
Talking to you made me realise how many advantages there are to doing a masters. I am still not 100% sure I will do it but I am more set on it now. I will leave it for now and see how I feel at the end of this year.

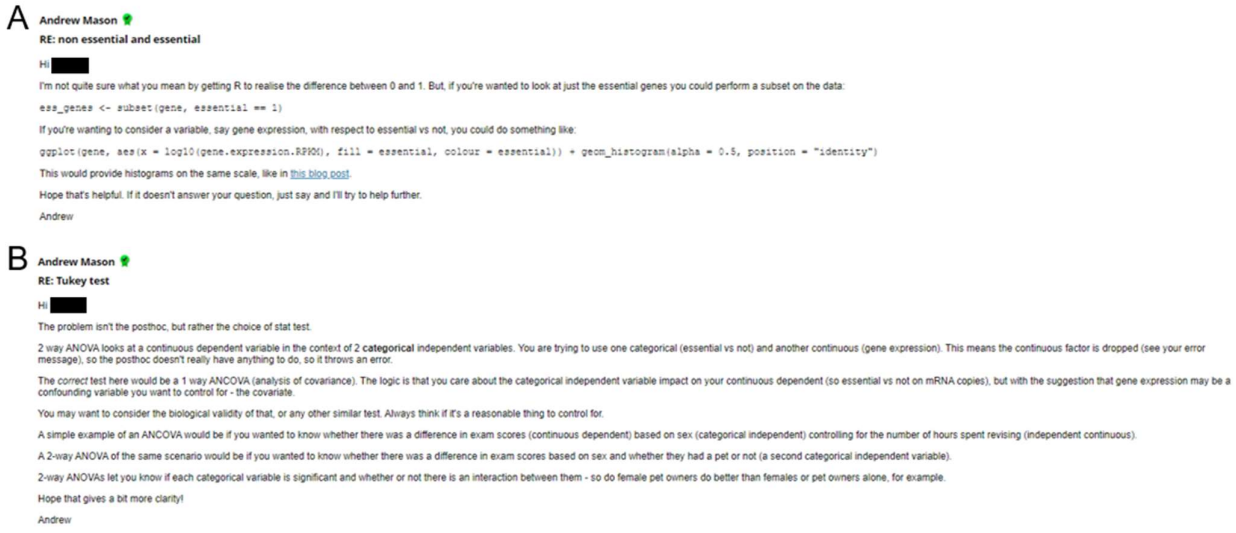
Thank you
[redacted]

Figure 1. Email from one of my stage 2 supervisees who was ready to change programme away from the integrated masters, largely because of a lack of information on what the final year entailed and the worry of an extra year of debt. We discussed pros and cons of both options, and after the meeting I confirmed cut off dates for decisions so the student had all the information, with offers of further support if needed. The student has since decided to stay on the integrated masters programme in order to complete a "proper" research project before graduating.

A specific case where I had to switch between supportive/counselling style supervision and a more challenging/directive role in a single meeting, was with a final year student who had been offered a PhD position but the offer had been withdrawn after an unpleasant and very unexpected reference from the

year-in-industry supervisor. Unsurprisingly the student was incredibly upset, particularly as they felt the negative reference was unjustified. Initially we simply talked through what had happened in an unrushed and non-judgemental space. I then gradually moved the meeting into a directed exercise where we discussed next options and came up with an action plan, following the GROW framework [6] I first came across through LTAHE. We've continued to develop this tutor/student relationship, with further career discussions, CV advice and mock interviews. I think it's been highly rewarding for both of us.

This personalised approach to teaching always feels to me like the "right" way to teach; to fully engage students and facilitate their learning. I have been particularly conscious of this during large computer labs (c.100 students) in the stage 2 "Big Data" module. A minority of students do not feel comfortable to ask anything during these sessions, but some have used the VLE discussion boards instead. In order to be as inclusive and helpful as possible I have been very responsive to messages, to ensure students using this different, but encouraged, learning environment are not at a disadvantage. However, I've struggled to set boundaries, often replying late at night, in great detail (e.g. Figure 2A,B), and then feeling flat at the lack of response, or a simple "Thanks".


A

Andrew Mason

RE: non essential and essential

Hi [redacted]

I'm not quite sure what you mean by getting R to realise the difference between 0 and 1. But, if you're wanted to look at just the essential genes you could perform a subset on the data:

```
ess_genes <- subset(gene, essential == 1)
```

If you're wanting to consider a variable, say gene expression, with respect to essential vs not, you could do something like:

```
ggplot(gene, aes(x = log10(gene.expression.RPKM), fill = essential, colour = essential)) + geom_histogram(alpha = 0.5, position = "identity")
```

This would provide histograms on the same scale, like in this blog post.

Hope that's helpful. If it doesn't answer your question, just say and I'll try to help further.

Andrew

B

Andrew Mason

RE: Tukey test

Hi [redacted]

The problem isn't the posthoc, but rather the choice of stat test.

2 way ANOVA looks at a continuous dependent variable in the context of 2 **categorical** independent variables. You are trying to use one categorical (essential vs not) and another continuous (gene expression). This means the continuous factor is dropped (see your error message), so the posthoc doesn't really have anything to do, so it throws an error.

The *correct* test here would be a 1 way ANCOVA (analysis of covariance). The logic is that you care about the categorical independent variable impact on your continuous dependent (so essential vs not on mRNA copies), but with the suggestion that gene expression may be a confounding variable you want to control for - the covariate.

You may want to consider the biological validity of that, or any other similar test. Always think if it's a reasonable thing to control for.

A simple example of an ANCOVA would be if you wanted to know whether there was a difference in exam scores (continuous dependent) based on sex (categorical independent) controlling for the number of hours spent revising (independent continuous).

A 2-way ANOVA of the same scenario would be if you wanted to know whether there was a difference in exam scores based on sex and whether they had a pet or not (a second categorical independent variable).

2-way ANOVAs let you know if each categorical variable is significant and whether or not there is an interaction between them - so do female pet owners do better than females or pet owners alone, for example.

Hope that gives a bit more clarity!

Andrew

Figure 2. Example 'Big Data' VLE discussion board posts where I've given expansive and time consuming answers, with links to external material (A), separately typeset code (A), and very detailed answers with multiple conceptual examples (B). In both cases these responses were sent after 11pm. (A) received a "Thanks" and (B) had no response.

Whilst this does constitute specific, personalised and responsive feedback (LO4), it creates a somewhat unrealistic learning environment where students could consider us available at all hours. Giving disproportionate amounts of my time to students is likely to be a constant struggle in my teaching career. In planning future teaching I need to be mindful of the types of questions that we may get and ensure there are sufficient and varied (but dedicated) ways to ask questions (e.g. Q&A sessions, discussion boards and anonymous questioning in workshops through mobile apps etc.). At least posting on discussion boards provides a resource for the entire cohort. I've also received (multiple) direct emails from >10 students on the module, but when directed to the discussion boards they don't post. There is then a crucial balance between fairness, available time, and which desirable learning environments for the student are feasibly deliverable.

LO4 - Devise and apply assessment and feedback methods that inform and measure student learning

I consider assessment the most difficult and stressful part of teaching. This largely comes from my own experience where an undergraduate dissertation mark of 72 would have resulted in an overall 2.1, compared to the 74 I earned which resulted in a 1st. Now as a marker, I realise just how thin-approaching-arbitrary the margin is between these grades, but just how important it may be for the student. Consequently, intra/inter-marker reliability is always on my mind. A good concept to be mindful of, but it creates a lot of self doubt, particularly when second marking with more experienced and senior colleagues.

Whilst STEM subjects have a reputation for quantitative, objective testing, the majority of my marking is subjective. Our existing rubrics are broad and open to interpretation (despite positive implementation of weighted rubrics in literature [7]), often supporting a vague and dissatisfactory "it *feels* like a 2.1" approach, leading to common results seen in the literature: poor inter-marker consistency [8], and frustrated markers and students [9,10]. A further consideration is that project supervisors are also primary markers. Even with double blind marking, it is hard to objectively mark in isolation from student effort and attainment: the "halo" effect [11]. I have found this balance hard when you want a student to do well but, as with personal references and peer review, marking is highly indicative of academic integrity.

Despite (or perhaps because of) my fears, my marking has typically been highly consistent with my colleagues, regardless of career stage (Figure 3A,B). However, across the last 3 years I have consistently disagreed with a senior and experienced colleague in my department (Figure 3C,D). In 2021 I second marked 6 scripts from this supervisor, all of which were beyond acceptable tolerance for mark disagreement. For 5 scripts I was consistently a band below, scoring 2.1s and low 1sts rather than high 70s and 80s. In my opinion, the supervisor didn't really produce evidence that met "excellence criteria", instead relying on "the feeling". As I was inexperienced, we settled on more generous marks. More concerning was the 6th script where I had scored a mid 2.1 compared to the low 2.2 of the primary supervisor. After discussion it was apparent that this was penalising the student for perceived poor commitment to the project (halo effect); not a feature of the mark scheme.

This year I collaboratively marked a poster and blind second marked a final year project with the same academic. The poster marking was potentially the most uncomfortable experience of my academic career, trying to justify a fair and rubric-consistent mark for the student when the other marker wanted to borderline fail them against criteria outside the mark scheme which, in the academic's opinion, were more important. We settled on a low but OK mark, but I felt thoroughly undermined. Whilst the option was there, I felt uneasy asking for a third marker as it wasn't anonymous, the student had already found the experience very stressful, and I was reluctant to add additional work to my colleagues. My worries on reliability and consistency were compounded when our project marking came back 33 points apart (my 88 to their 55; Figure 3C solid line with circles), even though our specific feedback remarks were similar. The 3rd marker in this case scored 87, with the primary supervisor conceding they had "been a bit harsh".

Whilst this close 3rd marking, and other marking consistency (Figure 3A,B), has given me confidence in my own reliability, I remain concerned about inter-marker consistency across a cohort, particularly given strong opinions from long-standing and influential staff members. Collaborative marking meetings and compulsory 3rd marking would undoubtedly help with consistency [12], but are at odds with existing pressures on staff time. A promising avenue may be AI-based marking [13], but I think it is unlikely to be popular with students for subjective assessments.

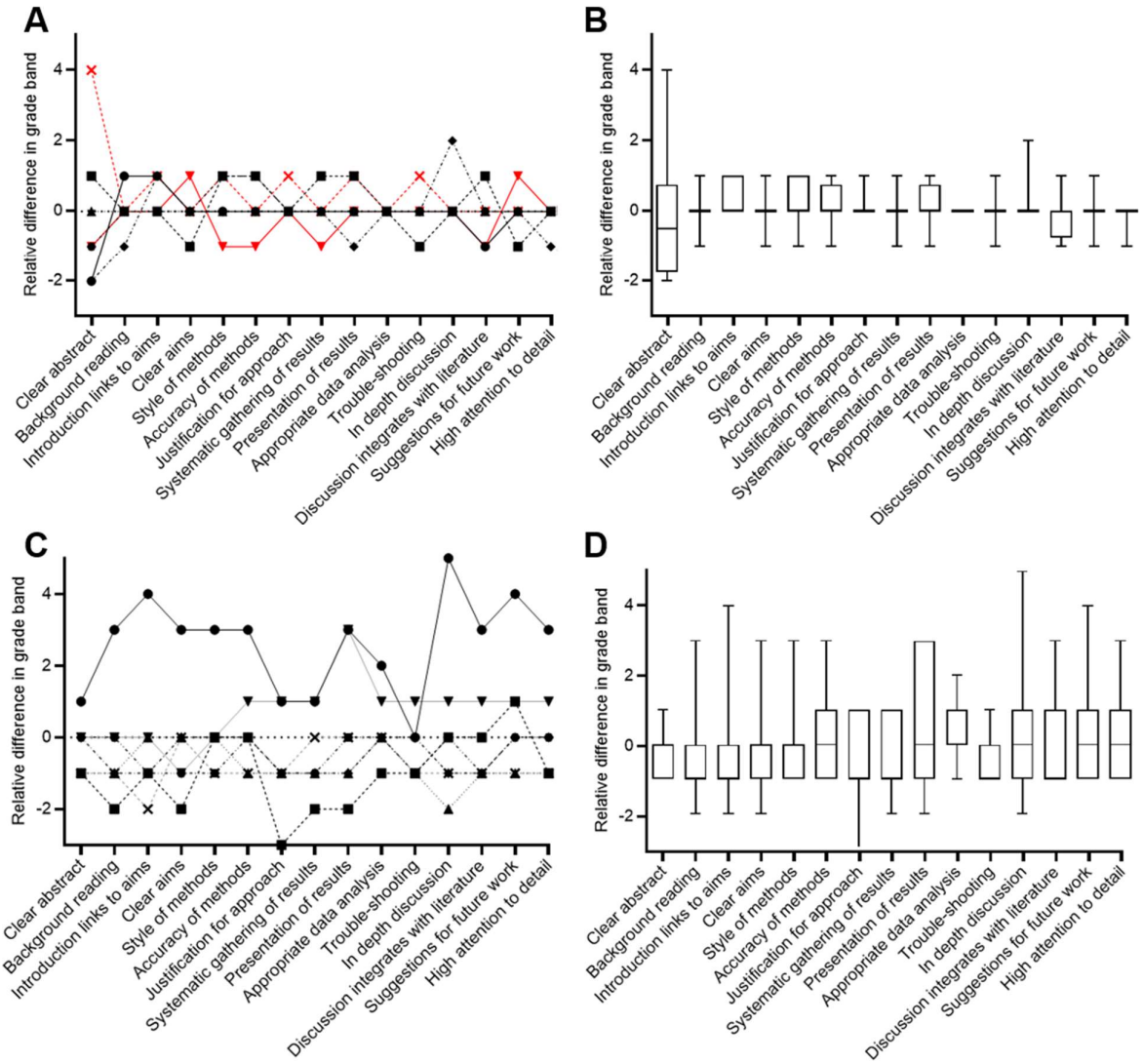


Figure 3. Evaluation of my marking consistency across final-year Biology BSc and MBiol student projects compared to a double blind marker. Each plot indicates the difference in 10-mark grade bands between me and the other marker, with each symbol set and connecting line reflecting a single student submission. Positive scores indicate where I scored a student higher than the other marker. Reduced titles of the marking rubric sections are indicated on the x axis. Marking is typically consistent with my colleagues (A), even with other ECRs (red lines), and the overall grade band typically averages out. Boxplot representation of the same data (B) suggests consistency across the rubric, but suggesting I am more generous with introduction and methods, and my appreciation of a good abstract is more variable. In stark contrast, there is high variability with the senior academic referenced in the main text (C), seen throughout the rubric (D). Consistently, absolute values of summed disparities were significantly more variable with the senior academic (Mann-Whitney U; z=-2.95; p=0.0032).

Summary of personal learning and development

Before this academic year, I already felt confident in the physical delivery of teaching material in an open, inclusive and clear manner, supported by good student feedback in various settings. Increasing both the amount and diversity of teaching this year has brought new challenges, particularly designing new material, but I've felt very supported by the LTAHE/TSLHE cohort, and have found the peer observation and seminar discussions especially enriching. I have a busy summer of module design ahead and hope to draw on my experiences and peers through this process.

Assessment remains the area in which I have least confidence, and I will continue my development, initially through discussions with my departmental teaching mentor, and by suggesting roundtable cooperative marking sessions as part of our monthly departmental teaching meetings.

Word count: 2874

References

1. Prosser M, Trigwell K. Qualitative variation in approaches to university teaching and learning in large first-year classes. Higher Education. 2013;67: 783–795.
2. Stuart J, Rutherford RJ. Medical student concentration during lectures. Lancet. 1978;2: 514–516.
3. Burke LA, Ray R. Re-setting the concentration levels of students in higher education: an exploratory study. Teaching in Higher Education. 2008. doi:10.1080/13562510802334905
4. Fox D. Personal theories of teaching. Studies in Higher Education, 2006;8(2): 151-163.
5. Morss K, Murray R. 'Theory and Practice" in Teaching at University: A Guide for Postgraduates and Researchers. SAGE; 2005;8-27.
6. Whitmore J. Coaching for Performance: GROWing People, Performance and Purpose. Nicholas Brealey International; 2002.
7. Hansson EE, Svensson PJ, Strandberg EL, Troein M, Beckman A. Inter-rater reliability and agreement of rubrics for assessment of scientific writing. Education. 2014;4: 12–17.
8. Williams L, Kemp S. Independent markers of master's theses show low levels of agreement. Assessment & Evaluation in Higher Education. 2019;44: 764–771.
9. de Kleijn RAM, Mainhard MT, Meijer PC, Pilot A, Brekelmans M. Master's thesis supervision: relations between perceptions of the supervisor–student relationship, final grade, perceived supervisor contribution to learning and student satisfaction. Studies in Higher Education. 2012;37: 925–939.
10. Koris R, Pello R. We cannot agree to disagree: ensuring consistency, transparency and fairness across bachelor thesis writing, supervision and evaluation. Assessment & Evaluation in Higher Education. 2022; 1–12.
11. McKinstry BH, Cameron HS, Elton RA, Riley SC. Leniency and halo effects in marking undergraduate short research projects. BMC Med Educ. 2004;4: 28.
12. Hunter K, Docherty P. Reducing variation in the assessment of student writing. Assessment & Evaluation in Higher Education. 2011;36: 109–124.
13. Vijaya Shetty S, Guruvyas KR, Patil PP, Acharya JJ. Essay Scoring Systems Using AI and Feature Extraction: A Review. Proceedings of Third International Conference on Communication, Computing and Electronics Systems. Springer Singapore; 2022. pp. 45–57.