FATE: Alternative Assessment

To make recommendations on assessing students in the age of Covid and Artificial Intelligence, we took a multi-faceted approach that included surveys of the UCI community, a review of the current literature on assessment, and creation of a repository containing examples of effective assessment practices used by UCI faculty.

Surveys

In order to gauge the practices being currently used from both faculty and student perspectives, we created two surveys: one sent out to UCI faculty and one sent out to UCI undergraduate students.

Results of Faculty Survey (n=72)

- We had broad representation in terms of School within UCI, size of classes taught, time at UCI and course level taught (**Appendix A, Figure 1**).
- The majority of respondents had some formal training in pedagogy, with almost 50% reporting at least two different forms of training (**Appendix A, Figure 2**).
- Respondents report using a variety of different assessments, but no single type of assessment is particularly universal, and for almost every form of assessment (except oral presentations) digital submissions are preferred (**Appendix A, Figure 3**).
- Only approximately 25% of respondents report using summative exams to count for more than 60% of a student's grade in that class (**Appendix A, Figure 4**). This usage pattern does not correlate with class level, class size or years of teaching (**Appendix A, Tables 1 to 3**).
- The majority of respondents use some form of group assessment and patterns of assessment usage for these are very similar to what is seen for the individual assessments (**Appendix A, Figure 5**). Use of group assessments also does not correlate with class level, size of class, or years of teaching (**Appendix A, Tables 4 to 6**).

Results of student survey (n=922)

- Student respondents reported that in almost 75% of their classes the majority of their grade was determined by a summative exam (**Appendix B, Figure 1A**).
- Respondents report very low adoption of best practices such as flexible deadlines, dropping of some assignments and allowing assignments to be retaken (**Appendix B, Figure 1B**).
- Respondents did not have many concerns about academic integrity violations, and only a handful of students were bothered about the use of ChatGPT as a means of cheating (**Appendix B, Figure 2**).
- Student comments about what they found to be effective in their assessments broadly fell into six (6) topics (using Natural Language Processing methods for identification of latent topics in textual data). When broken down into sub-topics, almost 50% of the student comments encompassed only nine (9) sub-topics (**Appendix B, Figure 3**).
- From these sub-topics (summarized in **Appendix B, Table 1**), it is clear students appreciate assessments that:
    - Match content covered in class

- ○ Have prompts and questions clearly written and not overly wordy, particularly as some students are English language learners
- ○ Are more than just memorization tests and apply the content
- Students were asked about what assessments were ineffective. They are not frustrated by exams or papers in theory. Much of the frustrations were about exams:
  - ○ Too few exams for a high proportion of the grade
  - ○ Questions are vague or confusing
  - ○ Questions that seem to have no relationship to the course content
  - ○ Timed multiple choice exams were stressful
  - ○ Written exams were often difficult to complete in the time provided

Literature review

We found it interesting that some students value assessments that closely match class content, while others appreciate assessments that go beyond memorization to apply the content in new ways. These seemingly contradict one another. However, this has been discussed in the education literature (see **Appendix C** for a full review). For example, education scholars concede that in most "normal teaching and learning situation[s]" faculty will not be able to know whether their assessment practices are having "a positive impact on learning" (Boud 2007, 19). In an effort to make evidence-based decisions, instructors may turn to student surveys or course evaluations, but these can provide conflicting information. As our student survey suggests, students can view the same assessment both positively and negatively, and this is often tied to a perception of whether or not the assessment was "fair." The perception of unfairness has many sources across higher education, but perhaps the most fundamental issue is that students believe they are not adequately informed of what is expected of them (Flint and Johnson 2011, 38).

Overall, this can leave faculty questioning what type of assessments should be used in their classes, how the assessments should be implemented, and how their efficacy can be measured. To aid in this effort, we collected examples of proven assessment practices used by UCI faculty and asked them how they measured their efficacy.

Repository of Effective Assessments used at UCI

We collected 11 examples of assessments used by UCI faculty and created an online repository that can be a resource for the teaching community: https://bit.ly/fate-assessment The repository includes examples of group quizzes, video presentations, "choose your own adventure" project, open prompt writing, learning communities, and group quizzes.

Summary

Based on surveys of both instructors and students, a clear divergence of opinion regarding the use and effectiveness of assessments emerged between what instructors believe versus what students perceive. This split is also consistent with published research indicating a definitive conclusion on the success of any specific assessment may never be known by an instructor. As such, because of these contrasting opinions, we believe our provision of a repository of many different types of assessments

can provide a broad resource for instructors to evaluate and reach their own conclusions as to what may be effective in their own courses.
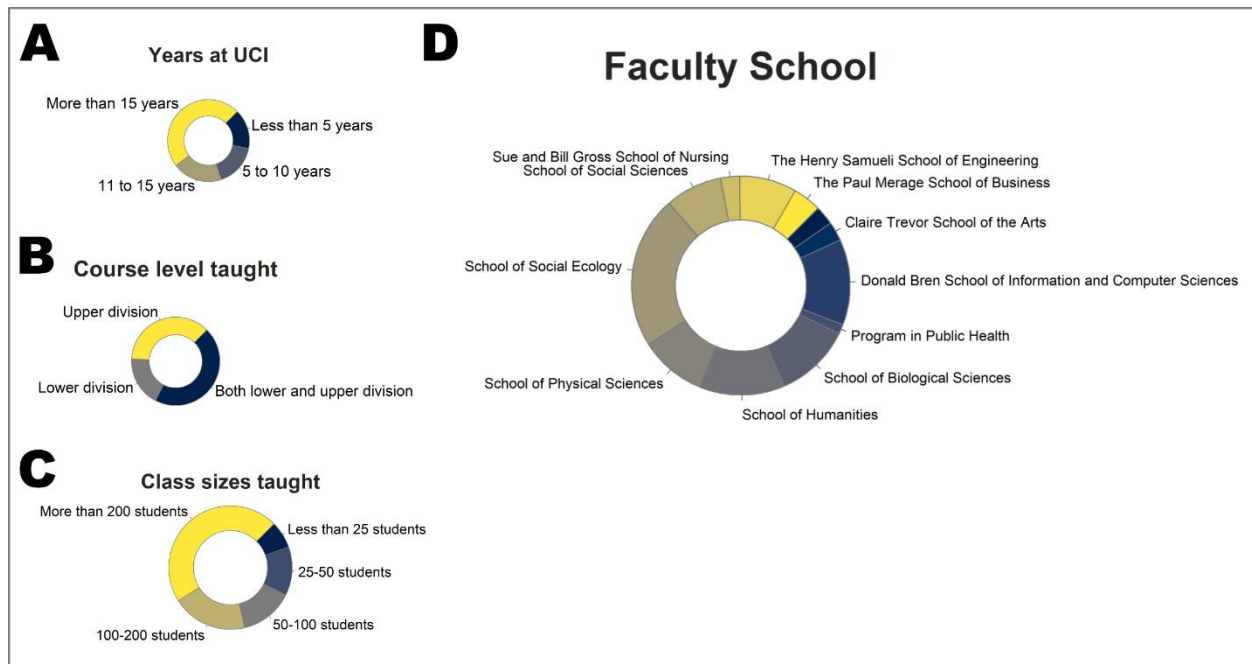
Figure 1: Diversity of faculty respondents. Our survey respondents (**n=72**) were relatively representative of faculty at UCI, based on number of years at UCI **(A)**, course level taught **(B)**, class sizes taught **(C)** and the Schools represented by the survey repondents **(D)**.
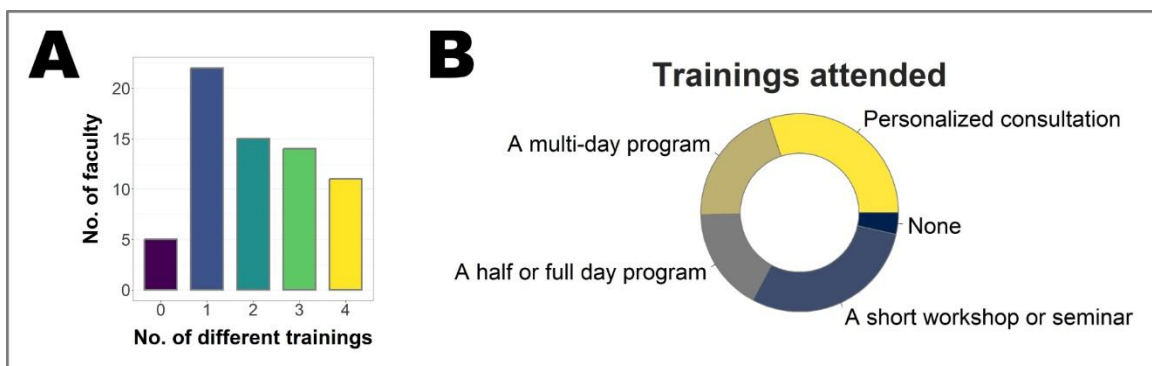

Figure 2: Faculty respondents have formal pedagogical training. The vast majority of respondents (**93%**) had at least one form of pedagogical training **(A)**, and the majority (**56%**) had two or more forms of training (**A**). Faculty indicate that they participated in a variety of different training programs at UCI **(B)**.
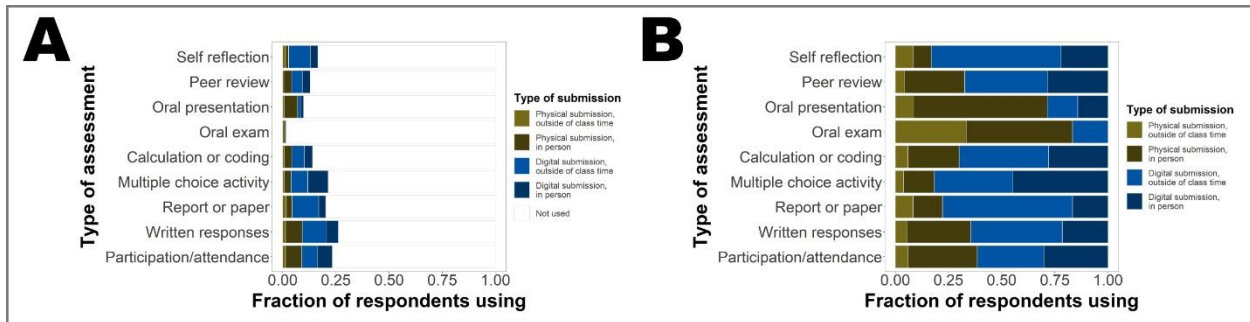
Figure 3: Respondents use a variety of different assessment approaches. Faculty report using several different methods of assessing student work, and no one approach is universal **(A)**. However, for almost all methods (except for oral presentations/exams), faculty use some form of digital submission for the assessment **(B)**.

Figure 4: Respondents do not heavily weight summative exams in determining student grades. Only about 25% of faculty report using exams to determine more than 60% of a student's grade in their class.
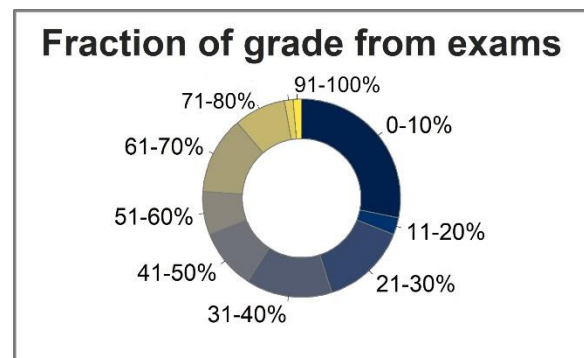
Tables 1-3: Use of heavily weighted exams does not correlate with any obvious factors. Logistic regression analysis showed that using exam scores as the primary determinant of a student's grade does not correlate with class level taught **(Table 1)**, class size taught **(Table 2)** or years of teaching experience **(Table 3)**.

Table 1: Exam use by class level taught

	Odds	p Value
Lower division	0.860	0.790
Upper division	0.942	0.899

Comparison group: Both levels

Table 2: Exam use by class size taught

	Odds	p Value
25-50 students	0.490	0.449
50-100 students	0.876	0.887
100-200 students	2.114	0.402
More than 200 students	1.726	0.491

Comparison group: 0-25 students

Table 3: Exam use by years teaching

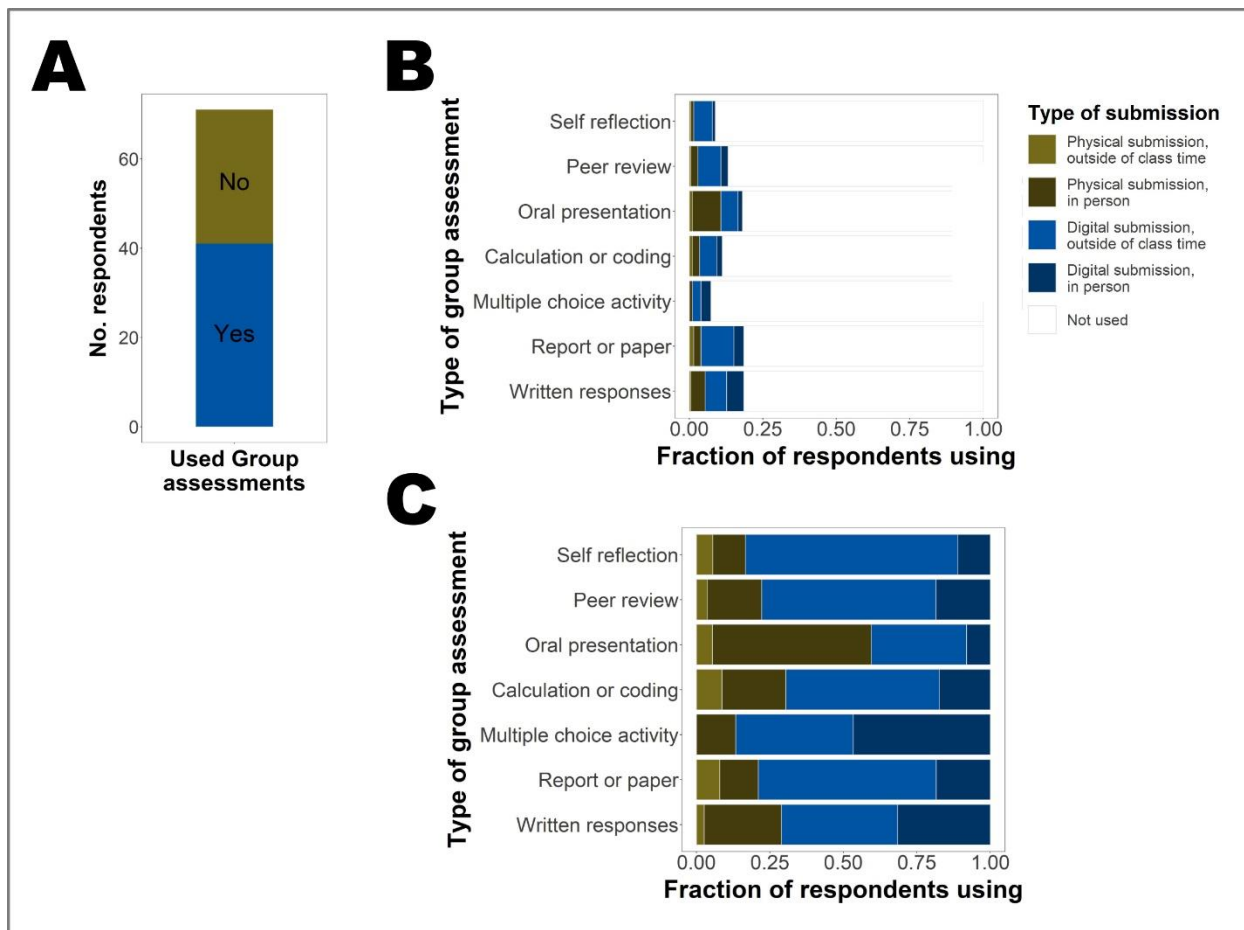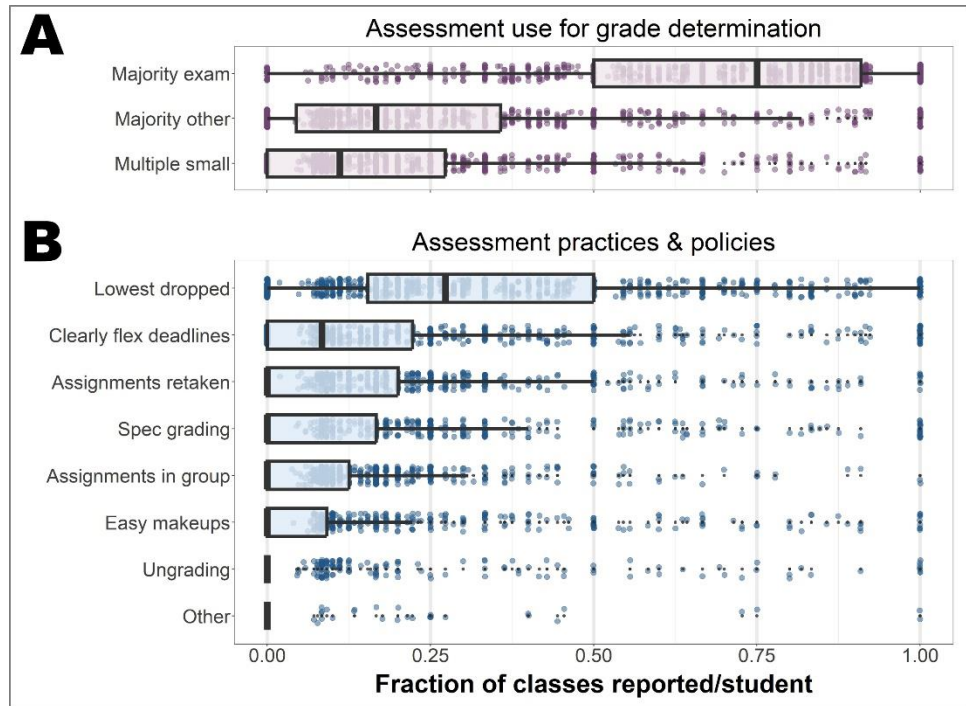|	Odds	p Value
5 to 10	1.026	0.970
11 to 15	2.029	0.294
More than 15	1.295	0.655

Comparison group: 0-5 years

Figure 5: Pattern of usage for group assessments is very similar to that seen for individual assessments. The majority of respondents **(55%)** report using some form of group assessments **(A)** and faculty report using several different methods of assessing group work, and no one approach is universal **(B)**. However, for almost all methods (except for oral presentations/exams), faculty use some form of digital submission for the assessment **(C)**. This is very similar to what is seen for assessment usage patterns for individual assessments.

Table 4: Groupwork use by class level taught

	Estimate	p Value
Lower division	0.595	0.375
Upper division	0.936	0.091

Comparison group: Both levels

Table 5: Groupwork use by class size taught

	Estimate	p Value
25-50 students	−0.629	0.579
50-100 students	0.000	1.000
100-200 students	0.511	0.639
More than 200 students	−0.223	0.819

Comparison group: 0-25 students

Table 6: Groupwork use by years teaching

	Estimate	p Value
5 to 10	0.134	0.879
11 to 15	0.357	0.679
More than 15	−0.677	0.343

Comparison group: 0-5 years

Tables 4-6: Use of group assessments does not correlate with any obvious factors. Logistic regression analysis showed that use of group assessments does not correlate with class level taught **(Table 4)**, class size taught **(Table 5)** or years of teaching experience **(Table 6)**.

Figure 1: Students report the majority of classes use heavily weighted exams and low adoption of best practices. Survey respondents (**n=922**) indicated that almost 75% of their classes had exams that counted for the majority of their grade **(A)**. They also reported that adoption of many best practices in terms of assessment policies was relatively low, with the dropping of the lowest assignment score being one of the most used practices **(B)**.

Figure 2: Student concerns about cheating. Students generally are not very concerned about various forms of cheating (including the use of ChatGPT in assignments).
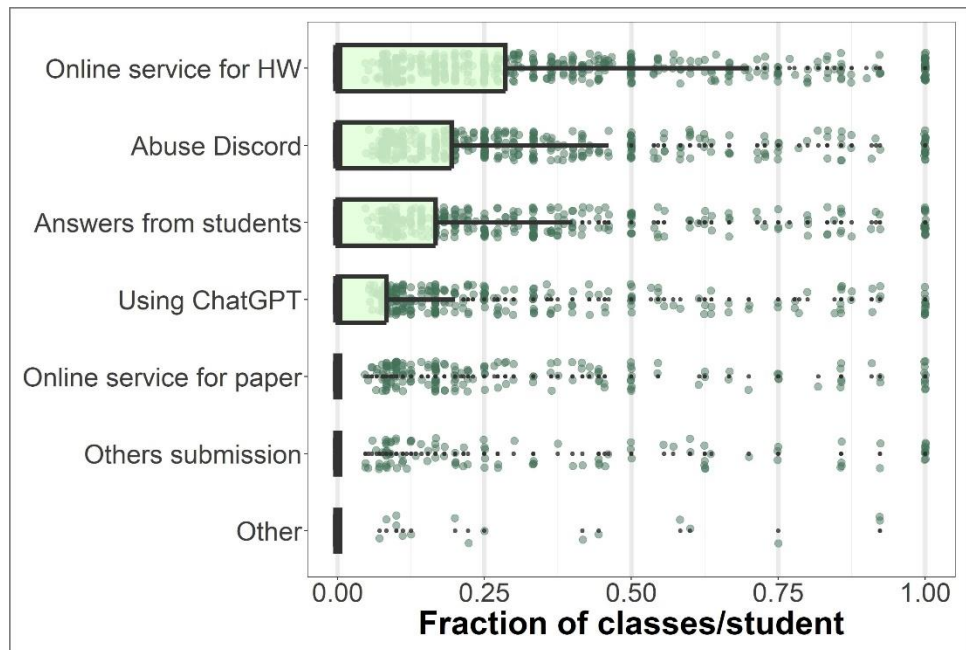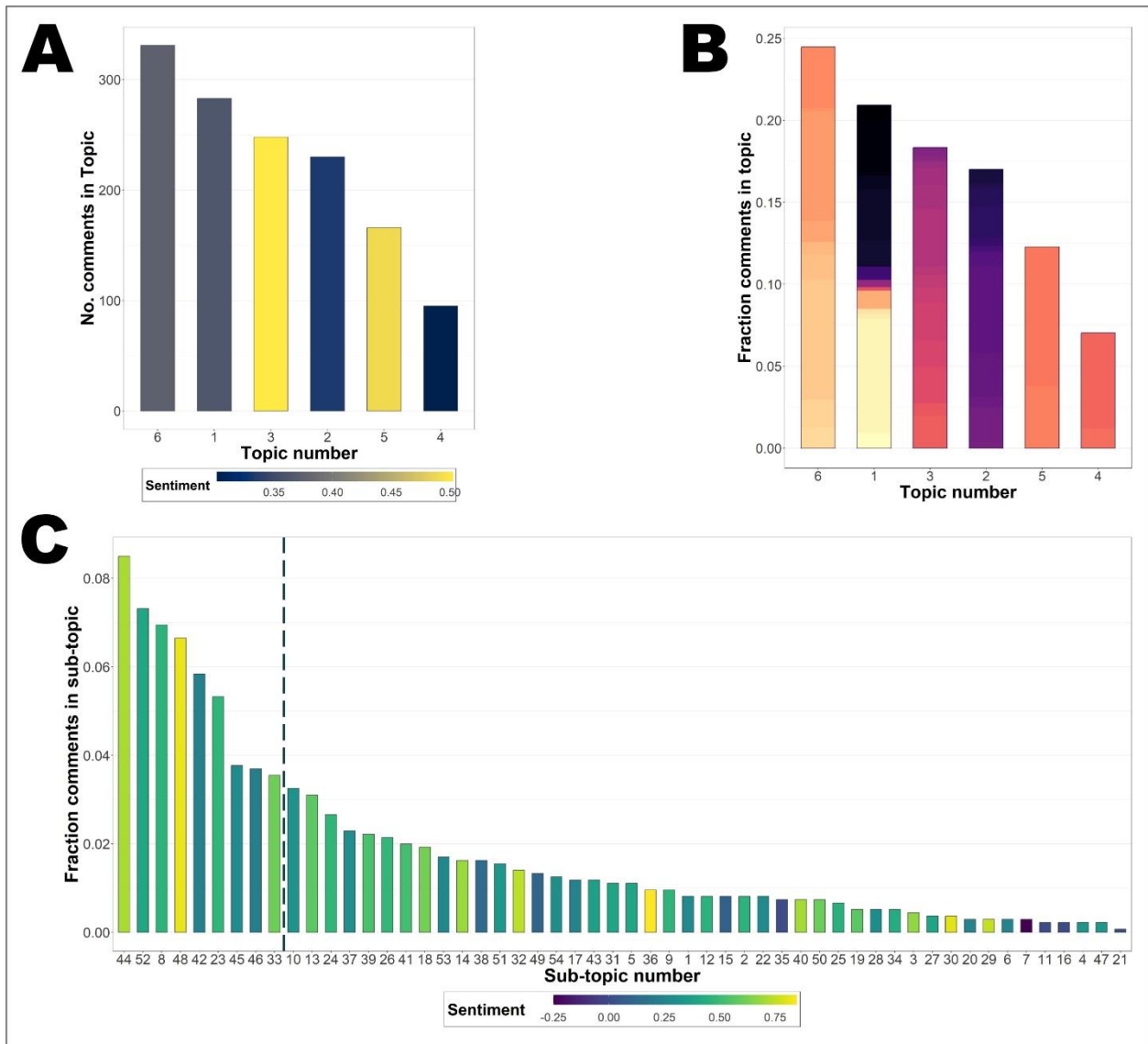
Figure 3: Analysis of student comments about effective assessments. Natural Language Processing (NLP) of student comments about what they consider effective assessments identified six major latent topics with a variety of sentiments associated with the topics **(A)**. Further analysis showed that these topics contained a variety of sub-topics **(B)**, and that more than 50% of the comments were encapsulated by only 10 sub-topics **(C)**.

Table 1: Major sub-topics from student comments and examples of each.

Summary	Example
Course was effective at measuring learning (knowledge and skills)	*The assessments was a great way for us to measure our knowledge.*
Multiple quizzes, test on what was in lecture	*My XXX course had very fair exams, very similar to what we learned in class.*
Assessments based on application, not just memorization	*This helped me understand the concepts rather than focusing on route memorization.*
Continuous and multiple modes of assessment	*XXX taught by XXX did a good job of testing our understanding of material through weekly pop quizzes and a cumulative final essay.*
Had midterms and final	*It has four midterms, lowest one is dropped, no final.*
Tested on what was covered in lecture	*The exams were fair and were materials that the professor has taught in class before.*
Tested on what was covered in lecture	*What we learn and are expected to study is exactly on the assessments.*
Importance of homework	*The assignments were hw every week at the end of the week and going over problems in class.*
Assessment of understanding, thinking and skills	*I felt like his test covered everything in his lectures and my ability to use critical thinking on his tests.*
Testing on application	*His exams were based on detailed data analysis where you had a couple days to analyze without knowing questions.*

FATE Assessment Group – Appendix C

Assessment has been described as "the core business of universities," but it is also considered by many education scholars to be "ineffectual, limiting, irrelevant, and blatantly unfair" (Flint and Johnson 2011, 12). "Assessment" is generally defined as a broader concept than "grading" (Suskie 2009, 10), but narrower than "evaluation" (Felder and Brent 2016, 155). Assessment can—and, arguably, should—be conceived across a curriculum or even a student's complete university experience or entire education, but many faculty and students tend to apply it only to a single course (Suskie 2009, 58).

Assessment is powerful. Assessment drives how curriculum is defined, what students learn, how they learn it, how they are motivated, how they spend their time, and how they judge their achievements as students (Flint and Johnson 2011, 8).

Studies show that many students perceive university assessment as unfair. The perception of unfairness has many sources across higher education, but perhaps the most fundamental issue is that students believe they are not adequately informed of what is expected of them (Flint and Johnson 2011, 38).

Assessment can have many purposes and benefits, but, again, faculty and, especially, students tend to focus on one: the awarding of grades, certificates, degrees or other "warrants" (Knight 2007, 72; Fink n.d., 5). This is a limitation not merely because grading is narrower than assessment. Rather, education scholars argue persuasively that grades and other "warrants" are often insufficient assessments because they incompletely communicate what the student has learned (Knight 2007, 74; Suskie 2009, 10). Indeed, some scholars go so far as to question universities' commitment to warranting in general. They suggest instead that, rather than treating warranting as the fundamental "product" that society expects from them, universities should warrant selectively, warranting only those learning achievements that they believe they can assess reliably (Knight 2007, 76-84).

Many education scholars argue that truly accurate assessment is not possible (Suskie 2009, 37). For this reason, education scholars concede that in most "normal teaching and learning situation[s]" faculty will not be able to know whether their assessment practices are having "a positive impact on learning" (Boud 2007, 19). However, scholars argue that "sufficiently truthful" assessments can still inform pedagogical decisions (Suskie 2009, 38).

Education scholars distinguish "contemporary" from "traditional" approaches to assessment (Suskie 2009, 4). "Contemporary" approaches are understood to reflect similarly "contemporary" approaches to education in general, such as, most importantly, "a new paradigm" —the shift of emphasis from "teaching" to "learning" (Suskie 2009, 58). In the broadest sense, "contemporary" assessment would have the following attributes:

- alignment of assessment with learning goals;
- an emphasis on skills rather than memorized knowledge
- evidence-based;
- be used to inform educational improvements as well as assign grades;
- integrated into the mission of the institution (Suskie 2009, 5; Fink n.d., 13).

One overarching normative theory of assessment holds that it should allow students to "demonstrate capability." Students assess whether they have been enabled to demonstrate capability based primarily, but not exclusively on the grade they receive. Other considerations include "procedural" issues, such as access to instructors, the adequacy of feedback, procedures for improving and appealing grades, policies for such things as late work and compliance with rules like word limits, and so on (Flint and Johnson 2011, 37-39).

Assignments

Given the realities of the situation, many faculty and students, while acknowledging in principle the breadth of the notion of assessment, may continue to think primarily about assignments when thinking about assessment. Several guides for "alternative" assessments are available (e.g., Elkhoury 2022; Gordon n.d.). Education scholars emphasize the following broad points for faculty seeking to improve assignments in a manner informed by best practices:

- align assignments with learning goals;
- justify the time assignments will require;
- break large assignments into smaller ones;
- address academic honesty consistently (Suskie 2009, 155-164);
- vary assignments forms and consider "non-traditional" forms (Suskie 2009, 155-164; Flint and Johnson 2011, 133);
- convey expectations;
- make assessments relevant to skills and careers;
- empower students (Flint and Johnson 2011, 133);
- count all assignments toward the final grade;
- inform students about test-preparation and test-taking strategies;
- provide feedback on assignments as quickly as possible;
- scale grades, rather than curving them;
- address class-wide poor performance on assessments (Felder and Brent 2016, 164-183).

References

Boud, David. 2007. "Grading, Classifying and Future Learning." In *Reframing Assessment as if Learning Were Important*, edited by David Boud and Nancy Falchikov, 14-25. London: Routledge.

Elkhoury, Eliana. 2022. "A Guide to Alternative Assessments." Last Modified Jan. 31. alternative-assessment.com.

Felder, Richard M., and Rebecca Brent. 2016. *Teaching and Learning STEM: A Practical Guide*. San Francisco: Jossey-Bass.

Fink, L. Dee. n.d. *A Self-Directed Guide to Designing Courses for Significant Learning.* https://www.bu.edu/sph/files/2011/06/selfdirected1.pdf.

Flint, Nerilee R.A., and Bruce Johnson. 2011. *Towards Fairer University Assessment: Recognizing the Concerns of Students*. London: Routledge.

Gordon, Damian. n.d. *The Hitch-hiker's Guide to Alternative Assessment.* http://www.damiantgordon.com/Guide.pdf.

Knight, Peter. 2007. "Grading, Classifying and Future Learning." In *Rethinking Assessment in Higher Education: Learning for the Longer Term*, edited by David Boud and Nancy Falchikov, 72-86. London: Routledge.

Suskie, Linda. 2009. *Assessing Student Learning: A Common Sense Guide*. 2nd ed. San Francisco: Jossey-Bass.