G1 Fairness in Blended Assessment in Higher Education – A Quantitative Analysis of Students’ Perception

Anne Jantos¹, Sebastian Hüttemann²
¹ Technische Universität Dresden
² Hochschule für Wirtschaft und Recht Berlin

1 Introduction
In higher education, the current method of awarding grades and degrees through summative evaluation is didactically outmoded, leaving students frustrated and dissatisfied (Jantos 2021; Traub & MacRury 1990). Educators in higher education have been reprimanded for failing to adequately address new challenges while continuing to model traditional lecture-based knowledge transfer instruction (Alt 2018; Assen et al. 2016; Alt 2014). As a result, there is a demand for teacher education programs to provide didactics that will improve teachers’ abilities to apply sensible course design and assessment strategies, and so provide their future students with lifetime learning skills that they can use throughout their careers (Major & Mulvihill 2018).

The curriculum in higher education offers mainly summative examinations (Andrade 2010). Yet, research shows that how students interpret results from assessment influences the nature of resulting cognitions, emotions, and behaviour (Brown 2011). Therefore, by assessing students solely after a course deprives teachers and students of the opportunity to use the effects of the feedback given by any assessment method. Especially self-regulation and self-efficacy cannot be addresses sensibly by summative assessment alone (Boud & Falchikov 2006). Yet, it is necessary for both achievement and motivation (Andrade & Heritage 2018). Teachers in higher education aspire to develop strategies to combine assessment methods and the usage of virtual tools to achieve fair and objective assessment strategies (Joshi & Klausner 2022; Gupta et al. 2020; Care et al. 2018; Griffin & Care 2015), for contextual assessment to address higher taxonomy levels (Zoller 2012), for guidance to conduct sensible assessment and improve learning output and to innovate assessment strategies generally (Kaup et al. 2020).

Our overarching goal is to enable teachers to use and combine online technologies to provide more flexibility and individuality in assessment to address the diversity of learners in higher education and therefore use digital transformation to reach a broader audience and break down barriers between teachers and learners.
Formative assessment, in contrast to summative assessment, has a transformative effect on the learning process since it allows for feedback throughout the learning process rather than just at the conclusion (Andrade 2010; Andrade & Cizek 2010). However, because this requires a significant amount of effort from the educator, it cannot be done thoroughly due to a lack of time and resources. A blended approach combining formative and summative assessment and the use of digital media to measure students’ performance in a meaningful and thorough way could be the answer. Approaches to this strategy are now being investigated in a variety of courses, but there is no standard system for constructing evaluation forms and suggesting a combined approach (Bazvand & Rasooli 2022; Zlatkin-Troitschanskaia et al. 2019; Flores et al. 2015; Heritage 2007). It is well recognized that expanding assessment options would boost student engagement and satisfaction (Pereira et al. 2022; Ragupathi & Lee 2020; Schütze et al. 2018). Yet, there is a lack in understanding the impact of combining different assessment methods to foster learning results (Vander Schee & Birrittella 2021; Flores et al. 2015; Heritage 2007).

We found that perceived fairness is regarded as a central component for motivation and can prevent attempts to cheat during exam situations (Azizi et al. 2022; Baniasadi et al. 2022; Andrade 2010). However, we have yet to discover which combinations of different assessment methods present the most fairness, as well as how educators might be provided with information to help them create and implement courses. This leads us to the question: RP – Which assessment combinations are perceived as fair by students?

2 Methodology
We present here a quantitative study analysing students’ perception of fairness in blended assessment. We researched and identified a variety of different assessment methods. Based on this work, a framework for enriching the course design was proposed: The Blended Assessment Cube, which now needs to be evaluated. To measure the perceived level of fairness for specific assessment method combinations, we conducted a survey with 52 students from German universities. Using the Blended Assessment Cube, we proposed six combinations of different assessment methods and presented them to the students to rate them. The collected data were statistically analysed and interpreted to design guidelines for the application of the Blended Assessment Cube. Figure 1 provides an overview of the research process.
Blended Assessment
Summative evaluation, according to Shepard (2006), should meet its primary purpose of documenting what students know and can do. If properly developed, it should also serve the additional aim of giving learning assistance. If the test’s content, format, and design give a sufficiently rich domain representation, studying for the summative test can be a beneficial learning experience (Shepard 2006). According to further research, taking a test can help you remember more and forget less by reinforcing the presentation of information you learned during the test (Rohrer & Pashler 2010). The results of summative evaluations are commonly used to compare the overall value of an educational program to various alternatives. Formative assessment, according to Bloom (1969), is used to provide feedback at various stages of the teaching-learning process, whereas summative evaluation is used to determine what a student has completed at the end of a course or program. As a result, formative assessment is not a test, but a process (Popham 2006) used in class by teachers and students to provide feedback on ongoing teaching and learning to improve students’ achievement of desired instructional outcomes (McManus 2008) and adjust instruction to the needs of the students (Black & Wiliam 1998). Nonetheless, we find that both assessment methods have distinct merits and that it needs both summative and formative assessment to achieve all levels of Bloom’s Taxonomy (Adams 2015) and that providing mixed assessment fits the learners’ habits best (Vattøy et al. 2021). We therefore recommend a combination of assessment approaches to enhance performance for instructors and learners because testing formatively alone is not feasible. A well-thought-out combination can complement the respective disadvantages of one method with the advantages of the other and thus lead to a holistic assessment experience for students.
Blended Assessment Cube

To assist teachers in putting together appropriate assessment combinations, we devised the Blended Assessment Cube, which depicts certain assessment methods that can be combined. We refocused the axes towards the following three dimensions, based on the Blended Learning Cube by Schoop et al. (2006):

- **Personal Dimension**: Addresses whether a person is assessed individually or as part of a group.
- **Physical Dimension**: Addresses whether the learning situation, respective assessment, is taking place virtually or in person.
- **Methodological Dimension**: Distinguishes between formative and summative assessment.

The following figure shows the Blended Assessment Cube:
The purpose and scope of the Blended Assessment Cube is the mapping of learning goals to feasible and fair assessment method combinations in higher education. It provides suitable assessment combinations related to a learning objective and be adaptable with minor changes to fit different teaching contexts regardless of the content, age of the learner, and the number of participants. Teachers can now utilise the Blended Assessment Cube to determine which assessment methods are possible based on the characteristics of their course goals and the general teaching settings. Furthermore, they can review the various assessment forms and select the broadest possible combination to provide students with a full assessment experience that is tailored to their individual needs. A teacher selects evaluation methods based on the parameters of their courses using various qualities or combinations. Teachers who want to offer work in groups in a virtual area and formatively assess their students might use online presentations or speeches. Formative assessment methods like Andrade’s Rubrics matrix (Andrade 2010) help students to examine themselves, each other, or to support teachers in assessing the quality of the learning outcome. The Blended Assessment Cube therefore enables teachers to broaden the evaluation possibilities and allows them to experiment with new tools. The asynchronous techniques provide more flexibility to students and teachers and leaves time and space for activities to deepen the acquired knowledge in face-to-face meetings. It is therefore very much suitable for Flipped Classroom and Virtual Collaborative Learning Arrangements, Case Study work and other group-centred learning activities.

**Questionnaire**

We created a questionnaire to investigate the perceived fairness of selected assessment combinations. To do this, we used the Blended Assessment Cube to create combinations of three methods each. The following six combinations were strategically selected. Combinations 1, 2 and 4 were selected as the most promising as they represent the most variation according to the Blended Assessment Cube. Combination 6 represents the exact opposite with a very low variation. Combinations 3 and 5 show a medium level of variation. The students were then asked to rate how fair they perceived all combinations they had been given. For each combination, a picture with a description was offered. It was additionally described that all group work and diaries would be taking place in a mix of virtual and face-to-face meetings. No further data such as age or gender was surveyed. The following figure shows an overview of the given combinations:
Table 1: Assessment Combinations

The figure also demonstrates how varied the combinations are according to the Blended Assessment Cube. To answer they were given a Likert scale with the following options: totally disagree, disagree, neutral, agree and totally agree. In the second part, the same combinations were displayed for sorting. Using drag and drop, the students could sort the combinations in descending order of fairness. The questionnaire was offered virtually and send out to Master students enrolled in Germany. It was available in English with MS Forms and took about 8 minutes to fill out.
3 Results

52 students answered the questionnaire. In the first part, the six combinations offered had to be assessed according to perceived fairness. The following figure shows the median:

![Figure 3: Median showing the results of the perceived fairness](image)

We created a ranking based on these data. Here we can already identify combinations that tend to be perceived as not fair and those that are perceived as fair. To verify this impression, we now compare the results of the different questions. In the second step, the same combinations were to be sorted according to their fairness. Based on the median we ranked the combinations and compared them with the median rank created by the sorting exercise. The results and the comparison between the two ratings show clear differences here:

![Figure 4: Box Plot perceived fairness](image)

<table>
<thead>
<tr>
<th>Combination</th>
<th>Median according to Likert questionnaire</th>
<th>Rank according to Box Plot based on Likert questionnaire</th>
<th>Average rank sorted by students</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2/3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2/3</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>
The results show the perceived fairest (Rank 1) and the unfairest (Rank 6) combinations are undisputed and unambiguous. There is no agreement in the results for the second to fourth places. However, it was possible to calculate an average order using the average of the calculated ranks. The results show that combinations with a high degree of diversity (Combinations 1,2,4) are perceived as fair. Combinations with similar methods (Combinations 5,6) are considered less fair. We can therefore answer the research proposition: RP – *Which assessment combinations are perceived as fair by students?* The following list shows the assessment combinations sorted descending by perceived fairness:

1. Group Work Assessment + Group Presentation + Learning Process Diary
2. Group Work Assessment + Consultation with Teacher + Written Exam
3. Group Presentation + Learning Progress Diary + Virtual Oral Exam
4. Group Work Assessment + Peer Review by Fellow Students + Oral Exam
5. Group Presentation + Learning Progress Diary + Written Online Exam
6. Written Exam + Oral Exam + Written Online Exam

We additionally identified general statements that describe perceived fairness in the shown assessment combination and inferred indication to implement blended assessment in higher education.

- **Diversity** – Combinations with a high degree of diversity are perceived as fair. Combinations that address all manifestations of the dimensions of the Blended Assessment cube are particularly diverse and therefore particularly fair. Teachers should therefore look for complementary formats that have exactly the corresponding characteristics.

- **One-dimensionality** – combinations that show only one level in the Blended Assessment Cube, and thus only address one manifestation of the dimensions and are therefore not diverse are considered to be particularly unfair.

- **Balance** – Students do not shy away from summative testing. It is the combination that makes the difference, not the individual methods. Formative assessment is not objectively superior, and teachers should not rule out using summative assessment methods to provide fair assessment but rather combine it with opposite methods to find a balance.
Individuality – students disagree about which combination is clearly the best. So, there is no optimal combination that everyone prefers. A teacher can therefore only strive to offer as broad a range as possible to meet the needs of all learners.

In addition, it could be identified that oral examinations are perceived as fairer than written examinations, because combination 2 and 4 differ only regarding this aspect and combination 4 was perceived as fairer than combination 2. Yet this is contradicted in observation of combinations 3 and 5. Here we see that students perceive online written exams fairer than virtual oral exams. This contradiction cannot be explained by the data gathered here.

4 Discussion
The results of this analysis show that students particularly perceive those combinations as fair that show a high level of diversity. Despite the clear inconsistencies in the data, we were able to formulate statements that provide information about the perceived fairness in assessment in higher education teaching.

It should be a concern of every university and every teacher in higher education to offer fair assessment to all students and thus also to promote student success through individualization and flexibilization of teaching. But teachers in higher education might feel lost in the virtual opportunities and methods (Coman et al. 2020) and asking for help or making use of trainings needs resources and courage. However, technical abilities alone are insufficient. Teachers must also adapt their teaching approaches to the online world (Coman et al. 2020). The Blended Assessment approach offers guidance to teachers and can promote improved assessment strategies and help to enable teachers to use this extension of assessment to improve their course design in the long term.

However, this support is only suitable for smaller groups of learners. Large events, such as lectures with 400 participants, can only be supported by blended assessment with difficulty and under special circumstances because it requires a great deal of effort on the part of the teacher to provide formative assessment for large groups, and most university teachers do not have assistants or technical and didactical support (Griffin & Care 2015).

Another limitation of this research is the homogeneity of the respondent group. Only students who are enrolled in Germany are surveyed. They may well have a different cultural background, but it can be assumed that they have primarily western cultural backgrounds. Furthermore, they are primarily enrolled in economics courses and are already more advanced in their studies, because they are currently studying for a master’s degree. Experienced learners may have different needs for fair assessment than inexperienced learners. This study does not consider this aspect.
Nevertheless, the present findings can be used to make a decisive contribution to the creation and implementation of blended assessment at universities. Both course design and the use of e-portfolios in existing courses can be structured and prepared in a more targeted manner using the insights gained here. Moreover, with the perspective of the students gained here, it can also be argued in the long term that portfolio examinations and other complex, combined performances should be expanded and offered more frequently if the goal of the university is to create a fair examination situation.

Literature


