An Automated Feedback System for Written Exams

Stefan Ullmann and Mareike Schoop
University of Hohenheim, Stuttgart, Germany

Stefan.Ullmann@uni-hohenheim.de
Schoop@uni-hohenheim.de

Abstract: Summative assessments, particularly exams, are the dominant method for evaluating student performance in higher education. Although feedback is the single most influential factor in promoting learning, exams are often excluded from the good practices highlighted in literature. The gap stems from challenges such as time constraints, large student numbers, and institutional policies. While automation could improve feedback processes, most current automated feedback systems (AFS) are teacher-centred and do not align with good feedback practices. The present paper uses a design science research approach to develop a student-centred AFS for exams in different domains. The system provides a structured setup process to guide lecturers in providing high-quality feedback. We utilise expert knowledge in the form of Bloom's taxonomy and task types, as well as student data such as points, to generate feedback for each student providing a performance overview and suggestions how to improve their exam performance and learning strategies. In addition, the system delivers detailed feedback on topics on which the student performed poorly. In an initial evaluation of the student-centred AFS, three lecturers successfully used the AFS to generate feedback reports to 1323 students from two large-scale bachelor courses. Survey results indicate that the student-centred AFS delivers high-quality, timely, and personalised feedback at scale, helping students to adapt learning strategies and to identify deficits. The present work thus contributes to solving the challenges of feedback in higher education in general and to solving the exam-specific feedback gap. Additionally, our AFS that is easy to adapt for non-experts. We demonstrate the various design features required for such an AFS, including an adaptable domain model utilising Bloom's taxonomy and customisable task types, to ensure applicability across diverse educational domains.

Keywords: Constructive Feedback, Assessment Feedback, Automated Feedback System

1. Introduction

Assessment is of vital importance for learning and teaching in higher education (Mekonen and Fitiavana, 2021; O'Donovan, Rust and Price, 2016). Apart from suggestions to replace traditional assessment methods, summative assessments are still the most common assessment method (Alquraan, 2012; Rawlusyk, 2018; Ray et al., 2018).

The process of assessment using exam, consists of setting a learning task for the student to perform followed by grading the performance and providing the student with their mark. Assessment feedback is an effective means for students to improve their performance and future learning (O'Donovan, Rust and Price, 2016). However, feedback practices, contextual constraints and individual challenges can be challenging (Henderson, Ryan and Phillips, 2019).

From the students' perspective, quality of the comments is often problematic in *feedback practices*. Students ask for more and specific feedback and suggestions for improvement. Lecturers are concerned *with contextual constraints* including time constraints and the scalability of feedback practices especially for large courses which leads to students not receiving well-crafted and personalised feedback. *Individual challenges* include the lecturers' expertise and understanding of the student. There are also some specific challenges regarding written exams (Henderson, Ryan and Phillips, 2019; Scoles, Huxham and McArthur, 2013). The distribution of marked exams as feedback prevents their re-use. Furthermore, students could receive assessment content from previous students, limiting their own revision of the material (Halinen et al., 2014; Medina and Yuet, 2013; Ray et al., 2018).

The feedback process could benefit from technological innovations (Haughney, Wakeman and Hart, 2020). Automating some forms of feedback could save time for lecturers and thus decrease their workload (Debuse, Lawley and Shibl, 2008). Many of the current automated feedback technologies are teacher-oriented whereas students would potentially profit from a more student-oriented approach adaptable to their needs (Deeva et al., 2021).

The overall goal of the present paper is to develop and evaluate a student-centred automated feedback system for exam feedback which overcomes the challenges.

2. Theoretical Background

The relevant theoretical background in learning, assessment and feedback will now be discussed.

2.1 Bloom's Taxonomy

Bloom's taxonomy classifies mental learning and skills into six cognitive levels, as shown in figure 1, (Bloom et al., 1956; Scott, 2003) which we use to categorise exam questions in a hierarchy of complexity (Anderson and Krathwohl, 2001; Omar et al., 2012).

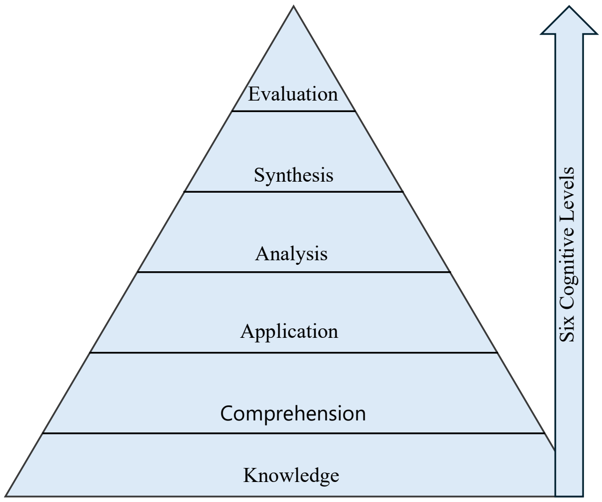

Figure 1: Six cognitive levels based on Bloom at al. 1956

The *knowledge* level refers to the memorisation. *Comprehension* refers to the meaning of information and explaining, classifying or interpreting. The *application* of concepts to specific scenarios means students understand and apply a concept. *Analysis* refers to breaking down information into simpler parts and analysing them including inferring relationships. *Synthesis* is about integrating and combining ideas or concepts into a new whole. E*valuation* alludes to rating and judging the values of ideas for given purposes (Bloom et al., 1956; Omar et al., 2012; Scott, 2003; Starr, Manaris and Stalvey, 2008).

2.2 Assessment and Feedback

Assessments are performed by lecturers to collect information about a student's performance and achievements (Gronlund, 1998). Formative assessments are carried out continuously throughout a course (Jacoby et al., 2014; Sambell, McDowell and Montgomery, 2012), whereas summative assessments evaluate a student's performance at the end of a course, e.g. in written exams (Gibson and Shaw, 2011; Gronlund, 1998). Summative assessments can be combined with supportive feedback and can thus work formatively (Carless, Joughin and Liu, 2006).

Good feedback needs to answer three questions to be effective for the learner: "Where am I going?" (clarifying the intended goal or task), "How am I going?" (informing about the performance in relation to the learning goal) and "Where to next?" (promoting continuous learning and improvement of performance) (Daka et al., 2021; Hattie and Timperley, 2007); it should provide useful information, be specific and personalised (Higgins, Hartley and Skelton, 2002), address the unique needs and contexts of the learner, be timely (James, McInnis and Devlin, 2002), consistent (Holmes and Smith, 2003), legible and detailed (Nicol, 2010) to foster understanding and facilitate improvement (Higgins, Hartley and Skelton, 2002).

3. Related Research

Automated feedback systems (AFSs) can both grade assessments and generate feedback automatically. However, most AFSs focus on the grading. Therefore, they mainly decrease the workload of lecturers. A typical AFS deduces the feedback from a domain model, expert knowledge and student data. This is done using a feedback generation model, which specifies the type and recipient of the feedback by using student data and/or expert knowledge. Most AFSs use feedback generation models that are data-driven (32.1%) or expert-driven

(48.6%); only 21% use a mixed approach. The domain model is a structured representation of the domain knowledge (Deeva et al., 2021).

AFSs can be open (allowing lecturers to set key parameters e.g. feedback information) or closed (preventing the change of parameters or process). Complex systems may require considerable technical expertise to adapt or modify the AFS. If an AFS can be adapted by the lecturer, options exist to adapt the feedback to the requirements of the circumstances and students (Shum et al., 2023).

Types of AFSs include learning management systems, recommender systems and intelligent tutoring systems (Schipper, Feskens and Keuning, 2021; Shum et al., 2023). Most AFSs are domain-specific; only 6% are non-domain-specific and therefore usable universally (Deeva et al., 2021). These non-domain-specific AFSs are designed for providing feedback in online assessments (Del Sánchez-Vera et al., 2012) , for providing formative feedback in large classes (Schaffer et al., 2017), and for use as part of a learning system for students (Tass Grigoriou, Christopher Cheong, and France Cheong, 2015). However, none of them is designed for providing feedback on written exams.

Generative AI can be used to generate feedback (McGowan, Anderson and Smith, 2024; Morales-Chan et al., 2024; Park, 2023) but lecturers need to review AI feedback and adjust it if necessary (McGowan, Anderson and Smith, 2024).

4. Research Goal and Methodology

The state-of-the-art shows that most AFSs are not student-centred, focusing primarily on automated grading and reducing lecturers' workloads; they require lecturers to possess technical expertise; they are domain-specific, which can limit their applicability (Deeva et al., 2021; Shum et al., 2023). Consequently, AFSs should support lecturers in implementing feedback element; AFSs should be applicable to exams in different subject areas; AFSs must fulfil the unique requirements of exam feedback, such as ensuring exams are not leaked during the feedback process. To this end, the following research questions are addressed:

- RQ1: Which properties does a student-centred AFS require to provide high-quality scalable feedback across different educational domains?
- RQ2: How can such an AFS be implemented?
- RQ3: How do students assess the clarity and actionable quality of the exam feedback generated by the AFS?

To answer these research questions, we follow a design science approach as shown in figure 2 (Hevner et al., 2004; Hevner, 2007; Peffers et al., 2020). The following chapters provide a detailed description of all the steps necessary to answer the RQs.

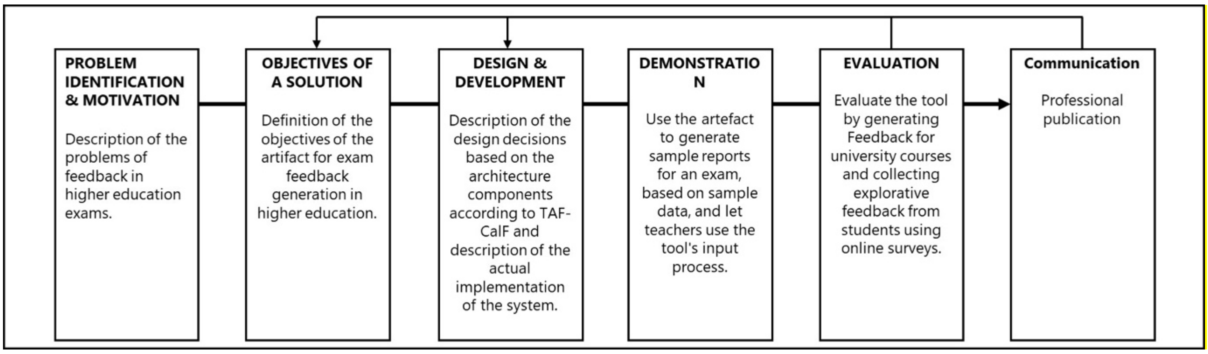

Figure 2: Design Science Process

5. Problem Identification and Objective Derivation

5.1 Problem Identification

A *detailed problem identification* is necessary to derive the objectives and justify the solution; it helps to understand the problem, which is vital for understanding the argumentation presented (Peffers et al., 2020). To this end, a case study in a German university with mostly paper-based exams was conducted. The problem analysis is based on literature, the authors' teaching experience and interviews with three experienced lecturers. The derived problems are numbered as P1 to P6.

Students receive their grade digitally following the grading of the exam and are offered an exam inspection on-site. Such events might take place much later than the publication of grades. Therefore, receiving timely feedback cannot be guaranteed (P1) but timeliness is an important factor for good feedback (Henderson, Ryan and Phillips, 2019; Li and Luca, 2014). As exam reviews are handled by experienced lecturers who can provide good feedback and advice on how to improve the student's learning strategy, this could close the previously described feedback gap. However, only a small number of students take advantage of the exam reviews and a small proportion of attending students ask questions about specific tasks or ways to improve. Furthermore, detailed face-to-face feedback in exam reviews is not feasible for large student groups (P2). Taken together, most students do not receive feedback that goes beyond the grade which is a general problem in higher education (P3) (Scoles, Huxham and McArthur, 2013).

Another important aspect concerns the distribution of graded exams to students which poses a potential threat to the integrity of subsequent exams, as there is a risk that students will not fully engage with the content, resulting in a narrow focus on marks rather than a broad understanding. Effective lecturers often reuse existing test questions (Appelhaus et al., 2023) due to a high workload which is also one of the most frequently cited challenges to good feedback (Henderson, Ryan and Phillips, 2019). Furthermore, there is a need to archive exams for several years depending on relevant legal regulations. Therefore, it is not possible to hand out the original exams as part of the feedback (P4).

AFSs are mostly designed for a specific domain (P5) (Deeva et al., 2021) and can thus not provide domain-independent exam feedback. AFSs are often lecturer-oriented rather than primarily addressing the quality of feedback given to students (P6) (Deeva et al., 2021).

5.2 Objectives of a Solution

Figure 3 summarises which *objectives* (called O1-O8) were derived from which problems. The problem analysis shows the need for a student-centred (O1) AFS usable in various domains (O2) that can provide high-quality feedback to a large number of students in a scalable and timely manner (O3) without the need of exam publication (O4). Such AFS needs to be easily adaptable to different exams (O5) and usable for a wide range of lecturers (O6). As the need for large data sets can make it difficult to use the system for different exams (Keuning, Jeuring and Heeren, 2019), the system should keep the required data simple (O7), to ensure good usability for various lecturers. The primary focus of student-centred AFSs should be the production of high-quality feedback for students (Deeva et al., 2021). Therefore, the AFS needs to ensure that effective feedback is provided addressing the intended goals and the students' performance in relation to these goals. Additionally, the feedback should provide instructions and suggestions on how students can further improve their learning and performance (O8).

Table 1: Mapping of Problems to Objectives

Problems	Objectives
P1: No timely feedback	O3: Scalable, timely feedback
P2: No practical solution for large groups	O3: Scalable, timely feedback
P3: Only exam feedback are often marks	O5: Usable for different exams, O6: Usable different lecturers, O7: Keeping required data simple, O8: High-Quality feedback
P4: No distribution of old exams	O4: No exam publication needed
P5: AFS domain specific	O2: Domain-independent usability
P6: lecturer-oriented/ feedback quality	O1: Student-centred AFS, O8: High-Quality feedback

6. Design and Development

In the following subsections, steps three and four of the design science research process are conducted.

6.1 Conceptual Design

The *conceptual design* is derived from the objectives (cf. section 5.2). The architecture component of the TAF-ClaF with its five essential AFS elements is used as a reference (Deeva et al., 2021). (1) The domain model is a structured representation of the domain knowledge (Deeva et al., 2021). Such model for the described tool must be easily adaptable for different domains. Bloom's taxonomy as an established method for classifying exam tasks in various domains (Rawat, Kumar and Singh Samant, 2023) is thus a good foundation for an easy-to-use and widely applicable AFS. To support lecturers in classification tasks, the AFS offers task types that can be defined by the lecturer (e.g. multiple-choice vs. free-response questions). The tool is thus able to provide feedback with suggestions for further improvement based on the individual student's ability to answer questions with different levels of thinking skills required or based on different task types. The feedback can, therefore, be personalised for each topic and for each student. (2) The expert knowledge used to derive rules for mapping the domain knowledge to student input can be provided by lecturers or educational theories. We use both to give students information for improving their learning strategies. (3) The student data required for feedback generation (Deeva et al., 2021) are the achieved points in each (sub)tasks. (4) The feedback generation model is a set of rules the system provides that uses data, expert knowledge or both to generate the feedback (Deeva et al., 2021). We use the achieved points in the different tasks and subtasks of an exam, combined with the knowledge of different levels of required thinking skills. The general exam feedback informs students of their performance on tasks of a specific thinking skill or specific task types and issues hints about what kind of preparation is important to improve. These hints are general in nature and can be used for all types of exams. At the task level, a description of the learning objectives for each task is provided. The subtasks are grouped according to the levels of thinking skills required or their different task types. A general description of the objectives for each subtask group is provided. Finally, the students are shown their achieved points in each task and subtask in relation to the set points. The lecturers are guided in providing feedback, ensuring that essential information is included. (5) The system must be easy to use. Once set up, our system is able to generate and send a structured feedback report attached to an email to each individual. The system can work offline and does not need to be integrated into a learning system. Figure 4 shows which objectives are fulfilled by which essential element design decisions.

6.2 Development

We *implemented* our AFS using Excel to organise information across several spreadsheets for ease of use. The AFS backend uses decision mechanisms and an algorithm to calculate and rank the relative and absolute point losses for the whole exam and for each task in relation to the classified thinking skill or the type of task, by analysing the points lost by a student in each individual subtask. The length of the feedback report can be customised by specifying the percentage of points that must be achieved in a task to receive task-specific feedback. For each exam, general exam information (e.g., date, title) as well as a threshold for task-specific feedback are provided by the lecturer. Based on the number of (sub)tasks and their topics, the AFS recommends text blocks for general feedback based on Bloom's thinking skill levels, which can be modified. Task-specific feedback includes goal descriptions and suggestions for improvement. The AFS provides a structured set-up process as shown in figure 5, containing examples, tips and instructions to the lecturer. Once all text blocks are complete, student data is entered, the individualised feedback reports and the emails with the reports attached are generated and sent.

Table 2: Mapping of Objectives to Design Solutions

Objective	Essential Elements	Design Solution
O1: Student-centred AFS	(1), (2), (4)	Blooms taxonomy/ task types, student points and the lecturer knowledge enable generation of personalised feedback
O2: Domain-independent usability	(1), (3)	Blooms taxonomy/ task types and student points are not domain specific, widely used and known from many lecturers
O3: Scalable, timely feedback	(5)	Automated generation and distribution without manual work after setup
O4: No exam publication needed	(1)	The use of task classification makes it possible to generate personalised feedback without a need for task/ exam publication
O5: Usable for different exams	(3), (4)	The adaptable domain model combined with easily available student data makes the tool usable for a wide range of exams
O6: Usable different lecturers	(1), (3), (5)	Domain model and student data are easy to use/ available. The structured, assisted set up process makes it easy usable
O7: Keeping required data simple	(3)	Point of the tasks/ subtasks
O8: High-quality feedback	(1), (2), (3), (5)	The structured and assisted setup process together with the student data and domain model ensure actionable, personalised feedback on the overall exam and on task level

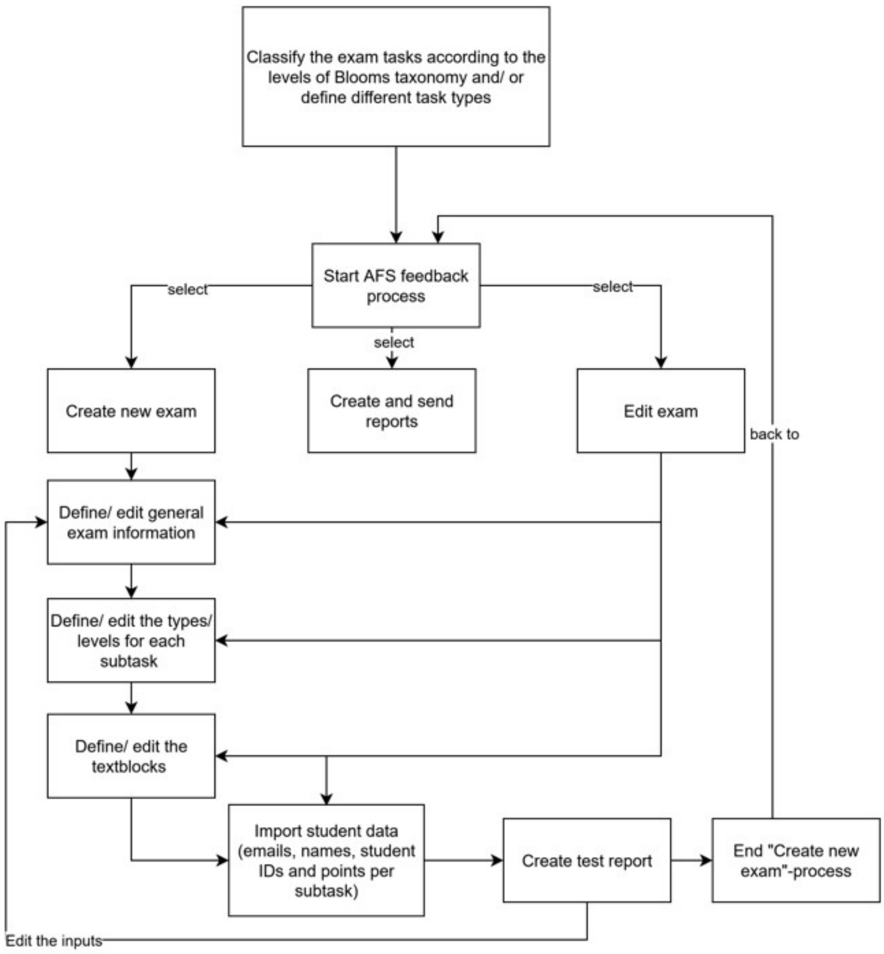

Figure 3: Setup Process Diagram

6.3 Demonstration: Feedback Report

A generated example feedback report is shown in figure 6. The personal address to the student is followed by general information about the exam, details of the grade, the total number of points achieved and the number of points necessary for passing the exam. The exam overview table provides an overview of the task information including the achieved points in relation the set points. The rows of each task are coloured green, yellow and red depending on the percentage of points achieved. This is done to make the report more legible. The second page of the report contains general feedback on the student's performance based on the overall exam performance for the different thinking skill levels or task types. It contains a description of the task type/thinking skill level and hints on how to improve preparation and performance. The feedback includes both general and task-specific comments. Currently, students receive feedback for tasks where less than 80% of the points are achieved; this number can be adapted for each exam. The detailed task-specific feedback contains a general description of the task goal in the exam, followed by descriptions and learning aid for the sub tasks of thinking skills/task types where the student lost points. For tasks with more than 80% of the points, appraisal feedback is provided.

Exam Report – XXX

Dear XXX XXX, this is your personal report for the XXX exam (Business Informatics/XXXX-XXX), which you took on XX.XX.XXXX (semester XX, supplementary examination period).

The report provides you with detailed information about your performance and potential areas for improvement. You achieved a total of 36.5 points (61%) out of 60 points. **Your grade: 3.3** The pass mark for the exam is 30 points (50%). The failure rate for the exam is 20.97% with an average grade of 3.36.

Notes on the table: Columns 1 and 2 show the task, the associated subtasks and the corresponding subject area. Columns 3 and 4 provide information about the maximum number of points that can be achieved and the number of points you have achieved. The last column shows the percentage of points you achieved in the subject area. The colour coding helps you to assess your performance: Red: < 50%; Yellow: ≥ 50%; Green: ≥ 80%

Tasks	subject area	achievable points	Your points	Points %
2	ARIS EPK	5	3	60%
2.1		2	0	
2.2		1,5	1,5	
2.3		1,5	1,5	
3	Organisational/ functional perspective	5	3	60%
3.1		2	0	
3.2		1,5	1,5	
3.3		1,5	1,5	
4	Excel	8	3,5	44%
4.1		1,5	1,5	
4.2		1,5	0	
4.3		1	0	
4.4		2,5	1	
4.5		1,5	1	
5	ARIS table scheme	5	3	60%
5.1		2	0	
5.2		1,5	1,5	
5.3		1,5	1,5	
6	xHTML	7	2,5	36%
6.1		1,5	1,5	
6.2		1,5	0	
6.3		4	1	
7	procedure models	5	1,5	30%
7.1		2	0	
7.2		1,5	1,5	
7.3		1,5	0	
8	ARIS ER-Model	20	16	80%
8.1		6	6	
8.2		7,5	7	
8.3		6,5	3	
9	SQL	5	4	80%
9.1		1,5	1,5	
9.2		1,5	1,5	
9.3		2	1	
	Total	60,0	36,5	61%

General feedback:

You were particularly successful in answering questions that required the application of known procedures and methods to solve a problem from a known problem pool. Therefore, understanding the content and possible practice questions are crucial for questions of this type. You achieved 71% of the points for questions of this type.

You have the greatest potential for improvement in tasks that require the reproduction of learned knowledge to answer. This means that you already know the content being tested from the lecture slides, e.g. theoretical basics. You must then reproduce this correctly in the tasks. You achieved 53% of the points in tasks of this type.

Task-specific feedback:

You achieved a high score (>80%) in the subject areas ARIS ER model and SQL. Very good!

Below are some detailed tips on the areas where you have the most potential for improvement.

You still have considerable room for improvement in the area of Excel. In the task for this topic, you should demonstrate your theoretical knowledge of Excel and apply formulas to solve specific problems using an example table.
You lost most of your points in this task, 3 out of a total of 4.5, on tasks that involved using formulas you learned in the exercises. You should therefore study the exercises in greater depth. In particular, practise working with Excel in a practical way, as this can be helpful for applying your knowledge in the exam. You also lost points, 1.5 out of a total of 4.5, on tasks in which you were required to demonstrate theoretical knowledge already familiar from the lecture. To improve in these areas, you should familiarise yourself more thoroughly with the basics of Excel and the content of the slides.
In this subject area, you lost points on tasks that included the following topics: Excel basics, function: HLOOKUP, function: DATE, function: CONCATENATE, function: SUM.

You still have considerable room for improvement in the area of xHTML. In the task for this topic, you should be able to independently create a given section of a website, find and correct errors in XHTML pages, and outline the content represented by an XHTML expression.
You lost most of your points in this task, 3 out of a total of 4.5, on tasks in which you had to apply expressions in XHTML and CSS that you learned in the lecture to a problem, for example, to reproduce a given section of a web page in the source code. You should therefore familiarise yourself with the practical use of XHTML and CSS using the exercises or examples of your own choosing. All you need is an editor or Notepad and your existing browser. Create a web page and check whether the display in the web browser meets your expectations, then add further content to your web page. You can use the W3C's XHTML validator to check whether you have used XHTML syntactically correctly. You also lost points, 1.5 out of a total of 4.5, on tasks in which you had to reproduce knowledge of XHTML learned in the lecture. To improve in these areas, you should take a thorough look at XHTML expressions and familiarise yourself with the structure of various types of lists, tables, image embedding, email links, etc. It can sometimes be helpful to remember the long name, such as 'table data' in this case, instead of abbreviations such as 'td'. The same applies to the structure and use of CSS files and the rules they contain.
You lost points in this subject area on tasks that included the following topics: lists in xHTML, CSS and document types, writing xHTML.

Figure 4: Example Feedback Report

7. Evaluation and Conclusion

To analyse its usefulness, the AFS needs to be *evaluated*.

7.1 Evaluation and Discussion

The AFS was used to generate feedback for exams of 1323 bachelor students in two different courses. A survey assessing the tool was completed by 65 students, cf. figure 7. Note that not all questions were answered by all of the 65 students. The survey utilised a 7-point Likert scale (ranging from 1 - 'I totally disagree' to 7 - 'I totally agree') and included two yes/no questions.

Table 3: Evaluation results

	n	Mean	Std. Deviation	Median
The design of the report is clear	64	6,4	1,0	7
The report and the information provided are easy to understand	65	6,3	1,2	7
The information in the report is selected sensibly	64	6,2	0,9	6
The information in the points overview table helped me to assess my performance	64	6,3	0,9	7
The task-specific feedback helped me to recognise my learning deficits	60	5,8	1,3	6
The task-specific feedback helped me to adapt my learning strategies for future exams	59	4,8	1,6	5
I would benefit from an exam report in other exams	65	6,4	0,9	7

The students found the design report clear (M=6.4; SD=1.0; Mdn=7), the information understandable (M=6.3; SD=1.2; Mdn=7) and sensibly selected (n=64; M=6.2; SD=0.9; Mdn=6). For 73.8% of the students the report had an adequate length. The remaining students found it too long which might be due to different reports having different lengths, depending on how the student performed (yes/no question). The points overview table helped them to assess their performance (n=64; M=6.3; SD=0.9; Mdn=7); the task-specific feedback helped them to recognise learning deficits (n=60; M=5.8; SD=1.3; Mdn=6). Most of the students agreed at least partially that the feedback report helped them to adapt their learning strategies for future exams (n=59; M=4.8; SD=1.6; Mdn=5). The students replied that they would benefit from such reports in other exams (M=6.4; SD=0.9; Mdn=7). 52,3% of students stated that the report would influence their future exam preparations which is a significant success given the report was the first one they received.

In a first evaluation of the setup process, three lecturers used the AFS for two different exams. All lecturers were able to set the AFS up and to provide high-quality feedback for their large courses and valued the structured process. They mentioned the still time-consuming process of text block definition.

AFSs can use several complex student data and complex techniques to generate feedback (Deeva et al., 2021). To design an AFS that can provide feedback across different educational domains it is important, to use student data that is available in different educational domains and design a system that enables the lecturer to modify the feedback to make it appropriate for their exam. Whilst some existing non-domain specific systems also use Blooms taxonomy (Barker, 2010), we added task types, that can be defined by the lecturer to ensure the AFS suits the needs of the particular exam; furthermore, the AFS is usable even if the lecturer is not familiar with Blooms taxonomy.

Whilst some AFSs need programming skills or technical expertise to be modified (Karkalas et al., 2016; Shum et al., 2023), our system is usable without technical expertise and even supports lecturers during its setup process, as our first evaluation of the setup process shows.

There are several possibilities to implement an AFS, using programming languages like C++ or tools like Rapid Miner (Deeva et al., 2021). We decided to use only tools commonly available in the education field, (namely Microsoft Word, Excel and Outlook), to ensure the AFS is usable for all lecturers.

The evaluation using student surveys as the most common type of evaluation AFSs (Deeva et al., 2021), demonstrates that the generated feedback is clear and understandable. We also show that the students asses the feedback as actionable, with a majority reporting they can adapt their learning strategies and preparations for future exams.

7.2 Conclusion

The evaluation shows that the AFS is able to provide student-centred feedback in a scalable and timely manner to large courses. Therefore, we have contributed to solving the challenges of feedback in higher education in general and to solving the problems of exam-specific feedback in particular. Further we designed an AFS that is easy to adapt for non-experts. We have shown different design aspects required for such AFS, e.g. an adaptable domain model utilising Bloom's taxonomy and customisable task types to ensure applicability across educational domains. The AFS integrates expert knowledge and available student data to generate personalised, actionable feedback for different types of exams. Additionally, the system provides a scalable, automated implementation to ensure timely, consistent feedback without manual intervention.

Our research faces some limitations. Only 65 of the 1323 students filled in the survey. Therefore, the evaluation is seen as a first indicator that our approach is promising. More extensive evaluations are to follow. The usage of GenAI for the text block generation is a potential way to enhance the efficiency of the setup process of the AFS. Furthermore, we will analyse whether feedback via chatbots can enhance the student experience.

Ethics Declaration

This study used anonymised data only. No personally identifiable information was collected, stored, or analysed. Therefore, no ethical approval was required.

AI Declaration

AI-based tools were used solely for language improvement.

References

Alquraan, M. F. (2012) "Methods of assessing students' learning in higher education", *Education, Business and Society: Contemporary Middle Eastern Issues*, Vol. 5, No. 2, pp 124–133.

Anderson, L. W. and Krathwohl, D. R. (2001) *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*, New York, Munich, Longman.

Appelhaus, S., Werner, S., Grosse, P. and Kämmer, J. E. (2023) "Feedback, fairness, and validity: effects of disclosing and reusing multiple-choice questions in medical schools", *Medical education online*, Vol. 28, No. 1, pp 1–7.

Barker, T. (2010) "An Automated Feedback System Based on Adaptive Testing: Extending the Model", *International Journal of Emerging Technologies in Learning (iJET)*, Vol. 5, No. 2, p. 11.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. and Krathwohl, D. R. (1956) *Taxonomy of Educational Objectives*, New York, Longmans, Green & Co.

Carless, D., Joughin, G. and Liu, N.-F. (2006) *How assessment supports learning: Learning-oriented assessment in action*, Hong Kong, London, Hong Kong University Press; Eurospan distributor.

Daka, H., Mulenga-Hagane, M. L., Mukalula-Kalumbi Mwansa and Lisulo, S. (2021) "Making Summative Assessment Effective", *European Modern Studies Journal 5(4)*, pp 224–237.

Debuse, J. C. W., Lawley, M. and Shibl, R. (2008) "Educators' perceptions of automated feedback systems", *Educational Technology 24(4)*, pp 374–386.

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M. and Weerdt, J. de (2021) "A review of automated feedback systems for learners: Classification framework, challenges and opportunities", *Computers & Education*, Vol. 162, p. 104094.

Del Sánchez-Vera, M. M., Fernández-Breis, J. T., Castellanos-Nieves, D., Frutos-Morales, F. and Prendes-Espinosa, M. P. (2012) "Semantic Web technologies for generating feedback in online assessment environments", *Knowledge-Based Systems*, Vol. 33, pp 152–165.

Gibson, K. and Shaw, C. M. (2011) "Assessment of Active Learning", in *Oxford Research Encyclopedia of International Studies*.

Gronlund, N. E. (1998) *Assessment of studies achievement*, 6th edn, Boston, Allyn and Bacon.

Halinen, K., Ruohoniemi, M., Katajavuori, N. and Virtanen, V. (2014) "Life science teachers' discourse on assessment: a valuable insight into the variable conceptions of assessment in higher education", *Journal of Biological Education*, Vol. 48, No. 1, pp 16–22.

Hattie, J. and Timperley, H. (2007) "The Power of Feedback", *Review of Educational Research*, Vol. 77, No. 1, pp 81–112.

Haughney, K., Wakeman, S. and Hart, L. (2020) "Quality of Feedback in Higher Education: A Review of Literature", *Education Sciences*, Vol. 10, No. 3, p. 60.

Henderson, M., Ryan, T. and Phillips, M. (2019) "The challenges of feedback in higher education", *Assessment & Evaluation in Higher Education*, Vol. 44, No. 8, pp 1237–1252.

Hevner, A. R. (2007) "A Three Cycle View of Design Science Research", *Scandinavian journal of information systems 19(2)*, pp 87–92.

Hevner, A. R., March, S. T., Ram, S. and Park, J. (2004) "Design Science in Information Systems Research", *MIS quarterly*, pp 75–105.

Higgins, R., Hartley, P. and Skelton, A. (2002) "The Conscientious Consumer: Reconsidering the role of assessment feedback in student learning", *Studies in Higher Education*, Vol. 27, No. 1, pp 53–64.

Holmes, L. E. and Smith, L. J. (2003) "Student Evaluations of Faculty Grading Methods", *Journal of Education for Business*, Vol. 78, No. 6, pp 318–323.

Jacoby, J. C., Heugh, S., Bax, C. and Branford-White, C. (2014) "Enhancing Learning through Formative Assessment", *Innovations in Education and Teaching International*, Vol. 51, No. 1, pp 72–83.

James, R., McInnis, C. and Devlin, M. (2002) *Assessing learning in Australian universities: Ideas, strategies and resources for quality in student assessment*.

Karkalas, S., Mavrikis, M., Xenos, M. and Kynigos, C. (2016) "Feedback Authoring for Exploratory Activities. The Case of a Logo-Based 3D Microworld_AAM".

Keuning, H., Jeuring, J. and Heeren, B. (2019) "A Systematic Literature Review of Automated Feedback Generation for Programming Exercises", *ACM Transactions on Computing Education*, Vol. 19, No. 1, pp 1–43.

Li, J. and Luca, R. de (2014) "Review of assessment feedback", *Studies in Higher Education*, Vol. 39, No. 2, pp 378–393.

McGowan, A., Anderson, N. and Smith, C. (2024) "The use of ChatGPT to generate Summative Feedback in Programming Assessments that is Consistent, Prompt, without Bias and Scalable", *Proceedings of the Cognitive Models and Artificial Intelligence Conference,* 25.05.2024. New York, USA, ACM, pp 39–43.

Medina, M. S. and Yuet, W. C. (2013) "Promoting academic integrity among health care students", *American journal of health-system pharmacy*, Vol. 70, No. 9, pp 754–757.

Mekonen, K. Y. and Fitiavana, A. R. (2021) "Assessment of Learning Outcomes in Higher Education: Review of literature", *International Journal of Research Publications*, Vol. 71, No. 1.

Morales-Chan, M., Amado-Salvatierra, H. R., Medina, J. A., Barchino, R., Hernández-Rizzardini, R. and Teixeira, A. M. (2024) "Personalized Feedback in Massive Open Online Courses: Harnessing the Power of LangChain and OpenAI API", *Electronics*, Vol. 13, No. 10, p. 1960.

Nicol, D. (2010) "From monologue to dialogue: improving written feedback processes in mass higher education", *Assessment & Evaluation in Higher Education*, Vol. 35, No. 5, pp 501–517.

O'Donovan, B., Rust, C. and Price, M. (2016) "A scholarly approach to solving the feedback dilemma in practice", *Assessment & Evaluation in Higher Education*, Vol. 41, No. 6, pp 938–949.

Omar, N., Haris, S. S., Hassan, R., Arshad, H., Rahmat, M., Zainal, N. F. A. and Zulkifli, R. (2012) "Automated Analysis of Exam Questions According to Bloom's Taxonomy", *Procedia - Social and Behavioral Sciences*, Vol. 59, pp 297–303.

Park, J. (2023) "Medical students' patterns of using ChatGPT as a feedback tool and perceptions of ChatGPT in a Leadership and Communication course in Korea: a cross-sectional study", *Journal of educational evaluation for health professions*, Vol. 20, p. 29.

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V. and Bragge, J. (2020) *Design Science Research Process: A Model for Producing and Presenting Information Systems Research*.

Rawat, A., Kumar, S. and Singh Samant, S. (2023) "A Systematic Review of Question Classification Techniques Based on Bloom's Taxonomy", *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT).* Delhi, India, 06.07.2023 - 08.07.2023, IEEE, pp 1–7.

Rawlusyk, P. E. (2018) "Assessment in Higher Education and Student Learning", *Journal of Instructional Pedagogies 21*.

Ray, E. M., Daugherty, K. K., Lebovitz, L., Rudolph, M. J., Shuford, V. P. and DiVall, M. V. (2018) "Best Practices on Examination Construction, Administration, and Feedback", *American journal of pharmaceutical education*, pp 1127–1133.

Sambell, K., McDowell, L. and Montgomery, C. (2012) *Assessment for learning in higher education*, Abingdon, Oxon, Routledge.

Schaffer, H. E., Young, K. R., Ligon, E. W. and Chapman, D. D. (2017) "Automating Individualized Formative Feedback in Large Classes Based on a Directed Concept Graph", *Frontiers in psychology*, Vol. 8, p. 260.

Schipper, E. de, Feskens, R. and Keuning, J. (2021) "Personalized and Automated Feedback in Summative Assessment Using Recommender Systems", *Frontiers in Education*, Vol. 6.

Scoles, J., Huxham, M. and McArthur, J. (2013) "No longer exempt from good practice: using exemplars to close the feedback gap for exams", *Assessment & Evaluation in Higher Education*, Vol. 38, No. 6, pp 631–645.

Scott, T. (2003) "Bloom's taxonomy applied to testing in computer science classes".

Shum, S. B., Lim, L.-A., Boud, D., Bearman, M. and Dawson, P. (2023) "A comparative analysis of the skilled use of automated feedback tools through the lens of teacher feedback literacy", *International Journal of Educational Technology in Higher Education*, Vol. 20, No. 1, pp 1–42.

Starr, C. W., Manaris, B. and Stalvey, R. H. (2008) "Bloom's taxonomy revisited", *ACM SIGCSE Bulletin*, Vol. 40, No. 1, pp 261–265.

Tass Grigoriou, Christopher Cheong, and France Cheong (2015) "Improving Quality of Feedback Using a Technology-Supported Learning System", *PACIS 2015 Proceedings*.