EXPLORING ITEM ESTIMATORS' PARAMETER AND BLOOMS TAXONOMY FOR MARKING AND GRADING ASSIGNMENTS IN HIGHER EDUCATION INSTITUTIONS: A CASE OF BOTSWANA

Kaone Bakokonyane
Institute of Development Management, Department of Educational Management
kbakokonyane@idmbls.ac.bw

Nkobi Owen Pansiri
University of Botswana, Department of Educational Foundations
pansirion@ub.ac.bw

Abstract

The purpose of this study was to explore item estimators' parameter and Blooms Taxonomy for marking and grading assignments in higher education institutions in Botswana. This approach helped in constructing, validating assignment items, as well as in marking, scoring, grading and analyzing the marks. The study examined how these parameters assisted the lectures in marking and grading of twenty (20) assignment papers. This study used a qualitative approach to gather non-numerical data from the eight lecturers who were purposively sampled as there were expertise in the area. Four lectures marked ten assignment papers for module A and the other four lectures marked ten assignment papers for module B. Twenty (20) assignment papers were selected from a total of forty-four (44) from two modules for the fourth-year students. The case study research design was used to produce a comprehensive understanding of the marking and grading process for twenty scripts from each module. Lecturer interpretations of the marks were conducted after document analysis of the marked scripts and used rubrics. Thematical analysis were then employed to examine the information gathered from the eight lecturers who were interviewed regarding the item estimators' parameter and Blooms Taxonomy approach in assignment marking and grading. Four themes emerged as; simplicity, flexibility, accuracy, and speed as compared to alternative approaches in rubric 1-3. This approach was found to be effective as it made the marking and grading task less demanding, fair, well-organized and simple to use. In conclusion, the study present solid evidence suggesting that the use of an approach for marking and grading assignment items based on Bloom's Taxonomy and Item Response Framework is a useful and effective tool in the context of marking, scoring, grading, and analyzing the marks. Its implementation in institutions of higher learning is recommended to improve the quality and reliability of assignment items and to promote more effective marking and grading.

Key words; Marking, Grading, Item estimator, Ability estimator, Blooms Taxonomy, Rubric

ISSN 1021-559X /12/2024
Mosenodi

Introduction

Grades serve as the essential currency within the educational system (Rockhead, 2019). They hold significant predictive power regarding future educational outcomes (Guskey & Link, 2019), including admissions to higher education, performance therein, and eventual graduation (Rockhead, 2019). Research indicates that educational systems heavily depended on high stakes testing and examinations to assess academic achievement for various purposes since the 20th century (Quinn, 2020). Furthermore, it has been noted that in the 21st century, there was a shift in focus towards classroom assessment, particularly assignments, which emerged as a distinct area of study (Gonzalez et al., 2015).

High-stakes tests, examinations, and assessments in higher education are evaluated and graded by lecturers to assess students' comprehension of the learning material. Over time, the processes of marking and grading have been scrutinised, revealing various challenges. These include the prevalent reliance on grades and the public's perception of them as measures of student success (Brookhart, 2015; Cai et al., 2016), their role in providing feedback to stakeholders (Watling & Ginsburg, 2019), their impact on student motivation and learning, as well as their implications for students' future academic achievements and their willingness to engage in educational programs (Boud et al., 2015). Ehlers and Schwager (2016) contend that "For many students, the most important aspect in their graduation process is not learning, understanding, and knowledge, but the received grade. They find themselves struggling for good grades since these are a prerequisite for later getting the best chances on the labour market" (pp. 506).

Haines (2021) explored the concept of grades as complex indicators of student achievement, effort, and engagement. Previous investigations into the inconsistencies of classroom grading (Rockhead, 2019; Samejima, 2016) and the questionable validity of grades (Boud et al., 2015) have also been undertaken. These earlier studies have highlighted the potential for grading student work to be both invalid and unreliable, which served as a motivation for the current research on grading and marking approach. This current study integrated elements from Item Response Theory (IRT) and Bloom's Taxonomy to create a more streamlined approach to the assessment and grading of assignments in higher education. Specifically, it employed item estimator parameters for the construction and validation of assignments, alongside ability estimator parameters for the evaluation, scoring, grading, and analysis of students' marks (Abedi et.al, 2021; Drake et al., 2023; Samejima, 2016). Furthermore, the study used Bloom's Taxonomy across all parameters to measure the cognitive complexity of content-level and ensure that students engage in comprehensive logical reasoning processes (Wilson, 2016).

Higher education lectures in Botswana have shown significant variation in the criteria they use in marking and grading students of assignments. The research conducted by Adedoyin and

Chisiyanwa (2018) indicates that the lecturers use a limited range of assessment strategies. Students expressed concerns that the assessment methods predominantly rely on traditional measures, which do not adequately reflect their capabilities. Furthermore, the study revealed that the feedback provided to students is insufficient; they reported that it is often delayed, overly brief, and unlikely to facilitate their development. Additionally, the research found that the test items lack challenge and fail to assess various dimensions of student understanding. Concerns were also raised regarding the fairness and validity of the assessments conducted by lecturers. Many students noted a lack of clarity and transparency regarding expectations, leading to perceptions of unfair and invalid assessment. The assessment items primarily concentrate on lower-order cognitive questions, and students concurred that the assignments do not require them to engage in analysis, synthesis, or evaluation. Other issues identified include the non-strategic timing of assessments and inconsistencies in the grading practices of lecturers. This study, based on insights from Adedoyin and Chisiyanwa (2018), investigated the parameters of item estimators and Bloom's Taxonomy in the context of marking and grading assignments within higher education institutions in Botswana. It examines how these item estimators and Bloom's Taxonomy can be used to construct and validate assignment items, as well as how they can facilitate the processes of marking, scoring, grading, and analyzing assignments in the higher education sector.

Theoretical background of the study

The issue of marking and grading student assignments in higher education has been recognised since the 1800s (Rockhead, 2019). In the early 1900s, Starch and Elliott (1912) conducted research that revealed the potential for subjectivity in the assessment of student writing, including essays, reports, and constructed response test items. Their investigation focused on the consistency of teachers' grading practices shortly after the introduction of percentage-based grading systems. In their study, 142 educators were tasked with grading the same piece of English work, resulting in grades that varied from 50% to 98%. Additionally, in 1930, Eells (1930) assessed the reliability of individual teachers' grading by having 61 teachers evaluate the same history and geography assignments on two separate occasions, with an interval of 11 weeks between the assessments. Following an analysis of reliability coefficients, he determined that the assignment scores were "slightly better than mere guesses" (p.52) and that "the variability of grading is comparable within the same individual as it is among different individuals" (p.52). He investigated two extensive trials conducted from 1912 to 1930. The researchers concluded that "the advancement or hindrance of a student largely relies on the subjective assessment of their teacher" (p.52) rather than the actual work produced by the student, due to the discrepancies in scores assigned by different teachers, which varied from failing to outstanding (Eells, 1930; Starch & Elliot, 1912). Drawing from the insights of these studies, an approach of marking reliability was conducted using two modules, four distinct rubric marking modes, and four lectures at a higher education institution, revealing a similar issue. Consequently, the study proposed an approach for establishing parameters in the construction, marking, scoring, analysis, and validation of assignments within higher education.

Item estimator parameter and Blooms taxonomy that help in constructing and validating of assignments in institutes of higher education

Item Response Theory encompasses a collection of mathematical models designed to analyze how individuals respond to assessment items, which is reflected in its nomenclature (Hori et al., 2022). Bortolotti et al. (2013) state that "Item response theory is utilised for the development, evaluation, and administration of standardised measurements; it finds extensive application in the fields of psychology and education" (p. 2341). This theory is predominantly applied in multiple-choice and short-form assessments (Hori et al., 2022). Nevertheless, despite its benefits in facilitating objective grading, scholars have expressed concerns regarding the indiscriminate application of multiple-choice and short assessments. Issues associated with these types of evaluations include their potential to inaccurately represent student comprehension, hinder critical thinking skills, and exhibit bias against specific student demographics (Hambleton, 2015).

It is essential to recognize that the item response theory framework comprises three distinct models. These models use the item estimator parameters in a flexible manner (Osterlind & Wang, 2017). As noted by Hambleton (2015), one of these models is the discrimination parameter, commonly referred to as the Single Logistic Model. This model measures the extent to which a question can differentiate between lower-performing and higher-performing examinees (Haines, 2021). Furthermore, there is the difficulty parameter, often identified as the dual or two-parameter logistic model (Hori et al., 2022). This model incorporates both discrimination and difficulty parameters to evaluate the probability of a correct response. Various items may facilitate different forms of discrimination (Guo, 2021).

The Triple Logistic Model, commonly referred to as the Pseudo-guessing parameter, is the third model in this context. It uses three key parameters—discrimination, difficulty, and guessing—to measure the probability of a response (Hansen et al., 2019). Notably, its inclusion of a guessing component results in a lower asymptote, suggesting that even students with very low ability may answer questions correctly by chance (Guo, 2021). In certain cases, depending on the evaluation type and data fit, it may be appropriate to utilize only one (b) or two (a, b) parameters (Bortolotti et al., 2013). Furthermore, when dealing with multipoint questions, such as those using Likert scales or partial credit items, the models can be adapted to include additional characteristics (Hori et al., 2022). There are instances where guessing is not a factor.

This research used a two-parameter model concentrating on discrimination and difficulty, as noted by Ayanwale et al. (2018). Additionally, the study incorporated Bloom's Taxonomy, which played a crucial role in the development and validation of assignments. According to Piza-Mir (2022), Bloom's Taxonomy is vital for the categorization and classification of various skill and knowledge levels necessary for effective teaching and learning (Bloom, 1956). The author further asserts that Bloom's Taxonomy is a powerful tool for analyzing and developing assessment tests.

Piza-Mir (2022) also illustrated a strong correlation between the taxonomy and assessments, suggesting that it is an invaluable resource for guiding the design of tests that accurately measure the content and skills taught in the classroom (Bloom, 1956). This importance arises from Bloom's Taxonomy being essential in the creation and validation of questions, as it includes a set of action verbs that facilitate the construction of assessment tests. The chosen verb acts as the central element of the test questions formulated by educators, ensuring a thorough representation of all components of Bloom's Taxonomy in a balanced manner (Bloom, 1956).

Ability estimator parameter and Blooms Taxonomy that help in marking, scoring, grading, and analyzing of assignments in institutions of higher learning.

The study introduced an approach for assessment and grading that combines the estimator parameter with Bloom's Taxonomy. In this framework, the ability estimator parameter enables the examiner to select items by employing an alternative Item Response Theory (IRT) model known as the Dual Logistic Model or the difficulty parameter (Hori et al., 2022). This model allows examiners to categorise items based on their levels of difficulty. Although the initial theory did not incorporate Bloom's Taxonomy, which is crucial for evaluating difficulty, the research recognised it as a solution to this theoretical shortcoming (Wilson, 2016). Bibi et al. (2020) argue that Bloom's Taxonomy is instrumental in formulating various types of questions that support diverse cognitive processes. This framework enables teachers to ask questions ranging from basic factual recall to more complex cognitive tasks, thereby fostering a variety of thinking processes among students. For each skill, Bloom utilizes active verbs that demonstrate how students apply their learned knowledge (Seaman, 2011).

The original Bloom's Taxonomy encompasses a range of skills that progress from the simplest to the complex. The first level is Knowledge, which focuses on the ability to recall or recognise information that has been previously acquired. Instructional verbs associated with this foundational cognitive domain include writing, listing, labelling, naming, and stating (Wilson, 2016). The subsequent level is Comprehension, which involves the ability to understand or interpret information obtained from previously acquired knowledge. Instructional verbs associated with this level encompass explaining, summarising, describing, and illustrating (Seaman, 2011).

The second component is an application, characterized by instructional verbs such as use, solve, illustrate, and apply. This aspect focuses on selecting and utilizing data concepts to independently address a problem (Wilson, 2016). The subsequent taxonomy is analysis, with instructive verbs including compare, contrast, and analyze. Analysis involves understanding or breaking down the assumptions inherent in a statement or question to reach conclusions (Forehand, 2010). The instructional verbs associated with the next level are create, design, invent, and develop. Synthesis refers to the process of combining ideas to formulate a new concept or plan (Seaman, 2011). Lastly, evaluation involves assessing students against established criteria, with instructional verbs such as judge, critique, and justify being employed in this context.

In summary, Bloom's revised taxonomy serves as an effective framework for developing assessment tests and structuring teaching-learning sequences. The application of the defined levels, aligned with the verb categories of Bloom's taxonomy, is substantiated as a valuable tool for creating assessment tests. This approach reflects various psychological processes associated with learning across different subjects and demonstrates a shared understanding among educators, whether experienced or novice, and their students.

The study's item estimators' parameter and Blooms Taxonomy marking and grading assignments approach in Higher Education Institutions

The researchers utilized the insights gained from the literature review and theoretical framework to devise an approach for marking and grading assignments in higher education. Subsequently, they tested this approach against alternative approaches to determine its applicability for lecturers in higher education institutions. The developed marking and grading approaches were designed to ensure that the assessment topics were aligned with the Single Logistic Model or discrimination parameter. This model includes an item estimator parameter that aids in the selection of topics and criteria for evaluation. Additionally, it comprises an ability estimator parameter that facilitates the marking, scoring, and grading of assignments. These parameters are organized into two phases, as elaborated in the following section.

This parameter for estimation assisted the researchers in formulating the topic during Stage 1 by utilising the learning outcomes associated with the modules used. This assertion is corroborated by Abedi, et.al (2021), who contends that the topics assigned in coursework should be in harmony with the learning outcomes. Haladyna (2018) defines learning outcomes as statements that delineate the knowledge or skills that students are expected to acquire by the conclusion of a specific assignment, class, course, or program. Furthermore, Haladyna (2016) posits that effective learning outcomes prioritise the application and integration of knowledge, rather than merely covering the material. They specify how students will be able to utilize the material, both within the classroom setting and in broader contexts (Ayanwale et al., 2018).

In Stage 2, the researchers developed the content that was addressed in the study. They created a rubric featuring binary response options: correct and incorrect, which were influenced by the Single Logistic Model-Discrimination parameter derived from Item Response Theory (IRT) (Hambleton, 2015; Hori et al., 2022). Osterlind and Wang (2017) emphasise the importance of designing and utilizing rubrics for grading assignments or tests, as this practice can minimise inconsistencies and enhance the objectivity of evaluating written work. Consequently, the model incorporated a rubric shaped by the insights of the aforementioned authors (Hambleton, 2015; Nieminen, 2022; Osterlind et al., 2017).

Furthermore, in Stage 3, the study focused on item selection. The researchers categorised the items in the rubric according to Bloom's Taxonomy levels of cognitive domains, employing

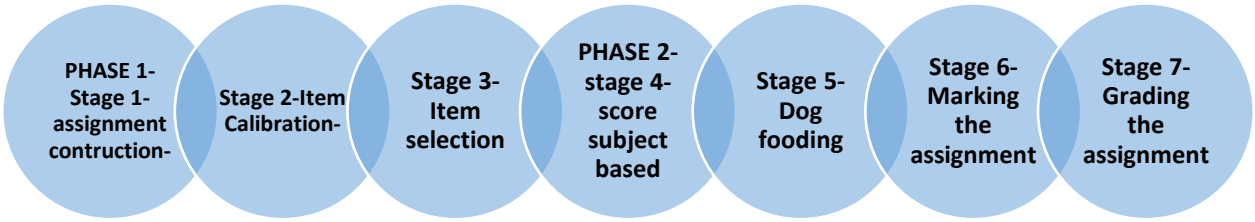
various criteria. This approach was supported by Nieminen (2022), who contends that the difficult parameter aids examiners in formulating items that effectively distinguish between lower and higher-performing examinees. In Phase 2, the ability estimator parameter was used in Stage 4, where researchers assigned scores to items according to their difficulty level, utilizing the Bloom's Taxonomy cognitive domain. In this framework, lower-level objective criteria received lower scores, while higher-level criteria were awarded higher scores (Wilson, 2016). The researchers established a scoring distribution along the Bloom taxonomy: Remembering was assigned 1 mark for each associated verb, Understanding received 2 marks per related verb, applying was allocated 3 marks for each relevant verb, Analyzing also received 3 marks per related verb, Evaluating was given 5 marks for each corresponding verb, and creating was awarded 6 marks for each related verb (Forehand, 2011; Seaman, 2011; Wilson, 2016).

In Stage 5, the concept of dog fooding was implemented. As noted by Damian et al. (2023), dog fooding refers to the practice of a company's employees using a newly developed product or service to evaluate it prior to its release to customers. Through this process, the researchers created and assessed two assignments, which enabled them to extract sufficient vocabulary from the synonyms feature in word processing software, utilized by the eight examiners during the evaluation of the students' assignments.

In Stage 6, the examiners were provided with four rubrics to evaluate ten assignments from third-year students. The assignments were selected through random sampling. The study employed the maximum likelihood approach derived from Item Response Theory (IRT), whereby a student receives a score upon meeting the established criteria and no score if the criteria are not met (Ayanwale et al., 2017; Bortolotti et al., 2013; Hambleton, 2015; Osterlind & Wang, 2017). During this process, the examiners using word processing features such as "speak," "find," and "navigation" to efficiently identify key points. The "speak" feature allowed examiners to highlight text for the computer to read aloud, enabling them to listen while reading simultaneously. The "find" and "navigation" features facilitated the rapid location of relevant points within the examinee's submissions by using specific keywords for marking.

In Stage 7, the examiners assessed the assignments using the Maximum Posteriori method from IRT theory to determine which criteria received the highest scores (mode) and the Expectation Posteriori method (mean) to calculate the average score (Ayanwale et al., 2017; Bortolotti et al., 2013; Hambleton, 2015; Osterlind & Wang, 2017).

Figure 1: Marking and grading approach



Phase 1-Item Estimator Parameter

Stage 1-Assignment construction (You come up with a topic).

Stage 2-Item Calibration (Content that you will cove, come up with a rubric which has binary items correct response and incorrect response)-Single Logistic Model-Discrimination parameter.

Stage 3- Item selection of a correct response-subject ability (Discriminate the items in the rubric using blooms taxonomy Difficult parameter (the depth of the content that you want to cover and align them with the Bloom taxonomy cognitive domain).

Phase 2-Ability Estimator Parameter

Stage 4-Score subject based- (Allocate scores to items based on their level of difficulty using the Bloom taxonomy cognitive domain low level objective criterions get low marks and high-level criterions get high marks).

Stage 5- Dog fooding (You write the assignment and mark it using word mining from word processing using synonyms ribbon).

Stage 6- Marking the assignment (You are using maximum likelihood- Get a mark when you have satisfied the criterion and no mark when you have not satisfied the criterion) In this one the examiner will use word processing ribbons such as speak, find and navigation to get the points quickly).

Stage 7-Grading the assignment (You use Maximum posteriori check which criterion has maximum marks-mode) (Expectation Posteriori-Mean what is the average mark).

Methodology

This research used a qualitative approach. Non-numerical data were collected and analyzed from eight lecturers who marked a total of twenty assignment papers, with ten papers sourced from each module. Four lecturers assessed ten assignment papers for Module A, while the remaining four lecturers marked ten assignment papers for Module B. In total, twenty assignment papers were selected from a pool of forty-four across the two modules for fourth-year students. A case study research design was implemented to gain an in-depth understanding of the marking and grading process for the ten scripts from each module. Purposive sampling was applied to select the twenty

assignment papers from the overall forty-four available from the two modules, noting that there were twenty-two assignment papers in each module (Hollstein, 2011).

In this study, eight lectures were purposefully sampled based on their expertise in the relevant field. Each participant had over five years of experience teaching the module, indicating their familiarity with the subject matter. These lecturers were included in the research to gather direct insights into their personal experiences regarding the marking and grading of assignments. Additionally, purposive sampling was employed to select fourth-year students, as assessments in the earlier years—first, second, or third—tend to carry lower stakes. In contrast, assessments conducted later in the program, especially in the final year, are likely to be of higher stakes, significantly affecting the students' outcomes.

The research employed a case study design to gain an in-depth understanding of the marking and grading procedures for ten scripts from each module of fourth-year students (Hay & Hastie, 2012). Following a thorough document analysis of the marked scripts and the associated rubrics, lecturers helped in uncovering and understanding of variations of marks within and among lecturers. This approach facilitated the identification and comprehension of discrepancies both within and between lecturers (Fetterman, 1988).

Data analysis and presentation

Twenty assignments were distributed among eight lecturers, each of whom specialized in two distinct modules, and all assignments were evaluated using a variety of marking rubrics. The ten assignments submitted by twenty-two students enrolled in a single module during their fourth year were assessed by four lecturers through several methods: initially, without any rubric; subsequently, with a non-detailed rubric that prioritized the students' writing style over the content; thirdly, with a rubric that evaluated comprehension of the material while neglecting the cognitive levels outlined in Bloom's taxonomy; and finally, with a comprehensive marking approach that incorporated a rubric considering Bloom's taxonomy cognitive levels along with certain computer applications. Additionally, another set of ten assignments from the same group of twenty-two students was graded by four lecturers in a manner analogous to the first set.

The thematic analysis developed by Braun and Clarke (2022) was used in this study to analyze the information obtained from eight interviewed lecturers concerning their experiences with assignment marking and grading. The interviews were conducted in a structured manner and subsequently transcribed. The researchers meticulously examined the data from the eight lecturers to identify recurring themes or concepts that were referenced multiple times.

Table 1
Results of marked assignment in Module A

Lecturer	Without rubric	Non-detailed rubric	Detailed with rubric without Blooms	Detailed with Blooms	Mean	Standard deviation
Lecturer 1	58.4	77.5	65.9	70.2	68	6.9
Lecturer 2	74.8	66.1	73.3	68.9	70.8	3.5
Lecturer 3	69.3	56.3	64.4	69.8	64.9	5.4
Lecturer 4	77.2	70.5	62.6	68.7	69.8	5.2
Mean	69.9	67.6	66.6	69.4		
Standard deviation	7.2	7.6	4.0	0.6		

Note: Analysis from the data using Braun and Clarke (2022)

Lecturer 1 in his marking had mean (averages marks for each rubric mode) of 68 and standard deviation of 6.9. This means that his marking variation were more spread out from the mean. When the lecturer was asked to what might have caused the huge variation between his markings, he said, "***In the first marking where I used my knowledge to mark, though it was easy to mark, it was not easy to score as there was nothing to guide me".*** Lecturer 2 had a mean of 78 and standard deviation was 3.5. This lecturer as compared to Lecture 1 did not have huge variation in his marking from the mean. The lecturer stated that, "***I did not have a huge variation because naturally I am a strict marker though it takes all of my time."***

The third lecturer has a mean of 64.9 and standard deviation of 5.4. The variation this lecturer is huge from her marking. The lecturer said, ***"I have huge variation because the first and second rubrics markings were more or less like guessing."*** The fourth lecturer had a mean of 69.8 and standard deviation of 5.2. This lecturer also had a huge variation in marking using different rubrics. The lecture stated that, "***I think two marking models for Rubric 1 and Rubric 2 are the ones that have contributed to this huge variation"***

The mean for all lecturers who marked without rubric is 69.9 and standard deviation is 7.2. The variation for marking without rubric has proved to be huge among all the four lecturers. This Lecture 4 stated that, "***There was a huge variation when marking without a rubric because students were not judged in a similar way***." Lecture 1 said that "***There was no consistency in marking students work."***

The mean for marking without a detailed rubric had a mean of 67.6 and standard deviation of 7.6. The variation among the lecturers who marked using a non-detailed rubric was also huge. Lecture 2 agreed by saying, "***There was a huge variation because judgements were not fair, reliable and were not as objective as possible".*** Lecture 4 also said that there was a variation since,

"***For rubric 2 students were penalized or rewarded for writing style as opposed to knowledge they have on the content tested."***

Moreover, the mean for a rubric which is detailed without Blooms Taxonomy cognitive level is 66.6 and standard deviation of 4.0. The standard deviation of this one is also huge but better that the one for the other two. Lecture 3 stated that, ***"The variation as compared to the ones for Rubric 1 and rubric 2 was getting narrower at least there was a guide to if the criteria was addressed was, I could give a mark and no mark where it was not addressed."*** Lecture 1 stated that, "***This one made it a bit easier for me to mark because it helped me distribute marks easier as I give a mark for a met criterion and no mark for not met criterion***."

Furthermore, the mean for rubric with Blooms Taxonomy cognitive level is 69.4 and the standard deviation is 0.6. This is a slight variation for lecturers who marked using a rubric with Blooms taxonomy cognitive level. This one was supported by Lecture 2 by saying, ***"There was a slight to no variation because the rubric helped me to capture more information about the students' performance."*** Lecture 4 also supported this to say, ***"This rubric helped me to marks as it helped me to record the extent to which the students achieved something not just whether the criteria was achieved and not achieved."***

Table 2

Results of marked assignment in Module B

Lecturer	**Without rubric**	**Non-detailed rubric**	**Detailed with rubric without Blooms**	**Detailed with Blooms**	**Mean**	**Standard deviation**
Lecturer 1	74.3	67.7	78.3	75.2	73.9	3.8
Lecturer 2	56.9	64.2	72.8	76.7	67.7	7.7
Lecturer 3	52.6	68.5	74.2	79.3	68.7	10.0
Lecturer 4	72.1	50.8	66.9	74.6	66.1	9.3
Mean	63.9	62.8	73.1	76.4		
Standard deviation	9.4	7.1	4.0	1.8		

Note: Analysis from the data using Braun and Clarke (2022)

Lecturer 1 in his marking had mean (averages marks for each rubric mode) of 73.9 and standard deviation of 3.8. This means that his marking variation was not that much more spread out from the mean. Lecture 1 when asked to why her variation was not spread much said, ***"I might have had a slight deviation because, although rubric 1 and 2 did not give more guidance and direction, I used my expertise as an experienced lecture."***

Lecturer 2 had a mean of 67.7 and standard deviation was 7.7. This lecturer has huge variation in his marking from the mean. The lecturer said, "***I am having a huge deviation because***

in Rubric 1 and Rubric 2 I was only using the best fit for the students and not necessarily what the students might have addressed the content." The third lecturer has a mean of 68.7 and standard deviation of 10.0. The variation this lecturer is huge from her marking. ***"I had this great deviation because most of our students cannot express their thoughts clearly hence, I struggled more especially where there was no rubric."*** The fourth lecturer had a mean of 66.1 and standard deviation of 9.3, this lecturer also had a huge variation in marking using different rubrics. The lecturer posits that, "***It has shown that there is a huge variation because in some instances more especially when I was marking using rubric 1 and 2 it has required much of my independent judgement coz it was not objective."***

The mean for all lecturers who marked without rubric is 63.9 and standard deviation is 9.4. The variation for marking without rubric has proved to be huge among all the four lecturers. Lecture 3, stated that, "***They might be huge variation because in most cases more especially in rubric 1, 2 and 3 lecturers were using their personal judgement since there were no subjective guidance and assessment criteria."***

The mean for marking without a detailed rubric had a mean of 62.8 and standard deviation of 7.1. The variation among the lecturers who marked using a non-detailed rubric was also huge. Lecture 2 argues that ***"There could be a huge variation because lectures marked without a rubric and this has given lecturers less information on what to mark correct and wrong."*** This was supported by Lecture 1 said, "***There was no reliable marking due because marking rubrics were uncluttered and pursuing clarity of criteria in order to increase the reliability of assessment, criteria can be altered in a way that threatens validity."***

Moreover, the mean for a rubric which is detailed without Blooms taxonomy cognitive level is 73.1 and standard deviation of 4.0. The standard deviation of this one is also huge but better that the one for the other two. This was supported by Lecture 4 by saying, "***The marking rubric given was able to allocate marks to objectively identifiable words or ideas that are listed in the marking rubric, but did not consider the levels of the objective."*** Furthermore, the mean for rubric with Blooms Taxonomy cognitive level is 76.4 and the standard deviation is 1.8. This is a slight variation for lecturers who marked using a rubric with Blooms Taxonomy cognitive level. Lecturer 2 stated that, "There ***was a slight variation or no variation because the marking rubric was explicit which then helped us a lot because it was able to help us distinguish better from poorer responses."*** This was supported by Lecture 3 by saying, ***"Rubric 4 was more reliable than other rubrics, due to the 'transparency' of criteria based on competences."***

Discussion

The marking and grading conducted through rubric 4, alongside the theoretical framework, literature review, and data analysis, facilitated the identification of a grading approach. From the data, four key themes emerged as; simplicity, flexibility, accuracy, and speed, particularly when

compared to alternative approaches. The marking and grading approach was determined to be user-friendly, as it divides the marking process into manageable stages, thereby simplifying the task for the marker. This approach's simplicity is attributed to its two phases, which are based on the Item Response theoretical framework. In the first phase, the theory employs item estimator parameters, which align with the study's initial objective. This objective primarily focuses on generating topics and establishing marking criteria (Bortolotti et al., 2013).

The second phase addresses the Item Response Theory framework and aligns with the study's second research objective. During this phase, the research concentrated on marking and scoring through the lens of Bloom's Taxonomy (Forehand, 2010), as well as grading approaches that include maximum likelihood, which identifies the criteria that receive scores, maximum posteriori, which establishes the highest possible marks (mode), and expectation posteriori, which calculates the average score (mean) (Cai et al., 2016). These elements are all considered in the grading process.

Anyanwale et al. (2018) emphases that an effective marking and grading system should be practical, avoiding excessive complexity or difficulty in implementation. Although all marking and grading methods rely on gathered data, a robust approach utilizes meaningful data collection to safeguard essential information regarding student performance (Brookhart, 2015). This was also reflected in the data, as the eight lecturers successfully used four distinct marking rubrics. It was discovered that the marking and grading approach was flexible. Given its versatility and ability to function effectively in both written assignments, it was discovered that it could be applied to any written assignment (Boud et.al, 2015). Its flexibility was most dependable when it came to the rubric, which outlined the (academic) expectations for students based on criteria and ensured consistency in assessments across assignments, courses, and individual students (Guo, 2021). Its adaptability contributed to its flawless validity and reliability, which resulted in assignments whose completion and marking leave no room for error and ultimately measure content in an unbiased manner (Guskey & Link, 2019).

The implementation of a marking and grading approach has been shown to enhance efficiency, especially when utilizing tools such as word processing, navigation, find, and speech ribbons. This approach allows educators to assess student assignments in a timely and efficient manner. Guskey and Link (2019) support this notion by asserting that ensuring the significance of grades is a crucial initial step in establishing effective grading practices. Achieving this requires educators to reflect on and acknowledge the various ways their values may have shaped their professional activities (Guo, 2021). Furthermore, a more detailed outline of expectations regarding content, structure, and grammar could be beneficial, as it aligns well with rubric development (Ayanwale, 2018).

Conclusion

The research revealed that when effectively applied, the marking and grading assignment approach can prove to be beneficial, as it alleviates the burdens associated with marking and grading tasks, ensuring fairness, organization, and ease of use. The study emphasizes that when selecting this marking and grading approach, whether in policy formulation or practical application, it is crucial to consider the item estimator parameter and ability estimator parameter. These factors enhance the simplicity, flexibility, accuracy, and efficiency of marking and grading assignments in higher education. Therefore, the study recommends of this marking and grading approach in higher education, as it provides independent and psychometrically robust measures for validating lecturers' marking and grading practices, thereby offering compelling evidence to support their grading decisions to both students and stakeholders.

Nonetheless, it is essential to acknowledge the methodological limitations of this research. The study used a qualitative methodology and focused on a single case, which limits the ability to generalize the significance of the findings. Furthermore, the context and methodology restrict the findings' generalizability. There is clearly an opportunity for further investigation, as this article presents only a preliminary analysis of the single case. Consequently, the study recommends that additional research could be conducted to explore the applicability of this approach for practical assignments.

Implication and recommendations

This study provides compelling evidence that the application of a marking and grading approach grounded in Bloom's Taxonomy and the Item Response Framework serves as a valuable and efficient tool for assessment purposes. It is recommended that higher education institutions can adopt this approach to enhance the quality and dependability of assignment items, thereby fostering more effective teaching practices. From a practical implication standpoint, this approach can enable lecturers to create more impactful marking and grading systems, as it facilitates the design of assessments that appropriately target the various levels of Bloom's taxonomy, ensuring that students acquire the essential skills and knowledge required.

References

Abedi, R., Nili Ahmadabadi, M. R., Taghiyareh, F., Aliabadi, K., & Pourroustaei Ardakani, S. (2021). The effects of personalized learning on achieving meaningful learning outcomes. *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, *12*(3), 177-187.

Adedoyin, O., & Chisiyanwa, L. (2018). Issues in Assessment Practices at Botswana Private Tertiary Institutions as Perceived by Undergraduate Students. *Asian Journal of Education and e-Learning (ISSN: 2321–2454)*, *6*(01).

Ayanwale, M. A, Adeleke, J. O, & Mamadelo, T. I. (2018). An assessment of item statistics estimates of basic education certtificate examination through classical test theory and item

response theory approach. *Inernational Journal of Educational Research Review*, *3*(4), 55-67.

Bibi, W., Butt, M. N., & Reba, A. (2020). Relating teachers' questioning techniques with students' learning within the context of Bloom's taxonomy. *FWU Journal of Social Sciences*, *14*(1), 111-119.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives.*

Bortolotti, S. L. V., Tezza, R., de Andrade, D.F., Bornia, AC., & de Sousa Junior, A.F. (2013). elevance and advantages of using he item response theory. *Qualiy and Quantity*, *47*(1), 2341-2360.

Boud, D., Lawson, R. & Thompson, D.G. (2015). The calibration of student judgement through self assessment patterns. *Higher Education Research and Development*, *34*(1) 45-59.

Braun, V. & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative psychology*, *9*(1), 3.

Brookhart, S. M. (2015). Gradedachievement, tested achievemen and validity. *Educational assessment*, *20(*4), 268-296.

Cai, L., Choi, K., Hansen, M., & Harrell, L. (2016). Item response theory. *Annual Review of Statistics and its Application*, 297-321.

Damian, A.L., Teixeira, L., Carrenho, B. C., Ferreira, B. B., Bentes, B. A., Tordin, M. G. & Pignatelli, M. R. (2023). Exploring UX Factors through dogfooding Approach: An Experince Report. *In proceedings of the XXII Brazillian Symposium on Software Quailiy*, 236-243.

Drake, A. P., Masur, P. K., Bazarova, N. N., Zou, E. W., & Whitlock, J. (2023). The Youth Social Media Literacy Inventory: Development and Validation Using Item Response Theory. *Journal of Children and Media*, *17*(4), 467-487.

Eells, N. C. (1930). Realiability of repeated grading of essay type examinations. *Journal of Educational Psychology*, *21*(1), 48-52

Ehlers, T. & Schwager, R. (2016). Honest grading, grade inflation and reputation. *Economic Studies*, 506-521.

Fetterman, D. (1988). Qualitative approaches to evalauting education. *Educational Researcher*, *17*(8) 17-23.

Forehand, M. (2010). Bloom's taxonomy. *Emerging perspectives on learning, teaching and technology*, 41(4), 47-56.

Gonzalez, M. D. L. O., Jareño, F., & López, R. (2015). Impact of students' behavior on continuous assessment in Higher Education. *Innovation: The European Journal of Social Science Research*, *28*(4), 498-507.

Guo, W. (2021). *Exploring Rating Quality in the Context of High stakes Rater-Mediated.* Alabama: The University of Alabama.

Guskey, T. K. & Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades. *Assesssmen in Education: Principles, Policy & Pracice*, 26 (3), 303-320.

Haines, C. (2021). *Assessing students' written work: marking essays and Reports.* England: Routledge.

Hambleton, R.K. (2015). *Handbook of item response theory*. CRC: CRC press.

Hansen, J., Sadler, P., & Sonnert, G. (2019). Estimating high school GPA weighting parameters with a graded response model. *Educational Meaurement Issues and Practice*, 38(1) 16-24.

Hastie, P., & Hay, P. (2012). *Qualitative approaches.* London: Routledge.

Hollstein, B. (2011). Qualitative approaches. *The SAGE handbook of social network analysis*, 404-416.

Hori, K., Fukuhara, H., & Yamada, T. (2022). Item response and its applications in educational measurement Part 1: Item response theory and its implementation in R. *Wiley Interdisciplinary Reviews: Computational Statistics*, *14*(2) 1531.

Nieminen, J. H. (2022). Disrupting the power relations of grading in higher education through summative self-assessment. *Teaching in Higher Education*, *27*(7), 892-907.

Osterlind, S. J., & Wang, Z. (2017). Item response theory in measuremen, assessment and evaluation of higher education. *In Handbook on mesurement, assessment and evaluation in higher education*, 191-200

Piza-Mir, B. (2022). Curriculum Analysis Accoring to Bloom's Revised Taxonomy in Science and Mathematics. *Journal of Positive School Psychology*, *6*(8).10125-10130.

Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, 42(3), 375-392.

Rockhead, I. G. (2019). *A Grounded Theory Exploration of Teacher Grading Practices*. Gardner-Webb University.

Samejima, F. (2016). Graded response models. In *Handbook of item response theory* (pp. 95-107). Chapman and Hall/CRC.

Seaman, M. (2011). Bloom's Taxonomy. *Curriculumn & Teaching Dialogue*, 13.

Starch, D., & Elliot, E. C. (1912). Reliability of grading of high-school work in English. *The School Review*, 20(7), 442–457.

Watling, C. J., & Ginsburg, S. (2019). Assessment, feedback and the alchemy of learning. *Medical education*, *53*(1), 76-85.

Wilson, L. O. (2016). Andrson and Krathwohl-Bloom'staxonomy revised. *Understanding the new version of Bloom's taxonomy*.