UNIVERSITEPARK bülten bulletin

CORRESPONDENCE
Kyffin Bradshaw
✉ *Kyffin.bradshaw@bcc.edu.bb*
✉ *Barbados Community College, Howell's Cross Road, St. Michael, Barbados, BB11058.*

AUTHOR DETAILS
Additional information about the author is available at the end of the article.

RESEARCH ARTICLE

Avoiding Fuzzy Letter Grading Systems: Designed Research to Ensure Fair Translation of Academic Performance into Letter Grades That Provide Tangible Measures of the Skills International Employers Seek from College Graduates

Kyffin Bradshaw [ID]

ABSTRACT

Background/purpose – College graduates from Small Island Developing States (SIDS) often seek regional and international employment. However, grading systems differ across countries and regions, making it difficult for international and regional employers to accurately interpret and translate foreign Letter Grades into the skills and personal attributes graduates need to secure regional and international employment. This study investigated the ability of SIDS to adequately represent on the global labor market the employability of their graduates after having completed a program specifically guided by learning outcomes, assessments, and Letter Grading.

Materials/methods – Bayesian statistics is considered robust in providing empirical evidence. Consequently, it was employed in this study to look for evidence linking the evaluation descriptors SIDS use in their Letter Grading System to the skills and levels their graduates actually attained in order to make them employable on equal grounds.

Results – The findings suggest that the Letter Grading System used by SIDS is inadequate for reliably communicating the level of academic mastery and competencies needed for graduating students to achieve employability on the international market.

Conclusion – Ultimately, SIDS must design and implement ambiguity-free, homogeneous Letter Grading Systems that include comprehensive qualitative indicators of the evaluation criteria used to rank their graduates if they are to convey valid and consistent messages to those who need to interpret them (i.e., prospective employers).

Keywords – Employability, Assessment, Bayesian statistics, Letter grade system.

1. INTRODUCTION

The academic grading process, as highlighted by Opstad (2020), is widely used in higher education for ranking graduates and providing information to employers about the employability and level of performance that can be expected from graduates. However, an evolving 21st century labor market has created an environment where employers are requesting even greater accountability from educational institutions in order to quantify employability in terms of traditional letter-grade ranking systems. Employers in the 21st century are emphasizing the point that hiring competent graduates, with the appropriate skills and personal attributes that are relevant to the workforce, is paramount to the success of their business, since poor recruitment decisions can prove costly in commercial terms. Therefore, it is reasonable to assume that, from an employer's perspective, attributes of employability should be given context within academic grading systems, in that grading should clearly measure the skills and personal attributes required by employers and are therefore needed by today's graduates in order to secure gainful employment (Ghani et al., 2018; Lisá et al., 2019).

In response to the general growing demand of a labor market which seeks competent graduates, many colleges and universities worldwide have sought to implement the concept of competency into their curricula. Competency-based education (CBE), according to the published literature, links theory with practice (Gervais, 2016). Competency, from the perspective of the employer, refers to the visible elements and underlying characteristics that drives superior job performance in their employees (Pang et al., 2019). According to Pazil and Razak (2019), employers seek versatile graduates who will be able to help drive their organization forwards in order to compete successfully in today's continually evolving business environment. Chan et al. (2018) shared a similar perspective by highlighting that college and university graduates worldwide are expected to have both the academic qualifications and corresponding skills in order to be considered employable.

One challenge with providing CBE when using a traditional letter-grade ranking system is ensuring homogeneous letter grading that will reap the expected rewards for both graduates and employers. It is known to be difficult to ensure that an **A**-grade awarded by one educational institution has the same weight and meaning across domestic, regional, and international educational systems. It is fair to assume, therefore, that differences in social, political, and economic standards, in most cases, may impact upon academic quality domestically, regionally, and internationally.

For many of the colleges and universities of the Small Island Developing States (SIDS) of the Caribbean, developing and employing competency-based curricula to meet the international standards set by developed nations are hampered mainly for two reasons. SIDS must first meet the international requirements for their educational institutions to become accredited or to maintain their accreditation. Currently, institutional accreditation and program accreditation are considered a global standard for educational institutions to show that they are capable of providing and delivering high quality education. Secondly, SIDS must provide proof of successful measures of assurance of learning to show that the assessment processes are conducted in a transparent way. It is therefore crucial for SIDS to secure international accreditation in order to make sure that their graduates are hirable globally, and to implement appropriate policies if they are to ever attain the status of being fully developed. As a result, SIDS must prove that they can adequately design and quantify the use of the Letter Grading System (LGS) in their educational systems, and to provide

assurance of learning measures that meet global demands if they are to stay in the hunt to gain fully-developed status. Against this backdrop, the current study attempts to identify the critical elements of the LGS that SIDS should focus on so as to be able to communicate effectively the achievement levels of graduates in terms of the skills and personal attributes that employers seek and that graduates need in order to secure employment globally

The objective of this study is to highlight some of the elements associated with Fuzzy Letter Grading Systems and to use this information in order to better understand what is needed for SIDS to create homogenous grading systems with quantifiable measures that show how their College and University graduates are adequately prepared to secure global employment. This research is believed necessary to provide the vital elements needed by educational administrators and policymakers in the SID islands so they may reflect critically on their practice and how to construct new understandings for Caribbean higher education in the context of competency-based education, pedagogy, and the tangible skills of learners within the framework of a global market.

2. LITERATURE REVIEW

The Letter Grading System is a widely accepted grading method used in education (Kafle, 2020). Still, several researchers have continued to investigate and evaluate grading systems with respect to student academic performance and the general theory of assurance of learning. Assurance of learning pertains to making continuous improvements to curricula so as to ensure that essential student competencies are obtained. Assurance of learning focuses on the knowledge and skills of students, rather than on the course content itself. Assurance of learning aims to ensure that educational institutions provide clear learning goals for their students, that they are able to appropriately assess student achievement of these goals, and to address any disparities between the goals set and what they were actually able to attain.

The European Credit Transfer System (ECTS), as highlighted in the works of Opstad (2020), is considered a student-centered LGS system that is based on the student workloads required to achieve the objectives of an educational program of study. The objectives themselves are specified in terms of the learning outcomes and competencies that the students are expected to acquire. The ECTS is considered a dual system. It is a system that uses credits to describe the duration of a program in as much as it is a 7-step grading system that indicates the performance of students. It may be inferred from Opstad (2020), that the ECTS is a type of open system that can be adapted to different LGS models, and can therefore also serves as an interpretative scale. Table 1 presents an overview of the ECTS grading systems, as depicted by Opstad (2020).

Table 1. ECTS Grading System

Grade	%	Description	General, qualitative description of evaluation criteria
F		Fail	A performance that does not meet the minimum academic criteria. The candidate demonstrates an absence of both judgment and independent thinking.
E	10	Sufficient	A performance that meets the minimum criteria, but no more. The candidate demonstrates a very limited degree of judgement and independent

			thinking.
D	25	Satisfactory	A satisfactory performance, but with significant shortcomings. The candidate demonstrates a very limited degree of judgement and independent thinking.
C	30	Good	A good performance in most areas. The candidate demonstrates a reasonable degree of judgement and independent thinking in the most important areas.
B	25	Very Good	A very good performance. The candidate demonstrates sound judgement and a very good degree of independent thinking.
A	10	Excellent	An excellent performance, clearly outstanding. The candidate demonstrates excellent judgement and a high degree of independent thinking.

According to Kafle (2020), teacher accountability is critical to implementing letter grade systems for the purpose of assessment. Kafle (2020) seemed to suggest that grading systems are intrinsically linked to assessment. It is a fair assumption to make that assessment forms the foundation of most academic grading systems, hence grading and assessment should in general be discussed simultaneously, especially where the goal is the translation of grading system output into quantifiable measures of graduates' skills and personal attributes. Fletcher et al. (2012) suggested that assessments are tied to a person's beliefs, meanings, and understanding of assessment; hence may therefore be subjective. As stated by McDonald et al. (2014), some faculty simply do not value or understand the purpose of assessment. Therefore, a case may be made that the structure of grading systems and the design of assessments can obscure the quality of the actual measures used in assessing skills and personal attributes of graduates, because all assessments are not created nor weighted equally. If these statements are accepted as true, then it is highly probably that an **A**-grade awarded by one institution may not signal the presence of the same associated abilities by a different student receiving an **A**-grade at another institution.

In order to illustrate this point, Table 2 shows a grading system used in the SIDS context. In a higher education system, the purpose of the LGS is primarily to communicate the level of achievement of graduates with respect to the institution's learning goals particular to a certain course of study. As a point of reference, the current study randomly selected a grading system currently used by some SIDS, which was then compared to the ECTS system (see Table 1) for the purpose of highlighting how the interpretation of Letter Grades may differ significantly in the way that institutions may use them, both domestically and internationally, to describe the abilities of their graduates.

Table 2. Generic Letter Grading System (GLGS)

Letter	Marks	Definition	Points
A+	90 - 100	Exceptional	4.00
A	80 - 89	Excellent	4.00
A-	75 - 79	Very Good	3.75
B+	70 - 74	Good	3.50

B	65 - 69	Fairly Good	3.00
B-	60 - 64	Fairly Good	2.75
C+	55 - 59	Satisfactory	2.50
C	50 - 54	Satisfactory	2.00
D	45 - 49	Passing	1.00
F	0 - 44	Failing	0.00

Through direct comparison of the ECTS (see Table 1) grading system and the GLGS (see Table 2), it is apparent that the GLGS does not provide a detailed qualitative description of its evaluation criteria in the same way that the ECTS systems does. Hence, it is reasonable to conceive that the GLGS, as presented in Table 1, does not adequately describe the value of a graduating student's abilities. The GLGS system, it seems, fails to quantify the quality of graduates in terms of the core, essential, or key skills as expressed in the ECTS grading system. It is generally accepted that grades should properly characterize levels of skill and mastery. Therefore, it may be argued that to simply use classifications such as "Satisfactory" and "Fairly Good" without providing a corresponding detailed descriptor of the evaluation criteria may result in employers most likely viewing ranking information from a Letter Grade System (as in the GLGS) to be vague and potentially misleading in providing adequate information to employers concerning the measures applied to the skills and personal attributes of graduates.

To provide greater context to evaluating the GLGS, further consideration was given to the works of Fletcher et al. (2012), in which they suggested that assessments are tied to a person's beliefs, meanings, and understanding of assessment. In contrast, McDonald et al. (2014) stated it as being the value of faculty understanding the purpose of assessment. According to DeLuca et al. (2016), it is important to evaluate teacher assessment literacy within an accountability framework. Assessment literacy, as defined by DeLuca et al. (2016), is the ability to construct reliable assessments, and to then administer and score these assessments in order to facilitate valid instructional decisions anchored to state or provincial educational standards.

For the purposes of this study, the principle extracted from the works of Fletcher et al. (2012), McDonald et al. (2014), and DeLuca et al. (2016) was that whilst high-quality skills and mastery was attainable, it could only be adequately reflected within a LGS if there was satisfactory alignment between what the students learned and what ultimately appeared on their assessments. It may be said that Contino (2012) and also Liu et al. (2009) reported similar perspectives concerning the notion of accurate alignment and assessment. In the view of Contino (2012), an aligned system's objective is to ensure that students receive a coherent and integrated learning experience in order that they may achieve a significantly high level of skills and mastery. It is reasonable, therefore, to assume that Contino (2012) believed that a coherent message and system will positively influence what is taught in the classroom, and thereby what students could eventually learn. Similarly, Liu et al. (2009) suggested that having a coherent message and a uniform language for describing instruction, assessment, instructional materials, and content standards makes it possible to construct meaningful measures that adequately reflect both the soft and hard skills that employers seek in today's graduates.

Based on these perspectives, it was assumed that there are several critical steps to measuring graduates' hard and soft skills and, if one of these steps is weak or missing, then

there is the potential for a Letter Grading System to misrepresent what skills were learned by the graduate and how these skills translate from the classroom to the prospects of a student being employable. In order to better quantify this assumption, the current study further investigated whether or not the GLGS, which was comparatively shown to lack precise qualitative descriptors of its evaluation criteria, was still sufficiently adequate to confidently and accurately communicate to employers, both domestically and internationally, the skills and personal attributes of its graduates according to what was taught in the classroom and what the students eventually learned after having completed their course of study.

3. METHODOLOGY

To better understand what is needed for SIDS to create homogenous grading systems with strong quantifiable qualitative descriptors that truly reflect the skills and personal attributes of their college and university graduates, Bayesian statistical analysis was applied to end-of-semester grades randomly selected from one of the SIDS island's colleges. The idea being to apply Bayesian statistics to look for evidence pertaining to the quality of academic assessments, and the potential impact that assessments may have on the adequate measurement of graduates' skills and attributes. The use of Bayesian statistics is considered a robust technique for the provision of empirical evidence. According to Kass and Raftery (1995) and also Stefan et al. (2019), the Bayes factors of Bayesian statistics measure strength of evidence, quantifying evidence in favor of one statistical model compared to another, and to thereby reveal the ratio of the two marginal likelihoods. That is, the likelihood of the data under the null hypothesis $(\mathcal{H}_0)$ and the likelihood of the data under the alternative hypothesis $(\mathcal{H}_1)$.

As such, the Bayes factor is actually the factor by which prior odds are multiplied in order to obtain posterior odds. For this reason, Bayes factors may be considered as the quantification of how the prior belief regarding a dataset should be adjusted according to a particular set of data. Stefan et al. (2019), asserts that Bayes factors can distinguish between absence of evidence and the evidence of absence. That is, Bayes factors focus on what level of evidence is deemed the most compelling. According to Stefan et al. (2019), Bayes factors provide researchers with a degree of freedom in formulating the relevant hypotheses since it does not require the two models to be nested. For example, Jacob and Levitt (2003) regard Bayesian statistical analysis as having the potential to detect illicit acts of teachers cheating despite the best attempts of the perpetrators to keep them clandestine.

The fundamental difference between traditional statistics and Bayesian statistics is the way they view the parameters of interest. In traditional statistics, variance is a fixed unknown constant with no randomness. A limitation with this perspective is that it is difficult to make confident statements about the uncertainty or degree of belief concerning the parameters of interest. In contrast, Bayesian statistics allow for the expression of randomness in parameters of interest; thus providing the advantage of expressing uncertainty through the use of probability distributions. Additionally, Bayesian statistics, unlike traditional statistics, offers the probability that a particular hypothesis is true given a certain dataset, $p(D|H_k)$. Traditional statistics cannot be used to define or compute $p(D|H_k)$ since assigning the probability of being true to the hypothesis is not allowed in traditional statistics. However, as Bayesian statistics allow for defining a degree of belief about some event by assigning a probability to that event, it underpins the heart of

the current research, which is, understanding the probability that a hypothesis is true before or after observing the data. For the current study, the data were analyzed under the null hypothesis $(\mathcal{H}_0)$: proper grading practices of well-aligned coursework and final exams allow students to attract comparable marks on both forms of assessment. The associated Bayes factors are then used as a measure of evidential strength, which is considered highly advantageous by various researchers since Bayes factors can quantify evidence for both $(\mathcal{H}_1)$ and $(\mathcal{H}_0)$.

In Bayesian statistics, Bayes factors are used as a tool to quantify evidence of the null hypothesis against the alternative, or vice versa. The Bayes factor is a ratio of the posterior odds to the prior odds. That is, a Bayes factor may be considered the factor by which the prior odds are multiplied by in order to give the prior odds. This means that because the prior odds and posterior odds are the ratio of beliefs about the hypothesis before and after observing the data, Bayes factors may be taken to represent how much of an adjustment should be made to the prior beliefs given the data. The Bayes factor robustness check allows for the visualization of how the Bayes factor changes for a wide range of prior widths. It acts as an indicator for the strength of evidence supporting $\mathcal{H}_1$ instead of $\mathcal{H}_0$. Bayes factors may essentially be considered an updating factor for prior beliefs (Stefan et al., 2019). For example, in this analysis it was assumed that both $\mathcal{H}_1$ and $\mathcal{H}_0$ are equally probable priori such that $\mathcal{H}_0 = \mathcal{H}_1 = 0.5$. In this case, $\mathcal{H}_1$ and $\mathcal{H}_0$ are point hypotheses as they assume that the effect size has a specific value. However, it was assumed that the parameter under investigation actually lies within a certain range of values. Under this assumption, the hypotheses reflect a distribution, called the prior distribution, that assigns a probability density to the parameter values.

The Bayes factor is then calculated according to Equation 1:

$$p(D|H_k) = \int p(D|\theta_k, H_k)\, \pi(\theta_k|H_k)\, d\theta_k \qquad (1)$$

where, θ_k is the parameter under H_k and $\pi(\theta_k|H_k)$ is its prior density (Stefan et al., 2019). Equation 1 represents the probability that the parameter value is equal to θ_k before observing the data, and that $p(D|H_k)$ is the posterior probability density of data D given a specific value of parameter θ_k after observing the data. The Bayes factor robustness check plots the Bayes factor against a range of values for the prior width. A Bayes factor above a certain upper boundary is considered sufficient evidence in favor of the alternative hypothesis $(\mathcal{H}_1)$, whilst a Bayes factor below the corresponding lower boundary represents evidence in favor of the null hypothesis $(\mathcal{H}_0)$. If the Bayes factor lies between these boundaries, then the evidence is deemed to be inclusive or anecdotal. By applying Bayesian updating, it is possible to update the probability that the parameter value is equal to each possible value of θ_k such that the posterior distribution is obtained. The posterior distribution is the updated probability or degree of belief about the parameter of interest after observing the data. Essentially, Bayesian statistics provides a way to compute the probability that a hypothesis is true given a certain set of data. For this preliminary investigation, eight datasets were analyzed in the current study; however, only the results for three datasets are presented so as to avoid repetitive results. It is important to note here that the data used was de-identified so as to avoid potential identification of the associated courses and origins of the data. As such, the data used remained completely random and

anonymous. As a result, any potential identifiable characteristics were replaced with generic labels such as A-C and the designators such as Alpha 1-Alpha 3.

The general underlying theoretical framework for the methodology chosen in the current study is alignment theory. Studies have successfully shown that when students are assessed via a combination of coursework and examinations, they tend to attract higher marks on coursework assessments than in their final exams (Alsuwaiket et al., 2019). Although this is one potential finding, alternative views suggest proper alignment of assessments should result where coursework and end-of-semester examinations test the same knowledgebase and skills, even if the final exams address the same skills at a higher or lower complexity than the coursework (Murdan, 2005). Assuming that this is true, and that the viewpoints of Murdan (2005) and Alsuwaiket et al. (2019) also holds, then it is reasonable to believe the argument that with proper alignment between the two assessment types, students should have a "true score" or long-term average assessment score providing learning had not occurred between the time of the various assessments. However, it is also quite feasible that even though proper alignment should yield scores that are fairly consistent, there will be occasions where students will attract scores other than their true level. For the purposes of this investigation, the current study assumes that faculty members adhered to the alignment protocols of the institution and that the students were given high-quality aligned coursework assessments and final examinations during the academic year. On this premise, it was assumed that a student should have a "true score" or long-term score average from a series of assessments.

4. RESULTS

In total, 246 completed questionnaires were analyzed following data cleansing. From the participant teachers purposively selected from secondary schools (I-IV) in the targeted study areas, of those who responded to the survey, 52.8% were female and 47.2% were male.

The three graphical representations depicted in Figure 1 represent the Bayesian correlation pairs and scatter plot of the coursework and end-of-semester exam, the Bayes factor robustness check plot, and the sequential analysis plot, respectively. Figure 1 reports a large effect (.63) correlation between the end of semester exam mark and the coursework mark. These two variables produced an associated Bayes factor $(\mathrm{BF}_{10} = 26.359)$ in favour of the alternative hypothesis.
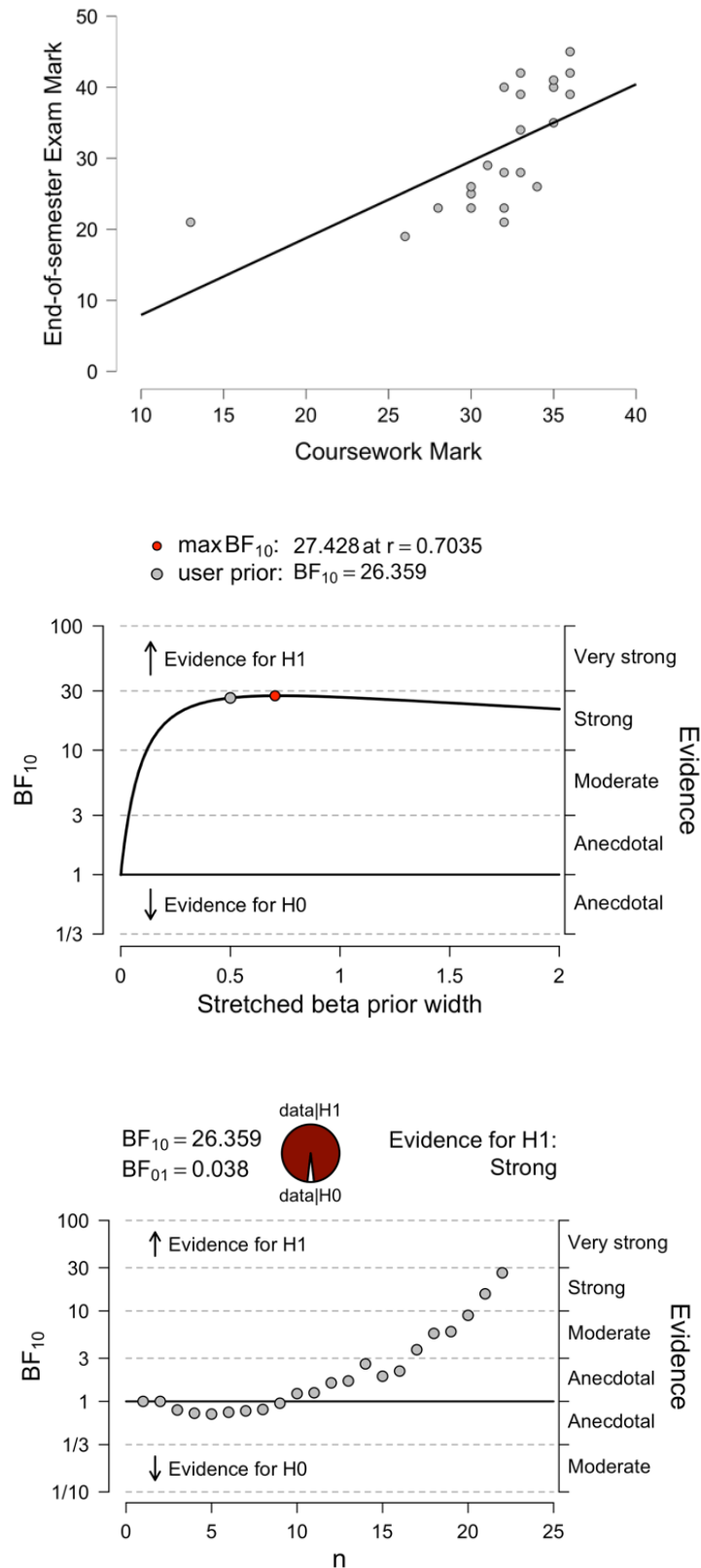
Figure 1. Bayesian Statistical Analysis of Alpha 1: Course A

In the case of Alpha 1, the alternative hypothesis predicts the data approximately 26 times better than the null hypothesis. This may be interpreted that in the absence of proper alignment, the coursework marks that a student attracts is not comparable to their end-of-

semester exam mark. Given this result, it was decided to look for evidence to support non-alignment in the data submitted for Alpha 1. In order to do this, the corresponding scatter plot, Bayes factor robustness check plot, and the sequential analysis plots were closely examined.

The scatterplot shown in Figure 1 indicates a positive trend between the two variables of coursework and end-of-semester exam mark. To see how robust these results are, the Bayes factor robustness check plot was utilized in order to look at how the Bayes factor changes for a range of priors. For Alpha 1, the Bayes factor robustness plot indicates that the Bayes factor does not vary significantly unless the prior width is very small. For most values of the prior width, the Bayes factor is shown to have a value exceeding 10. This is an indication of the existence of strong evidence in support of the alternative hypothesis. Since the conclusion does not change for a wide variation of the prior width, it may be concluded that the results are robust.

To find evidence pertaining to alignment between coursework and final exam assessments for Alpha 1, the sequential analysis plot was considered. The sequential analysis plot allows for the visualization of how the Bayes factor changes after every data point. In the case of Alpha 1, the development of the Bayes factor is fairly smooth for six of the 22 reported scores. This means that only approximately 27% of the data was considered to be robust, suggesting that the students did not achieve an average score from the two assessments. Approximately 73% of the data reported by Alpha 1 was therefore considered anecdotal. This means, according to the usages of alignment descriptors, that a relatively significant portion of Alpha 1's results do not measure any aspects of the reported standard. In this case, no inferences can be objectively made about 73% of the students' performances in that particular course given their coursework performance.

Figure 2 depicts the results of Alpha 2, and reports a large effect (.51) correlation between the two variables of coursework and end-of-semester exam mark. These two variables produced an associated Bayes factor $(BF_{10} = 40.628)$ in favour of the alternative hypothesis. In the case of Alpha 2, this means that the alternative hypothesis predicts the data 40 times better than the null hypothesis. Similar to the case of Alpha 1, this may be interpreted that in the absence of proper alignment, the coursework mark that a student attracts is not comparable to their end-of-semester exam mark. Given this result, it was decided to look for evidence to support non-alignment in the data submitted for Alpha 2. The scatterplot shows a positive trend between the two variables of coursework and end-of-semester final exam mark. For Alpha 2, the Bayes factor robustness plot indicates that the Bayes factor does not vary significantly unless the prior width is very small. For most values of the prior width, the Bayes factor exceeds a value of 40. This is an indication of the existence of very strong evidence in support of the alternative hypothesis. Since the conclusion does not change for a wide variation of the prior width, it may be concluded that the results are robust.

In the case of Alpha 2, the development of the Bayes factor is fairly smooth for four out of the 38 reported scores. This means that only approximately 11% of the data was considered to be robust, and suggests that students did not achieve an average score on the two assessments. Approximately 89% of the data reported by Alpha 2 was therefore considered anecdotal. This means that a significant portion of Alpha 2's results do not measure any aspects of the reported standard. In this case, no inferences can be objectively

made regarding 89% of the students' performance in that particular course according to their coursework performance.
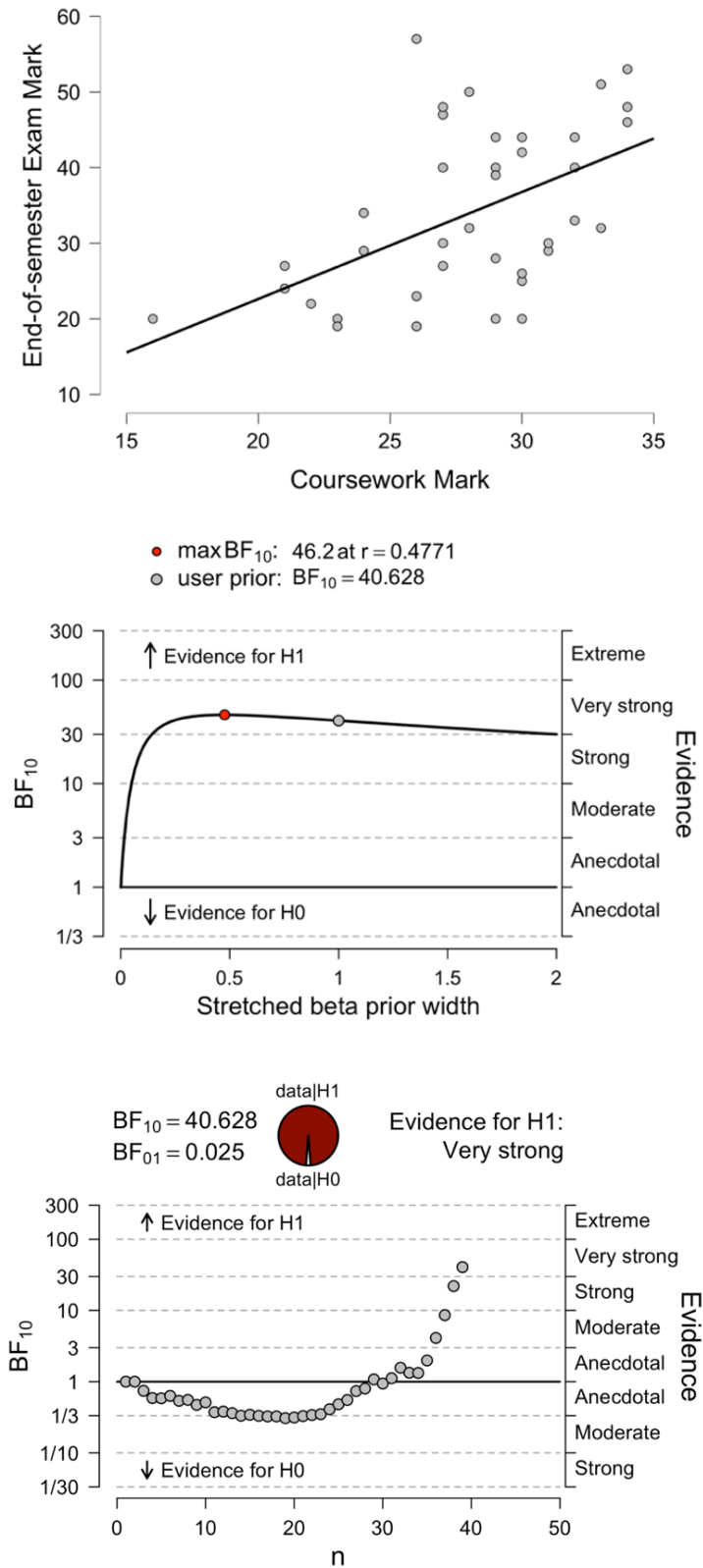

Figure 2. Bayesian Statistical Analysis of Alpha 2: Course B

For Alpha 3, the Bayesian correlation pairs report a large effect (.67) correlation between the two variables of coursework and end-of-semester exam mark. These two variables produced an associated Bayes factor $(\mathrm{BF}_{10} = 79.283)$ in favour of the alternative hypothesis.
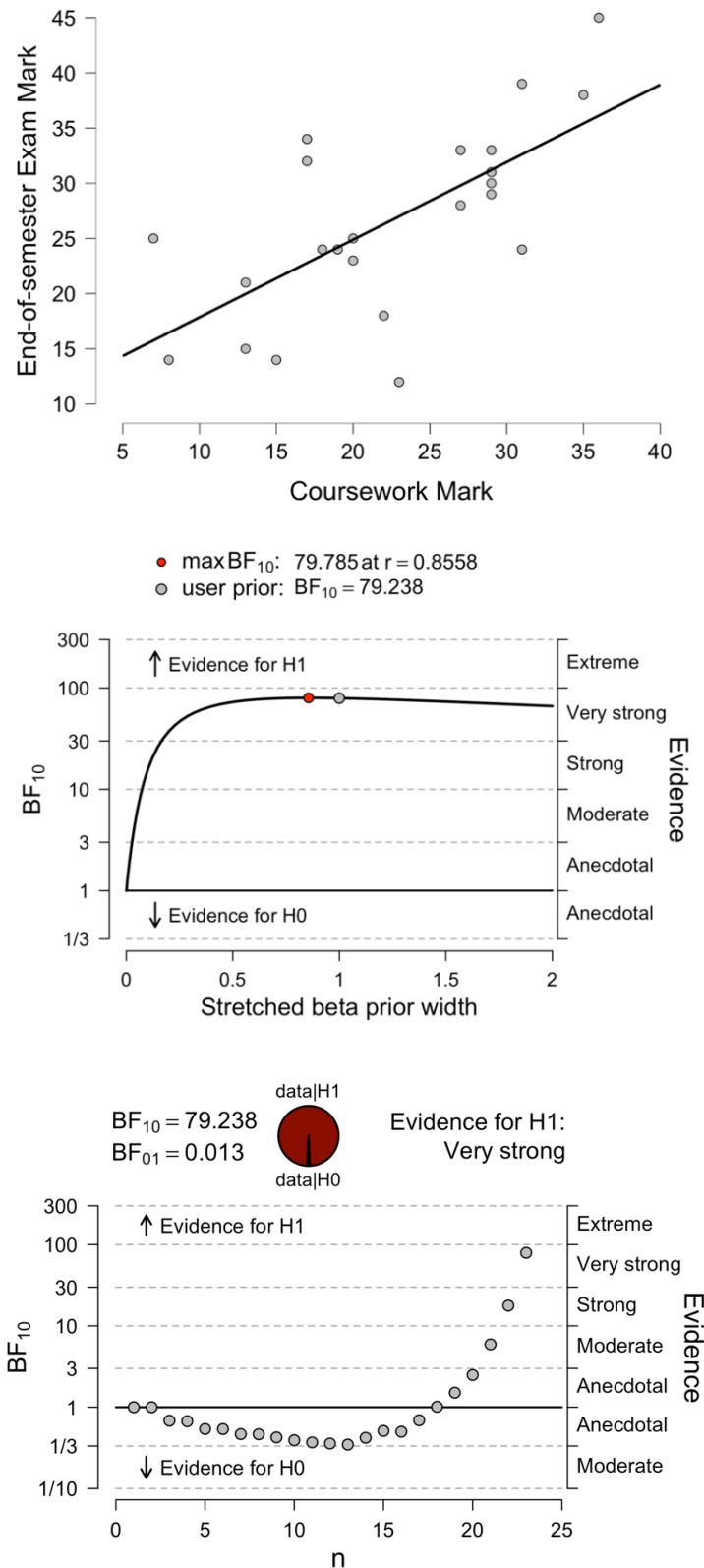

Figure 3. Bayesian Statistical Analysis of Alpha 3: Course C

This means that, in the case of Alpha 3, the alternative hypothesis predicts the data to be 79 times better than the null hypothesis. This may be interpreted that in the absence of proper alignment, the coursework marks that students receive may not be comparable to their end-of-semester exam mark. Given this result, it was decided to look for evidence to support non-alignment in the data submitted for Alpha 3. The scatterplot shows a positive trend between the two variables of coursework assessment and final end-of-semester exam mark. For Alpha 3, the Bayes factor robustness plot indicates that the Bayes factor does not vary significantly unless the prior width is very small. For most values of the prior width, the Bayes factor exceeds a value of 40. This is an indication of the existence of very strong evidence in support of the alternative hypothesis. Since the conclusion does not change for a wide variation of the prior width, it may be concluded that the results may be considered robust.

In the case of Alpha 3, the development of the Bayes factor is fairly smooth for four out of the 38 reported scores. This means that only approximately 13% of the data was considered to be robust, and suggests that the students did not achieve an average score on the two types of assessment. Approximately 87% of the data reported for Alpha 3 was therefore anecdotal. This means that a significant portion of Alpha 3's results do not measure any aspects of the reported standard. In this case, no inferences can be objectively made regarding 87% of the students' performance in that particular course according to their evaluated coursework performance.

5. DISCUSSION

The design of grading systems vary in their philosophies from country to country. However, despite the underlying viewpoint, the fair and sound interpretation of assessments and academic grading standards is essential to any graduates aiming to join a profession in today's global workforce. Employers of the 21st century often now require credential evaluation criteria of college and university graduate applicants that explain how the academic performance of the graduates were measured, and more importantly, how this translates into qualities they seek as employers of new graduates to hire. Specifically, employers want to see transcripts that convey a message of mutual trust and confidence in the fundamental needs of the labor market, rather than traditional transcripts that merely reflect the technical requirements that a student completed as part of their degree program or course.

In order to investigate what elements of the grading system SIDS should consider that are pertinent to developing reliable ranking systems that best reflect the academic and professional competences that today's international employers seek, Bayesian statistical analysis was applied to randomly selected and anonymous grading data. The idea behind the current research was that Letter Grading Systems (LGS) should be so designed that they convey a message of clarity, and as such a LGS chosen for ranking graduates should render that message homogenously in terms of the domestic employment labor market, as well as both regionally and internationally. The intention of the current study was not to present a solution to the challenges of universally interpreting Letter Grading Systems, but rather to study the highlights of the many critical areas that Small Island Developing States (SIDS) of the Caribbean region need in order to address as they pursue a path to becoming considered "fully developed." Therefore, the results of the current study are meant to be interpreted in the context of the needs of the Caribbean region's SIDS in addressing international

employment requirements so as to make them more recognizable and competitive for the international labor market.

For brevity, the article presents only three datasets. However, the study included analysis of eight datasets in total. The results from all eight datasets were shown to be anecdotal overall. This unusual and unexpected outcome raises questions pertaining either to quality of the participant students' academic abilities, or brings into question the grading system used to quantify their academic assessments. According to Kafle (2020), LGSs are supposed to lead to a better engagement of ideas, to help make classwork easier, and to provide information about learners' weaknesses and strengths.

An overall shortcoming of the LGS was highlighted by Kafle (2020), reporting results that are pertinent to those of the current study; that LGSs are susceptible to not being considered accurate in their performance measurement. It is believed that the results of the current study could be seen as an indicator of a lack in accuracy in the measurement of students' academic performance. The results, being anecdotal, appear to suggest that grading could have occurred on the basis of some other parameter such as the inflation or deflation of assessment grading marks. This statement acknowledges that there may be several other factors that can impact upon a grading system, whilst not intending to highlight, address, or speculate about the integrity or ethical conduct surrounding faculty members across various educational institutions. The statement is purely speculative and should be taken as supportive of the idea that grading is generally a subjective process, and that LGS results may include a certain degree of bias. Ornstein (1994) suggested that grading practices should be considered as inherently subjective evaluative judgements of student performance, regardless of the grading policies and practices applied. To illustrate this point, consider the scatter plot for Alpha 3 (see Figure 3). The courses carry a 40% coursework and 60% final examination mark weighting. For Alpha 3, it may be assumed that in the same cohort, the weakest coursework student, who scored approximately 7% out of the 40% coursework maximum, was able to attract 25% of the 60% allocation of marks for the final examination. While, at the same time, students scoring more than 20% of the allotted coursework marks could only achieve a little over 10% of the 60% final exam marks allocation. Yet, a few students obtained scores that were in agreement with what alignment theory predicts.

In the current study, a general trend appeared in the scatter plots for all eight faculty members, suggesting that the weaker students were able to transfer the knowledge and skills they acquired at the point of their coursework assessment to their final exam, while the stronger students were unable to transfer the same knowledge and skills from their coursework to their final exam. However, this general pattern does not support the logical educational argument, that a student who works diligently and consistently during the semester and achieves appropriately high coursework marks should be adequately prepared for their end-of-semester examinations, and thereby go on to attain a high overall final grade for their program of study, or vice versa. It is on this basis that it is believed that Bayesian statistical analysis of this particular dataset suggests that grading systems are too generic in their evaluation criteria; for example, the GLGS, are considered fuzzy and inadequate for reliably communicating the level of academic mastery and competence that today's international employers seek.

5. CONCLUSION AND SUGGESTION

The current research examined the reliability of the GLGS to translate the academic achievements of college graduates into hirable competencies that international employers may seek in today's labor market. It was found from the datasets analyzed that, in the context of the SIDS of the Caribbean region, the GLGS was inadequate for the purpose of reliably communicating the level of academic mastery and competence needed by graduating students for employability within the international labor market. Additionally, the results suggested that it was not possible to construct meaningful indices of alignment in the data that could quantify the quality of the assessments, instructional materials, and content standards associated with the use of the GLGS. This discovery is significant for SIDS to address, as it is widely believed that meaningful indices of alignment are important measures that may be used to translate what is taught in the classroom to what students ultimately are capable of learning.

Given these limitations, it is believed that SIDS should probably increase their emphasis on aiming for their graduates to be globally recognized, that since qualitative description of evaluation criteria for grades can vary significantly between different institutions and also regions and countries. It may all serve SIDS well if their graduates' grade interpretations are more guided by objective to be fair in terms of the technical merits of their program of study they recently graduated from rather than overly concentrating on achieving a correct equivalent conversion of the graduates' skills and attributes according to an international or regional framework.

Despite this argument, one pertinent question that still remains unanswered for the SIDS of the Caribbean, is how can fairness be measured and achieved between nations with different grading systems where differing evaluation criteria are used? It is believed that until domestic, regional, and international educational bodies implement some universal grade interpretation scheme, SID nations may continue to consider that their graduates are undervalued in academic terms, and probably misrepresented in the labor market when seeking employment either regionally or internationally.

In the current research, an attempt was made to address the need for reliable techniques in order to acquire the necessary data needed to potentially answer the underlying question of how do we really know what graduates have learned after having taken a certain course or program or study that was planned and specifically guided by learning outcomes, assessments, and traditional Letter Grading Systems? Specifically, the current study has investigated the attributes of the qualitative descriptive structure of today's LGS used by the SIDS, and the relationship between the content and structure of the assessment that students potentially then receive. The reported results suggest that the use of generic grading systems with broad quality labels that are loosely assigned to numerical grade categories are less successful in articulating the applied assessment criteria, and that they do not confidently capture the skills and personal attributes of graduates seeking employment in the regional or international labor market. As a result, the current study recommends that SIDS pay urgent attention to enhancing the various classes, categories, and quality labels currently in use for their Letter Grading System, so that critical information pertaining to the actual measured abilities and skills of their graduates can be correctly interpreted by domestic, regional, and international employers. Additionally, it is suggested that SIDS implement systems of assurance of learning so as to ensure that they have more comprehensive indicators available regarding students' academic performance that can

accurately convey valid and consistent messages to employers regarding the assessments that graduates receive, and the grade rankings that they obtain.

DECLARATIONS

Author Contributions The article was written by a single author, who read and approved the final published version of the article.

Conflicts of Interest The authors report no conflicts of interest.

Ethical Approval Chairperson Board of Management: BOM/06.28.001/DL/2021.

Funding None.

Data Availability Statement The datasets generated during and/or analyzed in the current study remain the property of the educational institution, and are therefore not publicly available due to student education records being considered official and confidential documents protected by the island's privacy protection law. However, data may be made available from the corresponding author through approval by the relevant educational institution upon reasonable request.

Acknowledgments None.

REFERENCES

Alsuwaiket, M., Blasi, A. H., & Al-Msie'deen, R. A. (2019). Formulating module assessment for improved academic performance predictability in higher education. *Engineering, Technology & Applied Science Research, 9*(3), 4287-4291. https://doi.org/10.48084/etasr.2794

Chan, S. W., Ahmad, M. F., Zaman, I., & Ko., W. S. (2018). Employers' perception on important employability skills in the manufacturing industry. *International Journal of Engineering and Technology, 7*(2-29), 170-175. http://dx.doi.org/10.14419/ijet.v7i2.29.13311

Contino, J. (2012). A case study of the alignment between curriculum and assessment in the New York State Earth Science standards-based system. *Journal of Science Education and Technology, 22*(1), 62-67. https://doi.org/10.1007/s10956-012-9376-x

Fletcher, R. B., Meyer, L. H., Anderson, H., Johnston, P., & Rees, M. (2012). Faculty and students' conceptions of assessment in higher education. *Higher Education, 64*(1), 119-133. https://doi.org/10.1007/s10734-011-9484-1

Gervais, J. (2016). The operational definition of competency-based education. *The Journal of Competency-Based Education*, *1*(2), 98-106. https://doi.org/10.1002/cbe2.1011

Ghani, E. K., Rappa, R., & Gunardi, A. (2018). Employers' perceived accounting graduates' soft skills. *Academy of Accounting and Financial Studies Journal, 22*(5). https://www.abacademies.org/articles/employers-perceived-accounting-graduates-soft-skills-7582.html

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics, 118*(3), 843-877. https://doi.org/10.1162/00335530360698441

Kafle, B. (2020). Teachers' Perception of Letter Grading System and Its Challenges a Qualitative Study in Vyas Municipality of Tanahun. *American Journal of Educational Research*, *8*(9), 622-632. https://doi.org/10.12691/education-8-9-3

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795. https://doi.org/10.1080/01621459.1995.10476572

Lisá, E., Hennelová, K., & Newman, D. (2019). Comparison between employers' and students' expectations in respect of employability skills of university graduates. *International Journal of Work-Integrated Learning, 20*(1), 71-82. https://www.ijwil.org/files/IJWIL_20_1_71_82.pdf

Liu, X., Zhang, B., Liang, L. L., Fulmer, G., Kim, B., & Yuan, H. (2009). Alignment between the physics content standard and standardized test: a comparison among US–NY, Singapore, and China-Jiangsu. *Science Education, 93*(5), 777-797. https://doi.org/10.1002/sce.20330

MacDonald, S. K., Williams, L. M., Lazowski, R. A., Horst, S. J., & Barron, K. E. (2014). Faculty attitudes toward general education assessment: A qualitative study about their motivation. *Research & Practice in Assessment*, *9*(1), 74-90. https://www.rpajournal.com/faculty-attitudes-toward-general-education-assessment-a-qualitative-study-about-their-motivation/

Murdan, S. (2005). Exploring relationships between coursework and examination marks: A study from one school of pharmacy. *Pharmacy Education, 5*(2), 97-104. https://pharmacyeducation.fip.org/pharmacyeducation/article/view/112

Opstad, L. (2020). Why are There Different Grading Practices Based on Students' Choice of Business Major? *Educational Process: International Journal, 9*(1), 43-57. https://doi.org/10.22521/edupij.2020.91.3

Ornstein, A. C. (1994). Grading practices and policies: An overview and some suggestions. *NASSP Bulletin, 78*(561), 55-64. https://doi.org/10.1177%2F019263659407856111

Pang, E., Wong, M., Leung, C. H., & Coombes, J. (2019). Competencies for fresh graduates' success at work: Perspectives of employers. *Industry and Higher Education, 33*(1), 55-65. https://doi.org/10.1177%2F0950422218792333

Pazil, A. H. M., & Razak, R. C. (2019). Perspectives of Asian employers on graduates' soft skills: A systematic review. *Universal Journal of Educational Research, 7*(11), 2397-2405. https://doi.org/10.13189/ujer.2019.071117

Stefan, A. M., Gronau., Q. F., Schönbrodt., F. D., & Wagenmakers, E.-J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods, 51*(3), 1042-1058. https://doi.org/10.3758/s13428-018-01189-8

ABOUT THE CONTRIBUTORS

Kyffin Bradshaw, PhD, is a Lecturer of Physics at the Barbados Community College (Barbados). His main research interests pertain to computational methods and experimental measurements with emphasis on their development and integration. His works have been published in a major international journal.
E-mail: Kyffin.bradshaw@bcc.edu.bb
ORCID ID: https://orcid.org/0000-0002-7449-2007