

## Original Research Article

### Demystifying the Need and Efficacy of Trial Examinations: Perspectives of Students and Teachers and Their Statistical Relationships with Examination Scores

#### Abstract

Trial or mock examinations for classes 10 and 12, conducted months before students appear for a high stake, externally administered and evaluated board examinations, came under scrutiny after an educational strategic document recommended a review on its need and efficacy. Due to a lack of literature on the need and efficacy of trial examinations in Bhutanese educational context, a qualitative inquiry employing interviews with principals and focused group discussions with teachers was conducted to identify potential theories. Teacher accountability, assessment feedback, and prediction of students' achievement in board examinations were used to justify the need and efficacy of trial examinations, while waste of instructional hours, teacher and student stress were the adverse consequences purported. The objectives of the research were to quantitatively test these assumptions and explore statistical relationships between trial examination and board examination scores. Primary data was collected from 261 teachers and 562 students. Trial and board examinations academic achievement results of 2432 class 10 and 557 class 12 students were collected. Descriptive statistics as well as inferential statistics were used to analyze the data. Results indicate that teachers perceive trial examinations as a necessary evil through which they learn more about students' knowledge gaps to act upon, rather than being anxious about student achievement. Both students and teachers perceived that trial facilitates improved student achievement, when measured through examination marks. Trial examination marks demonstrate positive, weak to moderate correlation to board examination marks and the relationships are statistically significant. The perceptions that the planning and conduct of trial examinations are a waste of instructional time, lead to teacher and student stress were obtained to be myths. This research contributes to existing literature that low stakes testing programs facilitate improvement in student scores.

#### Keywords

*Trial or mock examinations, high stakes board examinations, teacher and student stress, predictive validity, teacher accountability, instructional hours*

#### Introduction

The Constitution of the Kingdom of Bhutan stipulates that basic education will be free up to the tenth standard and higher education will be made available based on merit (Royal Government of Bhutan, 2008). Even before being pinned as a constitutional mandate, education was provided free of charge and students' performance in standardized examinations, also called

board examinations (BE), was used as the only criteria for selecting students for higher education including university education. Board examinations are high stakes, since students' performance in these examinations determine whether they qualify for government scholarships. Class 10 and class 12 students appear for the board examinations, which is administered and evaluated externally by Bhutan Council for School Examinations and Assessment (BCSEA). BCSEA sets the question papers and selected teachers, commissioned by BCSEA, marks the students' papers (BCSEA, 2022). Prior to the conduct of BE, schools also conduct trial examinations (TE) for reasons which has not been documented thus far. TE are usually conducted at the end of October and continues till the first week of November every year (BCSEA, 2023a; 2023b). To maintain uniformity throughout the schools in Bhutan, BCSEA sets the question papers and marking schemes, while the evaluation of the paper is conducted by teachers in their respective schools (Yuden, 2023). However, a recent notification from BCSEA (2024) requires that schools are responsible for the design of the question papers and the conduct of TE. The speculations regarding the very need of TE and the efficacy of the same and Bhutan education blueprint (Ministry of Education, 2014) recommended that an independent review be conducted. This research was conducted as a part of the review. It was only in 2023 that the Bhutanese education ministry came up with an assessment framework, which mandates that schools conduct TE (Ministry of Education and Skills Development, 2023). However, there is a lack of literature on the TE or mock examinations for class 10 and class 12 in the Bhutanese educational context.

This phase of the research was preceded by semi-structured interviews (Creswell, 2013; Creswell & Poth, 2018; Merriam, 2009; Merriam & Tisdell, 2016) of eleven principals and focus-group discussions (Cohen et al., 2018; Onwuegbuzie et al., 2009) with teachers teaching class 10 and class 12. Based on the semi-structured interviews and the focus group discussions with the principals and teachers respectively, in lieu of a lack of literature on the aims and purposes of trial examinations, nine themes were identified as the aims and purposes of TE: (1) TE familiarized students with the BE question pattern, (2) diagnostic test to identify students with learning gaps, (3) prepare students mentally for the BE, (4) to reflect on teaching and learning for both teachers and students, (5) provide feedback for remedial classes after TE, (6) motivate students to learn and to take things seriously, (7) determine teachers coverage of syllabus, (8) administratively determine if teachers and students are ready for the BE, and (8) to predict how students would perform in the BE. Notwithstanding the aims and purposes of TE, counter arguments on TE were also obtained from both principals and teachers. TE were perceived to be a waste of resources since some of the principals and teachers perceived that students did not take the TE seriously and was a waste of material resources and instructional time, extra work for teachers in setting the test papers and evaluating the papers, and additional stress for teachers and students.

### *Teacher accountability*

Teachers are held accountable for student performance and the latter directly reflects on teacher appraisal system. Despite cautions about the multiple use of students' standardized assessment marks, particularly for teacher accountability and appraisals (Klenowski, 2016; Wu, 2016), the practice is common in most of the education systems (Burns & Strunk, 2021; Trinidad, 2022). Bhutanese education system is no exception. The Bhutan professional standards for teachers (Ministry of Education and Skills Development, 2019), puts forth seven standards and thirty-seven indicators for teacher appraisal, which includes student learning. According to the principals and the teachers, who participated in the qualitative phase of this research, TE was

used to determine teacher accountability, such as completion of syllabus on time and improvement of student examination scores in BE.

#### *Waste of instructional time*

The conduct of TE impinges on the instruction hours. The TE for class 10 and class 12 students are scheduled for two weeks (BCSEA, 2023a; 2023b), which comes at the cost of instructional hours. Similarly, teachers have to spend time designing the question papers and mark students' answer scripts, which is also time consuming. However, do teachers and students perceive it as a waste of instructional hours?

#### *Teacher and student stress*

Tests and examinations are stressful, and studies report students feeling anxious about tests, in what is called test anxiety, especially in a high-stake testing program. Test anxiety negatively affects students' performance (Daly et al., 2011; Roos et al., 2021; Schillinger et al., 2021; Silaj et al., 2020). Although TE are not high-stakes examinations, the results obtained during the examinations can become a source of anxiety and stress, especially if students do not perform well (Chamberlian et al., 2011). Research in other educational settings have claimed that student's participation in mock or trial examinations reduce exam related stress (Hossan & Islam, 2019).

#### *Predicting BE performance*

One of the reasons purported for the need for TE was to predict school and individual students' performance in the BE. In other words, do TE scores have predictive validity? Predictive validity refers to the degree to which a test or an assessment precisely predicts future outcomes (Wainer & Sireci, 2005). Examinations scores have been found to predict future outcomes in similar tests (Allensworth & Clark, 2020; Edmunds et al., 2020). However, there is a lack of evidence in Bhutanese assessment literature about the predictive validity of trial examinations on other assessment outcomes.

### **Purpose of the research**

There is a dearth of literature on the aims, goals, purposes, and consequences of TE, thus rendering challenges in making data-driven decisions and evidence-based practices. The themes generated from principal interviews and teacher focus-group discussions through qualitative inquiry are plagued with concerns about generalizability (Cohen et al., 2018; Creswell & Creswell, 2017; Polit & Beck, 2010; Mertens, 2015; Newman, 2014). Therefore, to generalize the findings from the qualitative interviews and focus group discussions, quantitative research was undertaken. The central purpose of the research was to quantitatively determine how teachers and students perceived the role of TE in teacher accountability, loss of instructional hours, improving academic outcomes, and test anxiety. The secondary purpose of the research was to determine how TE scores predicted student performance score in BE.

### **Research questions**

This research is guided by four research questions:

1. Do teachers and students perceive the planning and conduct of TE as a waste of instructional time?

2. How does TE contribute to higher academic achievement?
3. Do the preparation, conduct, and consequences of TE stress teachers and students?
4. How does performance in TE relate to performance in BE?

## Data collection

Two survey instruments, consisting of Likert type statements, were developed to collect data from teachers and students. 261 teachers from across the country, who were teaching or have taught class 10 and or class 12 contributed data for the research. Similarly, 562 students who were in class 11, class 12, or first-year university or diploma programs participated in the survey in 2016. These students were particularly selected since they had the experience of appearing for trial examinations in the last one or two years. Cross-sectional surveys (Cohen et al., 2018; Mertens, 2015) were conducted which enables the collection of data in one go. Besides the survey data, students' TE and BE marks were also collected from schools. The examinations scores of 2434 students in class 10 and 557 science students in class 12 were collected. According to Krejcie and Morgan (1970), a sample of 30 or more participants in a group is more than sufficient to perform robust statistical tests.

### Teacher Survey Results

#### Validity and Reliability of the Questionnaire

To determine the internal consistency of the items in a scale or the reliability of the subscales, Cronbach's alpha was calculated for each of the scales. A Cronbach's (Cronbach, 1954) alpha of 0.7 and greater shows a higher internal consistency (Nunnally, 1978; Nunnally & Berstien, 1994; Taber, 2018). In other words, the scale is reliable. Two more scales were also included in the original questionnaire which dealt with policy implications as a result of trial examinations and modality of the conduct of it. However, both these scales were not reliable, as the Cronbach's alpha were less than 0.6. Therefore, these two scales were removed from further analysis.

#### Teacher accountability towards student learning

A growing body of literature suggests that teachers and their instructional approaches are key factors for the effectiveness and improvement of schools, particularly for promoting student learning in terms of academic achievement results as a key output variable (Day et al., 2020; Papay, 2012; Scheerens et al., 2007). Further, analysis of student academic achievement results is the easiest method to assign teacher accountability, since test results are measurable and statistical in nature (Lee, 2008).

Results indicate that teachers perceive trial examinations as a necessary evil for accountability purposes. Sixty four percent of teachers were of the perception that trial examinations were necessary for accountability ( $M = 3.66$ ,  $SD = 1.157$ ) against 17%, who did not feel it necessary for accountability purposes. This trend is however, despite their perception that school administrators believe that TE results reflect the quality of teachers' classroom instruction ( $M = 3.0$ ,  $SD = 0.95$ ). Teachers welcoming this model of accountability appears to be well founded in their belief systems: first, 75% against 7% of the teachers surveyed appears of the perception that students' become better test-takers as a result of trial examinations ( $M = 4.01$ ,

$SD = 0.918$ ); second, 65% of teachers were of the conviction that trial examination improves students' ability to think critically ( $M = 3.97$ ,  $SD = 0.943$ ); third, 79% of teachers believed that trial examination motivates students' to do better in the board examination ( $M = 4.02$ ,  $SD = 0.855$ ). The descriptive results for teacher accountability scale and percentages of responses who agree and disagree are provided in Table 1.

*Table 1: Descriptive results for teacher accountability scale*

Item no	Statement ( $\alpha = 0.744$ , $N = 261$ )	Mean	Std. Dev	Agree (%)	Disagree (%)
1	Trial examination is necessary for teachers' accountability.	3.66	1.16	63.6	17.2
7	Students become better test takers as a result of trial examination.	4.01	0.92	74.7	6.9
8	Trial examinations have improved students' ability to think critically.	3.79	0.94	65.1	10
12	Teachers view trial examination as an opportunity to learn about the content that students' have not mastered.	3.96	0.77	78.5	5
19	Trial examination motivates students to perform better in the board examination.	4.02	0.86	78.5	5.7
21	Low scoring students would do better in the board examination if they receive specific preparation after the trial examination.	3.85	0.84	75.1	6.5
35	School administrators believe students' trial examination marks reflect the quality of teachers' classroom instruction.	3	0.95	32.6	30.3

Teachers appear to be capitalizing on the results of the TE: Seventy nine percent of teachers reported using TE data to learn about the contents students had not mastered ( $M = 3.96$ ,  $SD = 0.769$ ). Teachers mining data from TE results appears to be based on their belief that students' performance in the BE could improve, if they received specific preparation on weaker content areas after the TE ( $M = 3.85$ ,  $SD = 0.836$ ). Seventy five percent of teachers appear to hold this conviction.

Teachers' perceptions of accountability towards student learning were further analyzed using demographic information. An independent-samples t-test indicated that scores on potential for improvement in the BE were significantly higher for teachers teaching in Rural locations ( $M = 4.04$ ,  $SD = 0.898$ ) than for their counterparts in Semi Urban locations ( $M = 3.08$ ,  $SD = 0.710$ ),  $t(188) = 2.56$ ,  $p = 0.011$ ,  $d = 0.37$ . This indicates that teachers in Rural locations feel more accountable for student learning than teachers in Semi Urban settings.

Similarly, an independent-samples t-tests indicated that teacher accountability scores for teachers teaching Class 12 ( $M = 3.93$ ,  $SD = 1.031$ ) were significantly higher than the same for Class 10 teachers ( $M = 3.56$ ,  $SD = 1.172$ ),  $t(76) = -1.950$ ,  $p = 0.039$ ,  $d = 0.33$ . This suggests that teachers' teaching Class 12 perceive a greater accountability towards student learning as compared to those teaching Class 10.

Teachers' perceptions about student motivations because of TE differed significantly based on the number of teaching experiences: it appears that as teachers gain more and more teaching

experiences, they become more and more skeptical about TE motivating students to do better in BE. An independent-samples t-test indicated that teachers' with less than 5 years of experiences' scores were higher ( $M = 4.21$ ,  $SD = 0.689$ ) than the scores of teachers with more than 15 years of teaching experiences ( $M = 3.86$ ,  $SD = 0.854$ ),  $t(130) = 2.052$ ,  $p = 0.042$ ,  $d = 0.454$ .

Similarly, statistically significant differences were obtained between teachers with less than 5 years and 10-15 years of teaching experience. The scores for teachers with less than 5 years were significantly higher ( $M = 4.21$ ,  $SD = 0.79$ ) than the scores for teachers with 10-15 years of teaching experiences ( $M = 3.65$ ,  $SD = 1.033$ ),  $t(47) = 3.734$ ,  $p < 0.001$ ,  $d = 0.65$ . Levene's test indicated unequal variances ( $F = 16.036$ ,  $p < 0.001$ ), so the degrees of freedom were adjusted from 146 to 47. Third, the scores of teachers with 6-10 years of experiences were also significantly higher ( $M = 4.01$ ,  $SD = 0.896$ ) than the scores of teachers with 10-15 years of teaching experiences,  $t(122) = 1.97$ ,  $p < 0.001$ ,  $d = 0.37$ .

Teachers' scores on students' improvement in critical thinking skills as a result of TE differed significantly as a result of increased years of teaching experiences. Independent-samples t-test indicated that teachers' with less than 5 years of experiences scores were significantly higher ( $M = 3.89$ ,  $SD = 0.947$ ) as compared to teachers with 10-15 years ( $M = 3.46$ ,  $SD = 1.016$ ),  $t(146) = 2.361$ ,  $p = 0.020$ ,  $d = 0.49$ . This indicates that as teachers gain more and more teaching experience, they become more critical of their own pre-dispositions.

Teachers' perceptions of TE result as an opportunity to learn about the content student have not mastered appears to decrease in momentum as they gain more and more teaching experience. Independent-sample t-tests indicated that the scores for teachers with less than 5 years ( $M = 4$ ,  $SD = 0.726$ ) were significantly higher than for teachers with 10-15 years of experience ( $M = 3.57$ ,  $SD = 0.959$ ),  $t(146) = 2.884$ ,  $p = 0.005$ ,  $d = 0.51$ . Similarly, teachers with 6-10 years of experience scores were significantly higher ( $M = 4.10$ ,  $SD = 0.683$ ) than that of teachers with 10-15 years of experience ( $M = 3.57$ ,  $SD = 0.959$ ),  $t(122) = 3.526$ ,  $p = 0.01$ ,  $d = 0.54$ . This seems to suggest redundancy, that as teachers gain more and more experience, they view this practice as routine and not something that has the potential to inform teaching and learning processes.

### Use of TE Results

Beliefs are said to be strong predictors of teachers' teaching behavior (Nespor, 1986; Pajares, 1992; Richardson, 1996). In general, four out of 5 teachers in the survey believed that students' performance in BE could be improved if they received "specific" preparation after the TE. The use of TE results provides a glimpse of the process of "specific" preparation referred to above. As indicated in the earlier section, almost 8 out of 10 teachers view the analysis of TE results to be a good and reliable source of information for the contents' students have not mastered. The knowledge of this gap in student achievement appears to be attempted to be filled using remedial measures. 79% of teachers reported using TE results to determine remedial measures. On the other hand, TE results also appear to be used to predict student performance in BE. The mean and standard deviations for this scale are provided in Table 2.

*Table 2: Mean and standard deviation for use of results scale*

Item no	Statements ( $\alpha = 0.611$ )	Mean	Std. Dev	Agree (%)	Disagree (%)
2	The results of trial examination is used for determining remedial measures.	4.02	1.004	78.9	10.7

11	Trial examination results are used to predict students' performance in the board examination.	3.95	0.835	76.6	5.7
12	Teachers view trial examination as an opportunity to learn about the content that students' have not mastered.	3.96	0.769	78.5	5
21	Low scoring students would do better in the board examination if they receive specific preparation after the trial examination.	3.85	0.836	75.1	6.5

One-way MANOVA was conducted to determine whether there was any statistically significant difference in the mean scores of items in the Use of result scale and demographic characteristics of the teachers. A statistically significant difference was obtained for items in the scale based on teachers' teaching experience,  $f(8,510) = 2.220$ ,  $p = 0.025$ ; Wilks'  $\Lambda = 0.934$ , partial  $\eta^2 = 0.034$ .

Post-hoc analysis showed that only the mean scores of TE results being used to determine remedial measures were significantly different between teachers with 6-10 years and 10-15 years of teaching experience ( $p < 0.05$ ). The mean scores for teachers with 6-10 years of experience  $M = 4.18$ ,  $SD = 0.983$  were significantly higher than that of teachers with 10-15 years of experience ( $M = 3.59$ ,  $SD = 1.279$ ),  $F(4, 256) = 2.629$ ,  $p = 0.035$ , partial  $\eta^2 = 0.039$ . This suggests that as teachers gain more and more experience, they become less enthusiastic about using TE results for determining remedial measures.

### Impact on Instructional hours

Education Policy documents such as Education Policy and Guidelines and Instruction ([EPGI], MOE, 2012) and strategic documents such as Bhutan Education Blueprint 2014-2024, Rethinking Education state "the school curriculum is to be delivered in a minimum of 180 instructional days in an academic year" (MOE, 2014, p. 32). The Department of Curriculum Research and Development (2012) suggests that extra classes are conducted especially for classes X and XII, which is because the syllabus is heavy and cannot be delivered in the stipulated instructional days. Also, one of the primary causes of the need for this study was because school administrators and policy makers were of the perception that the conduct of trial examinations impeded instructional days.

Table 3 summarizes the descriptive results for the impact on instructional hours due to conduct of trial examinations. The composite mean for this scale was obtained at 3.22, with a standard deviation of 1.093, suggesting that the scores among the participants and items largely differed.

*Table 3: Summary of descriptive results for impact on instructional hours scale*

Item	Statement ( $\alpha = 0.755$ , $N = 261$ )	Mean	Std. Dev	Agree (%)	Disagree (%)
14	The conduct of trial examination has negative effects on instructional hours.	2.82	1.143	42.1	29.1

15	After the trial examination, there is sufficient time to provide remedial classes to address students' needs.	3.26	1.161	30.7	53.3
23	Trial examination disturbs the instructional hours of other classes.	3.11	1.224	36.8	44.8
43	Invigilation duty for trial examination impede on instructional hours.	3.49	0.955	17.2	55.9
48	Setting trial examination question paper hinders instructional time.	3.18	1.055	33.3	48.7
49	Conducting trial examination reduces the number of instructional days in school	3.36	1.056	24.9	55.6
50	Marking trial examination papers hinders instructional time	3.29	1.062	29.1	52.5
51	Preparing consolidated result sheet for trial examination takes away instructional time	3.24	1.087	31.8	50.2

Contrary to policy makers and school administrators' views, teachers appear to believe that the conduct of TE does not adversely affect their instructional hours ( $M = 2.82$ ,  $SD = 1.143$ ). Forty-two percent of teachers surveyed were of the perception that it does not have negative impact on instructional hours, compared to 29.1% of participants who perceived it as a being detrimental to instructional hours. One-way MANOVA revealed that there was no significant difference between the scores based on teachers' demographic information. This suggests that teachers, irrespective of classes taught, location, subjects taught, teaching experience, and gender unanimously perceive that TE does not have negative impact on instructional hours.

Teachers reported that they get sufficient time to provide remedial classes to address students' specific needs after the TE ( $M = 3.26$ ,  $SD = 1.161$ ). Fifty three percent of teachers were of this perception against 31% who reported that the time available was insufficient. Document analysis, especially school calendars and student diaries, revealed that TE were usually conducted towards the third week of October. However, some schools, irrespective of being a Higher Secondary or a Middle Secondary conducted TE either for one week or two weeks.

In general, the conduct of trial examinations: setting question papers, conduct and invigilation duties, marking student papers, and preparing consolidated results, were perceived not to significantly impact instructional hours (means ranged from 3.11 to 3.49, and standard deviations from 0.955 to 1.224). It is important to make a crucial distinction here: the conduct of TE has obvious and visible impact on instructional hours; however, teachers' perceptions of these impediments do not appear to be negative. From Table 3, it can be surmised that the percentages of the participants who agreed to the statements did not differ much from the combined percentage of neutral and disagree. This suggests, once again, that teachers perceive TE as a necessary evil.

#### Teacher stress because of TE

Literature suggests that examinations or high stakes testing programs are generally very stressful for teachers (Abrams et al., 2003; Ballard & Bates, 2008). Teacher stress in the context of TE is used to refer to additional job responsibilities and accountability to raise students' marks. This begets a question: Do teachers stress out as a consequence of TE?

Contrary to established literature, teachers in Bhutan reported that they do not feel pressured to raise students' marks in or after the TE, either from school principals ( $M= 2.84, SD=1.047$ ) or parents ( $M= 2.63, SD= 0.954$ ). Similarly, they also reported that setting ( $M= 2.85, SD= 1.120$ ) and marking ( $M= 2.95, SD= 1.141$ ) TE papers were not undue burden for them. This corroborates with earlier findings about teacher accountability: teachers appear to be intrinsically motivated to improve student performance, and the sense of accountability is not imposed by external factors. However, unlike other countries, teachers in Bhutan are neither incentivized for students' good performance, nor sanctions are imposed if students' performance is under the mark. Although teachers feel accountable for students' performance, the absence of incentives or sanctions appears to be one of the reasons why teachers reported that the conduct of TE is not stressful. The descriptive results and the percentages of responses are shown in Table 4.

Thirty eight percent of teachers reported that they were not pressured to raise students' scores in TE by the school management, against 29 %. Similarly, forty four percent of teachers reported that parents did not pressurize them to raise students' score after the TE.

*Table 4: Summary of descriptive results for teacher stress scale*

Item	Statements ( $N= 261, \alpha = 0.70$ )	Mean	Std. Dev	Agree (%)	Disagree (%)
20	Teachers feel pressure from the school management to raise students' marks in trial examination.	2.84	1.05	28.7	38.3
26	Setting trial examination paper(s) is an undue burden for teachers.	2.85	1.12	33	44.8
27	Marking trial examination paper(s) is an undue pressure for teachers.	2.95	1.14	8.4	41.4
31	Teachers feel pressure from parents to raise students' marks after the trial examination.	2.63	0.95	17.6	44.1

Also, contrary to popular beliefs that setting and marking TE papers as being stressful for teachers, it appears that teachers themselves do not hold these perceptions.

To determine if there were any statistically significant differences between the samples, multiple one-way MANOVAs were conducted using school location, classes taught, teaching experience, and school types as dependent variables. All these MANOVAs revealed insignificant differences. This implies that teachers irrespective of the above variables generally perceived the conduct of TE as not stressful.

## Student results

### Learning

Results indicated that students were also of the perception that TE contributed towards their learning ( $M= 4.34, SD= 0.80$ ). Ninety four percent of students surveyed against 1% believe that TE enhanced their learning. Similarly, ninety one percent were of the notion that they became better test takers as a result of TE, against 2%. One-way MANOVAs indicated that there were significant differences between the samples based on their demographic characteristics. First, the scores of students in Semi-Rural schools ( $M= 4.60, SD= 0.610, N = 124$ ) were significantly higher than the scores of students or Urban ( $M= 4.42, SD= 0.725, N = 328$ ) and Semi Urban

Schools ( $M= 4.25$ ,  $SD= 0.859$ ,  $N = 61$ ),  $t(2, 510) = 4.557$ ,  $p=.011$ ,  $\eta^2p = .018$ . Second, the scores of Class 12 students on this item were significantly higher ( $M = 4.54$ ,  $SD = 0.627$ ,  $N = 252$ ) than that of Class 11 ( $M= 4.36$ ,  $SD = 0.794$ ,  $N = 261$ ) and first year degree or diploma students ( $M= 4.10$ ,  $SD= 1.00$ ,  $N = 49$ ),  $t(2,559) = 8.868$ ,  $p<.0005$ ,  $\eta^2p = .031$ . Third, science ( $M = 4.47$ ,  $SD= 0.769$ ,  $N= 218$ ) and commerce students ( $M= 4.47$ ,  $SD= 0.657$ ,  $N= 188$ ) perceived that they became better test takers as a result of TE more than their counterparts in Arts ( $M= 4.29$ ,  $SD= 0.835$ ,  $N = 156$ ),  $t(2, 559) = 3.201$ ,  $p =.041$ ,  $\eta^2p = .011$ ). In all the three cases partial eta squared were low.

*Table 5: Descriptive results of items in the Student Learning scale*

Item no	Statement ( $N= 562$ , $\alpha = 0.70$ )	Mean	Std. Dev	Agree (%)	Disagree (%)
1	Trial examination enhanced my learning.	4.53	0.69	94.13	1.42
2	Trial examination helped me to become a better test taker.	4.42	0.76	90.75	1.96
3	Trial examination improved my ability to think critically	4.24	0.82	83.81	2.85
4	Trial examination was an opportunity to learn about the content that I have not mastered.	4.17	0.92	79.54	5.87
<b>Composite</b>		<b>4.34</b>	<b>0.8</b>		

Eighty four percent of students surveyed held the perception that TE improved their critical thinking skills. Also, eighty percent of the students reported that TE results indicated the content that they had not mastered. The scores on this item also differed significantly among the samples based on demographic information. First, the scores of students in Semi-Rural schools ( $M= 4.37$ ,  $SD= 0.83$ ) were significantly higher than the scores of students in Semi-Urban Schools ( $M= 3.97$ ,  $SD= 1.032$ ),  $t(2, 510) = 4.479$ ,  $p= .012$ ,  $\eta^2p = .017$ . This corroborates with earlier findings that teachers in Semi-Rural Schools feel a greater sense of accountability towards student learning than their counterparts in Semi-Urban Schools.

Second, the scores of class 12 students ( $M= 4.31$ ,  $SD= 0.847$ ) were significantly higher than the scores of class 11 ( $M= 4.08$ ,  $SD= 0.947$ ) and first year degree/ diploma students ( $M= 3.94$ ,  $SD= 1.07$ ),  $t(2, 559) = .003$ ,  $\eta^2p= .020$ . Third, the scores of students in Commerce ( $M= 4.26$ ,  $SD = 0.854$ ) were significantly higher than the scores of students in Arts ( $M= 4.01$ ,  $SD= 0.957$ ),  $t(2, 559) = 3.368$ ,  $p= .035$ ,  $\eta^2p = .012$ . This indicates that students in class XII commerce perceive a greater opportunity to learn about their own learning compared to Science and Arts Students.

### Impact on Instructional Hours

Results from student data indicated that they did not perceive TE to have negative consequences on instructional hours. This finding corroborates with teachers' perceptions. Seventy two percent of the students surveyed did not have the perception that the conduct of TE has negative impact on instructional hours. Similarly, fifty eight percent of the students reported that TE did not disturb the instructional hours of other classes. However, students were of the perception that

teachers used instructional hours to set TE question papers (40%), mark and evaluate TE papers (37%), and prepare TE results (36%). Considering the higher standard deviation of the above three items above, it appears that the variations in students' scores were large. A substantial percentage of students (61%) were of the perception that the conduct of TE compelled teachers to rush through the syllabus. Table 6 summarizes the descriptive results obtained.

One-way MANOVA of the items on the impact on instructional hours and disturbance to other classes were found to be statistically significant based on location of the school.

*Table 6: Descriptive results of the impact on instructional hours scale*

<i>Item no</i>	<i>Statements (N = 562, <math>\alpha = 0.76</math>)</i>	<i>Mean</i>	<i>Std. Dev</i>	<i>Agree (%)</i>	<i>Disagree (%)</i>
27	The conduct of trial examination had negative effects on instructional hours.	2.16	0.93	9.25	71.17
29	Trial examination disturbed the instructional hours of other classes.	2.43	1.07	17.97	58.36
30	Teacher used instructional time to set trial examination questions.	3.17	1.04	39.68	24.91
31	Conduct of trial examination compelled teachers to rush through the syllabus.	3.66	1.09	60.68	16.37
32	Teacher used instructional hours to mark trial examination questions.	3.14	0.97	36.83	24.2
33	Teacher used instructional hours to prepare trial examination's result.	3.07	1	36.3	30.25
	Composite	2.94	1.02		

### Perceptions of Stress

Results indicated that generally students did not feel stressed due to TE ( $M = 2.98$ ,  $SD = 1.01$ ). Table 7 summarizes the descriptive results of this scale. Although eighty one percent of the students reported that they spent considerable time preparing for TE, only thirty nine percent felt pressured to perform well in the TE. About forty percent of the students reported that teachers emphasized performing well, and thirty three percent reported having been pressured to perform well in TE. Forty percent of the students reported being extremely anxious about taking the TE. However, on the contrary, 69% of the students were of the perception that preparing for and taking TE were not undue burden for them.

*Table 7: Descriptive results of Stress due to TE*

<i>Item no</i>	<i>Statements (N= 562, <math>\alpha = 0.73</math>)</i>	<i>Mean</i>	<i>Std. Dev</i>	<i>Agree (%)</i>	<i>Disagree (%)</i>
38	Students felt pressure from the teachers to perform well in trial examination	3.09	1.08	38.61	29.36
39	Students were extremely anxious on taking the trial examination.	3.22	0.91	39.5	18.68
40	Preparing for trial examination was an undue burden for students.	2.3	1.06	13	68.5

41	Writing trial examination was an undue burden for students.	2.25	1.04	12.99	68.51
42	Students felt pressure from parents to perform well in trial examination.	2.95	1.1	32.56	35.41
42	Students were under intense pressure to perform well in the trial examination.	3.12	1.09	39.32	28.83
20	Students spent a considerable amount of time preparing for trial examination.	3.95	0.78	81.32	5.52
<b>Composite</b>		<b>2.98</b>	<b>1.01</b>		

One-way MANOVA of the scale did not reveal any statistically significant difference between the samples. This indicates that despite the differences in demographic characteristics, the results do not differ by any significant means.

### Trial and Board Examination Marks

TE and BE marks of individual students were gathered from selected schools across the country, to empirically ascertain school administrators', teachers', and students' claim that TE results bear a relationship with students' performance in BE. Similarly, to ascertain the general perceptions and practice that TE results were used to predict students' performance, bivariate correlation analysis and linear regression analysis were conducted.

### Class 10 results

Pearson correlation coefficient ( $r$ ) measures the strength and direction of the association between two entities. The marks in all the subjects showed moderate positive correlation (0.4 - 0.6) except for History, Civics, and Geography and Optional Subjects (Economics and IT) which showed strong positive correlation coefficients. Also, all the coefficients were significant at  $p < 0.01$ . Table 8 shows a summary of the Pearson correlation coefficients of Trial and Board Examination marks of all the subjects.

*Table 8: Pearson correlation coefficients of trial and board examination marks.*

		<i>Board English</i>	<i>Board Dzongkha</i>	<i>Board Mathematics</i>	<i>Board His, Civ, Geo</i>	<i>Board Science</i>	<i>Board Optional</i>
<i>Trial English</i>	$r$	0.643					
	$N$	2434					
<i>Trial Dzongkha</i>	$r$		0.414				
	$N$		2430				
<i>Trial Mathematics</i>	$r$			0.674			
	$N$			2433			
<i>Trial His, Civ, Geo</i>	$r$				0.724		
	$N$				2433		
<i>Trial Science</i>	$r$					0.675	

	N	2432	
<i>Trial Optional</i>	r		0.724
	N		2396

*Remarks: All correlation coefficients were significant at 0.01 level (2-tailed). N represents the number of participants.*

Linear regression analysis was conducted on TE marks and BE marks to determine the predictive value of TE marks. The BE marks were used as dependent variables and TE marks as independent or predictor variables. A summary of results obtained is provided in Table 9.

Significant regression equations were obtained for all the subjects as follows:

A significant regression equation was obtained for TE marks and BE marks for English at  $f(1, 2432) = 1716.41, p < .001$ , with an  $R^2 = .414$ . Students BE marks increased by 0.923 for each increase in TE marks.

*Table 9 : Summary of Regression coefficients*

<i>Dependent variables</i>	$\beta$	<i>df</i>	<i>F</i>	<i>Sig.</i>	<i>R square</i>
Board English	0.923	2432	1716.414	0.001	0.414
Board Dzongkha	0.574	2428	501.545	0.001	0.171
Board Mathematics	0.738	2431	2021.366	0.001	0.454
Board His, Civ, Geo	0.923	2431	2674.374	0.001	0.524
Board Science	0.841	2430	2037.384	0.001	0.456
Board Optional	0.785	2394	2631.938	0.001	0.524

*Remarks: Predictor variables are TE marks*

A significant regression equation was also obtained for Dzongkha marks at  $f(1, 2428) = 501.545, p < .001$ , with an  $R^2 = .171$ . Students' BE Dzongkha marks increased by .574 for every increase in TE marks. A significant regression equation was also obtained for Mathematics marks at  $f(1, 2431) = 2021.366, p < .001$ , with an  $R^2 = .454$ . Students' Mathematics BE marks increased by .738 for every rise in their Mathematics TE marks. Also, a significant regression equation was obtained for History, Civics, and Geography.  $f(1, 2431) = 2674.374, p < .001$ , with an  $R^2 = .524$ . Students' BE marks increased by 0.923 for every increase in TE marks. For Science (Physics, Chemistry, and Biology) a significant equation was obtained at  $f(1, 2430) = 2037.384, p < .001$ , with an  $R^2 = .456$ . Students' BE marks in science increased by 0.841 for every increase in their TE marks. A significant equation was also obtained for Optional Subjects (Economics and Information Technology) at  $f(1, 2394) = 2631.938, p < .001$ , with an  $R^2 = .524$ . In other words, students' BE marks in these subjects increased by 0.785 for every increase in their TE marks.

## Class 12 Science

Results for class 12 science students indicated that both correlation coefficients and regression equations were significant at  $p < .001$  level.

TE and BE marks in English showed moderate correlation coefficient. This means that if TE marks increased, BE marks would also increase. Regression statistics showed that for every increase in TE English marks, BE marks would increase by 0.471 and that TE marks explained 45.5% of the total variance in BE marks. Table 10 and 11 show the summary of correlation coefficients and regression statistics respectively.

*Table 10: Correlation coefficients for BE and TE science subjects.*

		Eng_B	Dzo_B	Math_B	Phy_B	Che_B	Bio_B
Eng_T	r	0.675					
	N	557					
Dzo_T	r		0.505				
	N		555				
Math_T	r			0.745			
	N			335			
Phy_T	r				0.629		
	N				557		
Che_T	r					0.581	
	N					556	
Bio_T	r						0.545
	N						266

All correlation coefficients were significant at  $p < .001$  (2-tailed)

*Table 11: Summary of Regression statistics for Science*

	B	df	F	R Square	Sig
Eng_T	0.471	555	464.169	0.455	0.001
Dzo_T	0.357	553	189.768	0.255	0.001
Math_T	0.836	333	415.735	0.555	0.001
Phy_T	0.579	555	362.581	0.395	0.001
Che_T	0.435	554	282.469	0.338	0.001
Bio_T	0.42	264	111.718	0.297	0.001

*Independent variable: Board Examination marks, and Predictor variable: Trial Examination Marks*

A moderate positive correlation coefficient was obtained for TE and BE Dzongkha marks (0.505). TE marks predicted that for every increase in TE Dzongkha marks, BE marks would increase by 0.357. Also, with an  $R^2$  value of .255, TE marks explained 25.5% of the total variance. A strong positive correlation coefficient was obtained between TE and BE Mathematics marks (0.745). Beta value indicates that for every increase in TE Mathematics marks, BE mark would increase by 0.836. The regression equation explained 55.5% of the total variance. TE and BE marks for Physics were moderately correlated (positive, 0.629). Beta value indicated that for every increase in TE marks, BE mark increases by 0.579. R square value indicated that 39.5% of total variance was explained. Chemistry marks also showed moderate positive correlation. TE marks predicted that for every increase in chemistry marks in TE, BE marks increase by 0.435 and that 33.8% of the total variance were explained. Biology marks also

showed moderate positive correlation. BE marks increased by 0.42 for every increase in biology TE marks. 29.7% of the variance was explained.

## Discussions

### *Trial examinations are a necessary evil*

Class 12 teachers appear to hold greater accountability towards student learning compared to class 10 teachers. Perhaps this may be because at the time of this research, the majority of subjects taught in class 12 did not have continuous assessment marks which were added to the summative examination marks (BCSEA, 2022). Results indicate that while teachers are not under pressure to raise student scores from the school administration or the parents, the conduct of TE compels them to complete the coverage of prescribed syllabus. Teachers use students' TE performance to determine and conduct remedial classes to bridge the gaps in student learning, to improve student academic achievement during examinations. Although the school administration equates students' performance in TE to the quality of classroom instructions, the low-stake nature of TE and the diagnostic assessment data obtained provide impetus for interventions before the final, high-stakes BE.

### *Instructional hours are not wasted*

Despite the TE taking between a week and two weeks of instructional time to conduct and other compromises teachers have to make for preparation and evaluation of answer scripts, both teachers and students perceive that instructional hours are not wasted. Perhaps this is because TE are scheduled activities, which does not impinge on the 180 hours of classroom instructions (MoE, 2014). Although data suggests that some teachers use instructional hours to design question papers and evaluate student's answer script, this may be occurring since group moderation of test papers and group evaluations of answer scripts have become a norm in Bhutanese schools.

### *Contribution towards student achievement*

Students' performance in TE provides assessment feedback about the contents that the students have to learn before the final examinations, in order to better prepare for the inevitable day. Both teachers and students become aware of the gaps in learning (Bangert-Drowns et al., 1991), and there is a little more than a month to provide remedial classes and self-study before the commencement of BE. In addition to determining the learning gaps, students also become cognizant of the test paper formats (Lockshin & Zamkov, 2009). Phelps (2012) and Bangert-Drowns et al. (1991) purports that increasing the frequency of testing positively impacts student achievement. The conduct of practice tests has greater impact on student learning than restudying the concepts (Adesope et al., 2017). Furthermore, results indicate that the notion that TE may be inducing test anxiety in students is a myth.

### *TE performance predicts student's performance in the BE*

Academic achievement results from more than 3000 students indicate that there is a significant positive correlation between TE scores and BE scores. Regression results demonstrate that the relationship between TE and BE scores are statistically significant. Therefore, TE scores can and are being used to predict student achievement in BE.

## Conclusion

The research was conducted to quantitatively validate and generalize the findings from a qualitative inquiry regarding TE. The findings of this research contribute to Bhutanese assessment literature, regarding the assumed impacts on teacher accountability, teacher and student stress, student learning, and the predictive validity of TE on student achievement in BE. Results suggest that TE are necessary for several reasons, including facilitating the collection of feedback on student learning and making provisions for addressing learning gaps and improving student achievements. The results debunk the myths that the planning and conduct of TE are a waste of instructional hours and lead to stress among teachers as well as students.

Disclaimer (Artificial intelligence)

Option 1:

The author hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

## References

- Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: teachers' opinions of statewide testing programs. *Theory into Practice*, 42(1), 18-29.
- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659-701.  
<https://doi.org/10.3102/0034654316689306>
- Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher*, 49(3), 198-211.
- Ballard, K., & Bates, A. (2008). Making a connection between student achievement, teacher accountability, and quality classroom instruction. *The Qualitative Report*, 13(4): 560-580.
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research*, 85(2), 89-99.  
<https://doi.org/10.1080/00220671.1991.10702818>
- Bhutan Council for School Examinations and Assessment (2022). *Rules and regulations for the conduct of public examinations in Bhutan*. Ministry of Education and Skills Development. Royal Government of Bhutan.
- Bhutan Council for School Examinations and Assessment (2023a). *Trial examinations: BCSE (Class X)*. Ministry of Education and Skills Development. Royal Government of Bhutan.
- Bhutan Council for School Examinations and Assessment (2023b). *Trial examinations: BHSCE (Class XII)*. Ministry of Education and Skills Development. Royal Government of Bhutan.
- Bhutan Council for School Examinations and Assessment (2024). *Trial examinations question setting by schools*. Ministry of Education and Skills Development. Royal Government of Bhutan.
- Burns, J., & Strunk, K. O. (2021). School accountability. In B. P. McCall (eds) *The Routledge Handbook of Economics in Education* (pp. 107-126). Routledge.
- Chamberlain, S., Daly, A. L., & Spalding, V. (2011). The fear factor: students' experiences of test anxiety when taking A-level examinations. *Pastoral Care in Education*, 29(3), 193-205. doi:10.1080/02643944.2011.599856
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8<sup>th</sup> ed.). Routledge
- Creswell, J. W. (2013). *Qualitative inquiry and research design: choosing among five approaches* (3<sup>rd</sup> ed.). Sage.
- Creswell, J.W., & Creswell, J.D. (2017). *Research design: qualitative, quantitative, and mixed methods approaches* (4<sup>th</sup> ed.). Sage.

- Creswell, J.W. & Poth, C.N. (2018). *Qualitative inquiry and research design choosing among five approaches* (4<sup>th</sup> ed). Sage.
- Cronbach, L.J. (1954). *Educational psychology*. Brace.
- Daly, A. L., Chamberlain, S., & Spalding, V. (2011). Test anxiety, heart rate and performance in A-level French speaking mock exams: an exploratory study. *Educational Research*, 53(3), 321–330. <https://doi.org/10.1080/00131881.2011.598660>
- Day, C., Sammons, P, & Gorgen, K. (2020). *Successful school leadership*. Education Development Trust.
- Edmunds, J. A., Unlu, F., Furey, J., Glennie, E., & Arshavsky, N. (2020). What happens when you combine high school and college? The impact of the early college model on postsecondary performance and completion. *Educational Evaluation and Policy Analysis*, 42(2), 257-278.
- Hossan, MR, & Islam, N. (2019). Improving Students' Learning and Performance Using Mock Tests in Engineering Classes." *Proceedings of the ASME 2019 International Mechanical Engineering Congress and Exposition*. Volume 5: Engineering Education. Salt Lake City, Utah, USA. November 11–14, 2019. V005T07A007. ASME. <https://doi.org/10.1115/IMECE2019-11372>
- Klenowski, V. (2016). Questioning the validity of the multiple uses of NAPLAN data. In B. Lingard., G. Thompson, & S. Seller (eds). *National testing in schools* (pp. 44-57). Routledge
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30(3), 607–610. <https://doi.org/10.1177/001316447003000308>
- Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence from cross-state casual-comparative and correlational studies. *Review of Educational Research*, 78. 608-644. DOI:10.3102/0034654308324427
- Lockshin, J. & Zamkov, O. (2009). Predictors of academic achievement and their possible applications. *Policy Futures in Education*, 7(4). 399-409. <https://doi.org/10.2304/pfie.2009.7.4.399>
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. Jossey-Bass.
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: a guide to design and implementation* (4<sup>th</sup> ed.). Jossey Bass.
- Mertens, D. M. (2015). *Research and evaluation in education and Psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (4<sup>th</sup> ed). Sage.
- Ministry of Education (2012). Education policy guidelines and instructions. Royal Government of Bhutan.
- Ministry of Education (2014). *Bhutan education blueprint: 2014-2024*. Royal Government of Bhutan.

- Ministry of Education and Skills Development (2019). *Bhutan professional standards for teachers*. Royal Government of Bhutan.
- Ministry of Education and Skills Development (2023). *National student learning assessment framework: Assessment, evaluation, and reporting*. Royal Government of Bhutan.
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19, 317-328.
- Newman, L. W. (2014). *Social research methods: Qualitative and quantitative approaches* (7<sup>th</sup> ed.). Pearson
- Nunnally, J. C. (1978). *Psychometric theory*. McGraw Hill.
- Nunnally, J. D. & Bernstein, I. H. (1994). *Psychometric theory*. McGraw Hill.
- Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, A. G. (2009). A qualitative framework for collecting and analyzing data in focus group research. *International Journal of Qualitative Methods*, 8(3), 1-21. <https://doi.org/10.1177/160940690900800301>
- Pajares, F. M. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332.
- Papay, J. P. (2012). Refocusing the debate: assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123-141.
- Phelps, R. P. (2012). The effect of testing on student achievement, 1910–2010. *International Journal of Testing*, 12(1), 21–43. doi:10.1080/15305058.2011.602920
- Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: myths and strategies. *International Journal of Nursing Studies*, 47(11), 1451–1458. <https://doi.org/10.1016/j.ijnurstu.2010.06.004>
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Eds) *Handbook of Research on Teacher Education* (pp. 102-119). MacMillan.
- Roos, A-L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2021). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review*, 33, 579–618 (2021). <https://doi.org/10.1007/s10648-020-09543-z>
- Royal Government of Bhutan (2008). *The constitution of the kingdom of Bhutan*. Royal Government of Bhutan.
- Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). Review and meta-analysis of schools and teaching effectiveness. Enschede: Department of Educational Organization and Management, University of Twente.
- Schillinger, F.L., Mosbacher, J.A., Brunner, C., Vogel, S. E., & Grabner, R. H. (2021). Revisiting the role of worries in explaining the link between test anxiety and test performance. *Educational Psychology Review*, 33, 1887–1906 (2021). <https://doi.org/10.1007/s10648-021-09601-0>

- Silaj, K. M., Schwartz, S. T., Siegel, A. L., & Castel, A. D. (2021). Test Anxiety and Metacognitive Performance in the Classroom. *Educational Psychology Review*, <https://doi.org/10.1007/s10648-021-09598-6>
- Taber, K.S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48, 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Trinidad, J. E. (2022). Meaning-making, negotiation, and change in school accountability, or what sociology can offer policy studies. *Sociological Inquiry*, 93 (1). 153–178.
- Wainer, H. & Stephen G. Sireci, S. G. (2005). Item and Test Bias. In K. Kempf-Leonard (eds). *Encyclopedia of Social Measurement*, (pp. 365-371). Elsevier. <https://doi.org/10.1016/B0-12-369398-5/00446-1>.
- Wu, M. (2016). What national testing data can tell us. In B. Lingard., G. Thompson, & S. Seller (eds). *National testing in schools* (pp. 18-29). Routledge
- Yuden, S. (Aug 18, 2023). *BCSEA to set trial examination question papers for classes X and XII*. Bhutan Broadcasting Service. <https://www.bbs.bt/news/?p=190651>