RABSEL: the CERD Educational Journal

ISSN 2957-8019(Online) | ISSN 2077-4966(Print)|

24(1)88-106

Journal homepage: Journal.pce.edu.bt

Analyzing Test Data: A Mathematics Test Case Study

Author: Gembo Tshering (PhD)[1]* & Kezang Choden [2]

Paro College of Education, Royal University of Bhutan

http//doi.org/10.17102/rabsel.24.1.97| Accepted October 2023 | Published December 2023

Abstract

Most countries in the Global South conduct school examinations at national levels, generating massive data related to student performances in a range of school subjects. However, the data are often used in summative pass-fail decisions and seldom in formative decisions. One possibility is that countries must learn how to analyze the assessment data for formative decisions. Using Math test data collected from 1500 students across 60 schools in Bhutan as part of a Ph.D. study, this paper shows ways to use test data for formative decisions. The study results show that test data can profile students' knowledge and skills, classify performance into a hierarchy of thinking skills, determine the relevance of knowledge and skills to societal and technological change, and benchmark the knowledge and skills with international standards.

Key Words: Math test, knowledge profile, thinking skills, international benchmarks

Corresponding author: gembotshering.pce@rub.edu.bt

Purpose of Testing

As Tshering (2012) explored, testing encompasses both advantages and drawbacks. Nevertheless, the primary focus of this paper is to introduce an approach empowering schoolteachers to harness the positive aspects of testing, recognizing its enduring presence. Building on the assertions of Roediger et al. (2011), testing effectively exposes knowledge gaps, stimulates heightened learning in subsequent educational encounters, improves knowledge organization, facilitates knowledge transfer to novel scenarios, aids in the retrieval of non-tested information, refines metacognitive monitoring, guards against interference from previously learned material when tackling new subjects, and provides valuable feedback to instructors. Moreover, they posit that regular testing fosters student study engagement and contributes to the retrieval effect, ultimately supporting long-term retention. However, how schoolteachers can tap into these benefits remains to be determined. This paper seeks to present an approach for teachers to derive similar advantages, as asserted by Roediger et al. (2011).

Utilizing a mathematics test as a case object, this paper will illustrate strategies to (a) profile students' understanding of the school mathematics curriculum, (b) gather insights into students' readiness to confront future challenges, (c) assess students' knowledge and skills conducive to adapting to rapid societal changes; and (d) establish international benchmarks for students' proficiency in mathematics.

Mathematics Test Design and Development

Downing's (2006) comprehensive framework distills various test development procedures into 12 distinct steps: overall planning, content definition, test specification, item development, test design and assembly, test production, test administration, scoring test responses, establishing passing scores, reporting test results, item banking, and compiling a test technical report. Adhering to these 12 steps, a two-hour mathematics test for Grade 10 Bhutanese students was meticulously crafted, incorporating all 42 PISA mathematics items released by OECD (2009) for public use, as detailed in Tshering (2012). Notably, these 42 items had been employed by the OECD in either PISA 2000 or PISA 2003.

Crucially, the mathematics test was strategically aligned with the Grade 10 Bhutanese mathematics curriculum, ensuring the relevance of the test data despite its origins in the PISA item pools. This intentional mapping enables teachers to leverage the mathematics test data for targeted instructional interventions and facilitates the establishment of international benchmarks, as highlighted in the works of Tshering (2012) and Tsheing and Prain (2011).

Developing a Mathematics Proficiency Scale

Examining students' profiles in mathematical knowledge and skills can be effectively undertaken by analyzing their scores within the framework of proficiency levels. Leveraging item response theory modeling of students' response data, as advocated by Embretson and Reise (2000), allows for the independent estimation of student ability and item difficulty parameters. Crucially, these parameters can be situated on a unified measurement scale, enabling the mapping of item difficulty about student ability. As Embretson and Reise (2000) define, student ability encapsulates the knowledge and skills required to solve a test item correctly. This conceptualization articulates the mathematical knowledge and skills inherent in individual test items, facilitating the interpretation of student ability vis-à-vis the specific mathematical competencies embedded in the test items, as elucidated by Tshering (2016). This analytical approach provides valuable insights into the nuanced profile of students' mathematical knowledge and skills, offering significant potential to enhance teaching efficacy, promote meaningful learning for students, and empower school leaders to make well-informed decisions.

To illustrate, Table 1 presents a modified item map derived from PISA (OECD, 2005, p. 257), focusing on item 2a of the mathematics test. OECD (2005, 2009) has extensively documented the process of establishing performance levels for PISA test items, including those incorporated into its Mathematical Literacy test. Tshering (2016) offers a simplified presentation of the methodology for mapping item difficulty to the requisite task demanded by the item.

Table 1

Item Map for Item 2a Adapted from the OECD

Item ID	Item Difficulty	Comments-item demands
2a	611	Interpret and link picture, text and algebra; algebraic substitution; solve basic equation; single step; correct manipulation of expressions containing symbols

PISA employs a comprehensive framework of six proficiency levels from 1 to 6. Each level is accompanied by detailed performance descriptions outlining the specific types of mathematical knowledge and skills anticipated from students achieving that particular proficiency. The assignment of performance scores to each proficiency level is proportional to the complexity of the items, offering a nuanced measure of students' mathematical capabilities. Students are identified into distinct proficiency levels, determined by their performance scores about the benchmarks corresponding to each level. For a visual representation, please refer to Table 2, showcasing the PISA Mathematics Proficiency Levels (2005, pp. 260-261).

Table 2

Six Performance Levels Adapted from the OECD

Level	Score Points on the PISA Scale	Summary Descriptions for Six Levels of Overall Mathematical Literacy
6	Above 669	At Level 6 students can conceptualise, generalise, and utilise information based on their investigations and modelling of complex problem situations. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply their insight and understandings along with a mastery of symbolic and formal mathematical operations and relationships to develop new approaches and strategies for attacking novel situations. Students at this level can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments, and the appropriateness of these to the original situations.
5	607 to 669	At Level 5 students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare, and evaluate appropriate problem-solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterisations, and insight pertaining to these situations. They can reflect on their actions and formulate

		and communicate their interpretations and reasoning.
4	545 to 607	At Level 4 students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic, linking them directly to aspects of real-world situations. Students at this level can utilise well-developed skills and reason flexibly, with some insight, in these contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments and actions.
3	482 to 545	At Level 3 students can execute clearly described procedures, including those that require sequential decisions. They can select and apply simple problem-solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They can develop short communications reporting their interpretations, results and reasoning.
2	420 to 482	At Level 2 students can interpret and recognise situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures, or conventions. They are capable of direct reasoning and making literal interpretations of the results.
1	358 to 420	At Level 1 students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry our routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli.

Given that the mathematics test was derived from PISA (OECD, 2009), Table 2 serves as a valuable tool for deciphering mathematics test scores in the context of student ability. This interpretation, as outlined by Tshering (2012) and Tshering and Prain (2011), empowers teachers to design targeted instructional activities tailored to their students' specific needs and proficiencies.

Method

This section intricately explores using the mathematics test for in-depth data acquisition. It not only demonstrates techniques for profiling students' mathematical knowledge and skills but also delineates the distribution of students across PISA mathematics proficiency scale levels and delves into an examination of their thinking skills. The analysis spans student performance on the mathematics test at national, district, and school levels, unveiling geographic nuances. Navigating through statistical intricacies, the researchers delve into 'between- and within-school' variances, culminating in a comparative study against international benchmarks. This illuminates the global standing of the Bhutanese Education System.

A Profile of Students’ Mathematical Knowledge and Skills

Analyzing students' mathematical knowledge and skills on a per-topic or curriculum-strand basis offers the advantage of delivering diagnostic insights into their comprehension of specific mathematical areas. Profiling can be achieved through the use of a test specification. For the Mathematics test, two distinct specifications were formulated—one aligning with the PISA Mathematical Literacy Framework (refer to Table 3) and the other adhering to the Grade 10 Mathematics Curriculum (refer to Table 5).

Table 3

Specification of the Mathematics Test in terms of the PISA Mathematical Literacy Framework

Mathematics Domain	**Competency Cluster**												**Total**
	Reproduction				Connections				Reflection				
	MR	CR	OR	TO	MR	CR	OR	TO	MR	CR	OR	TO	
Space and Shape	9a	5a 22a	1a	**4**	13a		1b 3a 3b 4a 24a	**6**			3c	**1**	**11**
Quantity	23b	14a 16a 16b 23a		**5**		12a 19a 23c		**3**			16c	**1**	**9**
Change and Relationship	7b 7c	26a	2a 6a 6c	**6**	7a 7d	6b 15a	2b	**5**		15b	26b	**2**	**13**
Uncertainty	18a	8a 17a		**3**	17b		11a 21a 25a	**4**	20a		10a	**2**	**9**
Total	5	9	4	**18**	4	5	9	**18**	1	1	4	**6**	**42**

MR=Multiple-choice response; CR=Closed-constructed response; OR=Open-constructed response; TO= Total number of items in a competency cluster

Table 4 juxtaposes the strands of the Bhutanese Grade 10 mathematics curriculum with the domains outlined in the PISA Mathematical Literacy Assessment Framework to examine their alignment. The Table highlights the correspondence between the strands of the Bhutanese Grade 10 mathematics curriculum and the PISA Mathematical Literacy Assessment Framework domains, with shaded cells indicating alignment. The mapping identifies similar mathematical knowledge and skills emphasized in the Bhutanese Grade 10 mathematics curriculum strands and the PISA Mathematical Literacy domains. For example, the Numbers strand in the Bhutanese curriculum emphasizes understanding number meanings, ordering and representing real numbers, and applying various number theory concepts. Correspondingly, the Quantity domain in the PISA framework underscores skills such as understanding relative size, recognizing numerical patterns, using numbers to represent quantities and attributes of real-world objects, and estimation (OECD, 2004b). Given the similarity in the emphasized mathematical knowledge and skills in Numbers and Quantity, they are mapped together

Table 4
Comparison of the Bhutanese Grade 10 Mathematics Curriculum Strands and the PISA Mathematical Literacy Assessment Framework Domains

Bhutanese Grade 10 Mathematics Curriculum Strands	**PISA Mathematical Literacy Assessment Framework Domains**			
	Space and Shape	Change and Relationship	Quantity	Uncertainty
10A: Numbers				
10B: Operations				
10C: Pattern				
10D: Measurement				
10E: Geometry				
10F: Data Management & Probability				

Notes: 1. White color=does not match; 2. Blue colour= does match only in terms of the 42 PISA mathematics items; 3. Green colour= does match only in terms of the similarities in mathematical knowledge and skills; and 4. Purple colour= does match both in terms of the 42 PISA mathematics items and the similarities in mathematical knowledge and skills

Next, the mapping extends by allocating the 42 PISA mathematics items to the Bhutanese Grade 10 mathematics curriculum strands based on how closely the items measure the mathematical knowledge and skills embedded in the strands. The comparison results, as illustrated in Table 4, conclude that the strands of the Bhutanese Grade 10 mathematics curriculum align with the domains of the PISA Mathematical Literacy Assessment Framework. Consequently, a test specification for the Mathematics test was developed, focusing on the Grade 10 mathematics curriculum strands. Table 5 provides a detailed representation of the test specification.

Table 5

Test Specification of the Mathematics Test in Terms of the Bhutanese Grade 10 Mathematics Curriculum Strands

Mathematics Curriculum Strands	**Competency Cluster**												**Total**
	Reproduction				Connections				Reflection				
	MR	CR	OR	TO	MR	CR	OR	TO	MR	CR	OR	TO	
10A: Numbers			2a	**1**									**1**
10B: Operations		16a 26a		**2**			2b	**1**			26b	**1**	**4**
10C: Pattern	7c 7b	5a 14a 16b	6a 6c	**7**	7a 7d	6b	3a 3b 24a	**6**			3c 16c	**2**	**15**
10D: Measurement	23b	22a 23a	1a	**4**	13a	19a 23c	1b 4a	**4**					**9**
10E: Geometry	9a			**1**									**1**
10F: Data Management & Probability	18a	8a 17a		**3**	17b	12a 15a	11a 21a 25a	**6**	20a	15b	10a	**3**	**12**
Total	5	9	4	**18**	4	5	9	**18**	1	1	4	**6**	**42**

The alignment between the strands of the Bhutanese Grade 10 mathematics curriculum and the domains of the PISA Mathematical Literacy Assessment Framework suggests that the 42 PISA mathematics items possess the potential to effectively evaluate the mathematical knowledge and skills of Grade 10 Bhutanese students. However, more than mere similarities in objectives and domains are required to confirm the suitability of a set of test items designed for the PISA Mathematical Literacy Assessment Framework domains to test the Bhutanese Grade 10 mathematics curriculum strands. Ensuring the 42 PISA items align well with the Bhutanese Grade 10 mathematics curriculum standards is crucial. Utilizing Webb's alignment method (Webb, 1999, 2006), Tshering (2012) demonstrated that the mathematics test, developed with 42 PISA items and Grade 10 mathematics curriculum elements, exhibited a robust alignment. This alignment signifies that the mathematics test is adept at assessing the mathematical knowledge and skills of Grade 10 Bhutanese students. Furthermore, it implies that the mathematics test can effectively utilize the PISA Mathematics Proficiency Scale, as depicted in Table 2.

Table 6 provides a comprehensive profile of Grade 10 Bhutanese students' mathematical knowledge and skills, referencing the PISA Mathematical Literacy Framework.

Table 6

Grade 10 Students' Mean Performance Scores on the Four Domains of the PISA Mathematical Literacy Framework

Domain	Mean	S.E	Location on the PISA Mathematics Proficiency Scale
Shape & Space	355	3.04	Below Level 1
Quantity	450	2.53	Level 2
Change & Relationship	414	2.21	Level 1
Uncertainty	385	2.47	Level 1

Note: The scores were standardized with a mean of 100 and a standard deviation of 50.

In Table 6, the data reveals that, on average, Grade 10 Bhutanese students scored below Level 1 of the PISA Mathematics Proficiency Scale in the domains of Space and Shape, Change and Relationship, and Uncertainty within the PISA Mathematical Literacy Framework. Conversely, they attained Level 2 in Quantity on the PISA Mathematics Proficiency Scale.

Moving on to the second analysis, the mean performance scores of students in the specific strands of the Bhutanese Grade 10 Mathematics Curriculum are computed. The outcome of this computation is detailed in Table 7.

Table 7

Grade 10 Students' Mean Performance Scores on the Six Strands of the Bhutanese Grade 10 Mathematics Curriculum

Domain	Mean	S.E	Location on the PISA Mathematics Proficiency Scale
Numbers	465	4.73	Level 2
Operations	385	7.15	Level 1
Pattern	409	4.80	Level 1
Measurement	408	3.74	Level 1
Geometry	465	4.41	Level 2
Data Management & Probability	383	2.81	Level 1

Table 7 shows that Grade 10 Bhutanese students exhibit proficiency in showcasing the mathematical knowledge and skills aligned with Level 2 of the PISA Mathematics Proficiency Scale, specifically in the strands of Numbers and Geometry within the Grade 10 Mathematics Curriculum. Additionally, as depicted in Table 6, Grade 10 Bhutanese students showcase their mathematical capabilities at Level 1 of the PISA Mathematics Proficiency Scale, particularly in the strands of Operations, Pattern, Measurement, and Data Management and Probability within the Grade 10 Mathematics Curriculum.

Percentage of Students at each Level of the PISA Mathematics Proficiency Scale

The percentage of students at each PISA Mathematics Proficiency Scale level was meticulously calculated to offer a more detailed understanding of students' mathematical knowledge and skills. The outcome of this calculation is presented in Table 8.

Table 8

Percentage of Bhutanese Students at Individual PISA Mathematics Proficiency Levels

Proficiency Level	Percentage	SE
Below 1	27.03	2.03
1	35.16	1.67
2	26.62	1.44
3	8.092	1.26
4	2.7	0.65
5	0.30	0.14
6	0.10	0.08

Table 8 shows that roughly a quarter of Grade 10 Bhutanese students scored below Level 1 of the PISA Mathematics Proficiency Scale. Most students demonstrated performance at Level 1 and Level 2 on the PISA Mathematics Proficiency Scale. Notably, approximately one-tenth of the students exhibited proficiency at Level 3 or beyond on the PISA Mathematics Proficiency Scale.

A Profile of Students' Thinking Skills

Students' thinking skills are categorized into three overarching competency clusters: recollection, connection, and reflection, as outlined by OECD (2009). Table 9 provides a snapshot of the student's average performance scores across these three comprehensive competency clusters.

Table 9

Mean Performance Scores on the Three Competency Clusters

Competency Clusters			**Reproduction**	**Connection**	**Reflection**
	Mean		462	365	353
		SE	4.72	3.22	3.73
Reproduction	463	4.72	0	^	^
Connection	365	3.22	v	0	v
Reflection	353	3.73	v	^	0

Note: 1. The table is read across the row for a cluster to compare with the clusters listed along the top of the table.
2. v denotes less than and ^ denotes greater than.
3. The standard errors include link errors as well.

Multiple comparisons of students' mean scores on the competency clusters, employing the Bonferroni correction, established that the mean performance scores on these competencies exhibited statistically significant differences ($p<0.05$). This highlights the diversity among students in their proficiency in higher-order thinking skills, indicating that certain individuals can effectively employ such skills in solving intricate mathematical problems.

Student Performance on the Mathematics Test at National, District, and School Levels

When student achievement is analyzed at district and school levels, the information gleaned becomes more pertinent to stakeholders than an aggregation solely at the national level. Therefore, this section presents the mean performance scores of students at the national, district, and school levels.

Firstly, the national mean performance score of Bhutanese Grade 10 students is computed at 361 (S.E.= 4.1). Moving on, the mean performance scores of the 19 districts in Bhutan are calculated, as illustrated in Table 10. The one-way analysis of variance revealed a statistically significant difference in the mean scores of the Mathematics test across the 19 districts ($p<0.05$). Post-hoc comparisons employing the Bonferroni correction further indicated that the mean performance scores of certain districts were statistically significantly different from others ($p<0.05$). However, it is worth noting that most districts did not exhibit statistically significant differences in their mean performance scores compared to their counterparts.

Table 10

District Mean Performance Scores on the Mathematics Test

			A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**
	Mean		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
		SE	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
A	**389**	8.0	0	•	v	•	•	•	•	•	•	•	•	•	•	•	•	•	v	•	•
B	**412**	13.2	•	0	v	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
C	**468**	5.8	^	^	0	^	^	^	^	^	^	^	^	^	^	^	^	^	^	^	^
D	**415**	5.8	•	•	v	0	•	•	•	•	•	•	•	•	•	•	•	^	•	•	•
E	383	18.7	•	•	v	•	0	•	•	•	•	•	•	•	•	•	•	•	v	•	v
F	396	38.0	•	•	v	•	•	0	•	•	•	•	•	•	•	•	•	•	v	•	•
G	396	19.8	•	•	v	•	•	•	0	•	•	•	•	•	•	•	•	•	v	•	•
H	**413**	13.5	•	•	v	•	•	•	•	0	•	•	•	•	•	•	•	•	v	•	•
I	**396**	9.2	•	•	v	•	•	•	•	•	0	•	•	•	•	•	•	•	v	•	•
J	398	26.8	•	•	v	•	•	•	•	•	•	0	•	•	•	•	•	•	v	•	•
K	**410**	21.4	•	•	v	•	•	•	•	•	•	•	0	•	•	•	•	•	v	•	•
L	**395**	14.9	•	•	v	•	•	•	•	•	•	•	•	0	•	•	•	•	v	•	•
M	**400**	6.7	•	•	v	•	•	•	•	•	•	•	•	•	0	•	•	•	v	•	•
N	395	17.4	•	•	v	•	•	•	•	•	•	•	•	•	•	0	•	•	v	•	•
O	**407**	19.6	•	•	v	•	•	•	•	•	•	•	•	•	•	•	0	•	v	•	•
P	**377**	5.8	•	•	v	v	•	•	•	•	•	•	•	•	•	•	•	0	v	•	v
Q	**433**	5.8	^	•	•	•	^	^	^	•	^		•	^	•	^	•	^	0	•	•
R	**406**	10.7	•	•	v	•	•	•	•	•	•	•	•	•	•	•	•	•	•	0	•
S	**420**	15.6	•	•	v	•	^	•	•	•	•	•	•	•	•	•	•	^	•	•	0

Note:1. The table is read across the row for a district to compare its performance with the districts identified along the top of the table.

2. v denotes less than, ^ denotes greater than, * refers to the same means and standard errors shown in the columns under 'Mean' and 'SE' for the corresponding districts, and • denotes non-statistically significant difference.
3. The standard errors include the link errors as well.

4. Bold numbers indicate a statistically significant difference from the national mean performance scores at p<0.05.
5. Alphabets denote 19 school districts.

Thirdly, the mean performance scores of the 60 schools were computed, with a glimpse of the post-hoc test results provided for 10 schools in Table 11. The one-way analysis of variance unveiled a statistically significant difference at the p<0.05 level in the mean scores of the Mathematics test across the 60 schools. Post-hoc comparisons of schools, incorporating the Bonferroni correction, indicated that the mean scores of certain schools were statistically significantly different from others (p<0.05). It is noteworthy, however, that most schools did not exhibit statistically significant differences in their mean performance scores compared to their counterparts.

Table 11

School Mean Performance Scores on the Mathematics Test

			1	2	3	4	5	6	7	8	9	12	...
Mean			404	395	387	433	439	388	470	381	365	391	
		SE	14.6	15.1	11.6	13.6	13.9	11.6	12.1	16.2	15.1	17.4	
1	**404**	14.6	0	•	•	•	•	•	•	•	•	•	
2	**395**	15.1	•	0	•	•	•	•	•	•	•	•	
3	**387**	11.6	•	•	0	•	•	•	**v**	•	•	•	
4	**433**	13.6	•	•	•	0	•	•	•	•	•	•	
5	**439**	13.9	•	•	•	•	0	•	•	•	•	•	
6	**388**	11.6	•	•	•	•	•	0	**v**	•	•	•	
7	**470**	12.1	•	•	^	•	•	^	0	^	^	•	
8	381	16.2	•	•	•	•	•	•	**v**	0	•	•	
9	365	15.1	•	•	•	•	•	•	**v**	•	0	•	
12	391	17.4	•	•	•	•	•	•	•	•	•	0	
13	**401**	13.7	•	•	•	•	•	•	•	•	•	•	

Note: 1. The table is read across the row for a school to compare its performance with the schools identified along the top of the table.
2. v denotes less than, ^ denotes greater than, * refers to the same means and standard errors in the columns under 'Mean' and 'SE' for the corresponding schools, and • denotes non-statistically significant difference.
3. The standard errors include the link errors as well.
4. Bold numbers indicate a statistically significant difference from the national mean performance scores at p<0.05.
5. Counting numbers denote schools.

Performance of Students by Their Schools' Locale

The Ministry of Education categorizes schools into four groups based on their locale, specifically urban settings. These categories include (a) urban, (b) semi-urban, (c) semi-rural, and (d) rural (Ministry of Education, 2022). The mean performance scores of students, classified by their schools' locale in relation to urban settings, were meticulously calculated. The outcomes of this analysis are presented in Table 12. As indicated in Table 12, the mean score of urban school students surpasses that of semi-rural and rural school students. Similarly, the mean score of students in semi-urban schools outpaces the mean score in semi-rural schools.

Table 12

Mean Performance Scores of Students by Their Schools' Locale

Locale			Urban	Semi-Urban	Semi-Rural	Rural
	Mean		365	353	318	337
		SE	4.39	9.25	13.58	6.07
Urban	365	4.39	0	12	**47**	**28**
Semi-Urban	353	9.25	-12	0	**35**	16
Semi-Rural	318	13.58	**-47**	**-35**	0	-19
Rural	337	6.07	**-28**	-16	19	0

Note: 1. The standard errors include the link errors as well.
2. The significant differences are shown in bold, $p<0.05$.

The Between- and Within-School Variances

To analyze the performance scores of Grade 10 Bhutanese students, the between- and within-school variances were calculated using the VARCOMP procedure in SPSS (SPSS Inc, 2004). The results indicate that the within-school variance (37198.20) is notably more significant than the between-school variance (4597.53), with only 11% of the variance being attributed to the between-school factor.

International Benchmarks for the Bhutanese Education System

The national mean score of Grade 10 Bhutanese students in the Mathematics test can be effectively compared with the national mean scores of countries that participated in PISA 2003 due to the established linkage between the Mathematics test and the PISA 2003 Mathematical Literacy Test (see Tshering & Prain, 2011). This linkage facilitates the creation of international benchmarks for the Bhutanese education system. Utilizing the national mean scores from PISA 2003 participant countries and the national mean score of Grade 10 Bhutanese students, a league table has been formulated, as depicted in Table 13.

Table 13

Mean Mathematics Test Scores of Bhutan and the Countries that Participated in PISA 2003

Country	Mean	SE	Country	Mean	SE	Country	Mean	SE
Hong Kong-China	550	4.5	France	511	2.5	Italy	466	3.1
Finland	544	1.9	Sweden	509	2.6	Portugal	466	3.4
Korea	542	3.2	Austria	506	3.3	Greece	445	3.9
Netherlands	538	3.1	Germany	503	3.3	Serbia	437	3.8
Liechtenstein	536	4.1	Ireland	503	2.4	Turkey	423	6.7
Japan	534	4	***OECD average***	***500***	***0.6***	Uruguay	422	3.3
Canada	532	1.8	Slovak Republic	498	3.3	Thailand	417	3.0

Belgium	529	2.3	Norway	495	2.4	Mexico	385	3.6
Switzerland	527	3.4	Luxembourg	493	1	**BHUTAN**	361	4.1
Macao-China	527	2.9	Hungary	490	2.8	Indonesia	360	3.9
Australia	524	2.1	Poland	490	2.5	Tunisia	359	?
New Zealand	523	2.3	Spain	485	2.4	Brazil	356	4.8
Czech Republic	516	3.5	United States	483	2.9			
Iceland	515	1.4	Latvia	483	3.7			
Denmark	514	2.7	Russian Federation	468	4.2			

Note: 1. The national mean score data of the countries participating in PISA 2003 are adapted from OECD (2004, p. 358). The Bhutanese findings have been inserted.

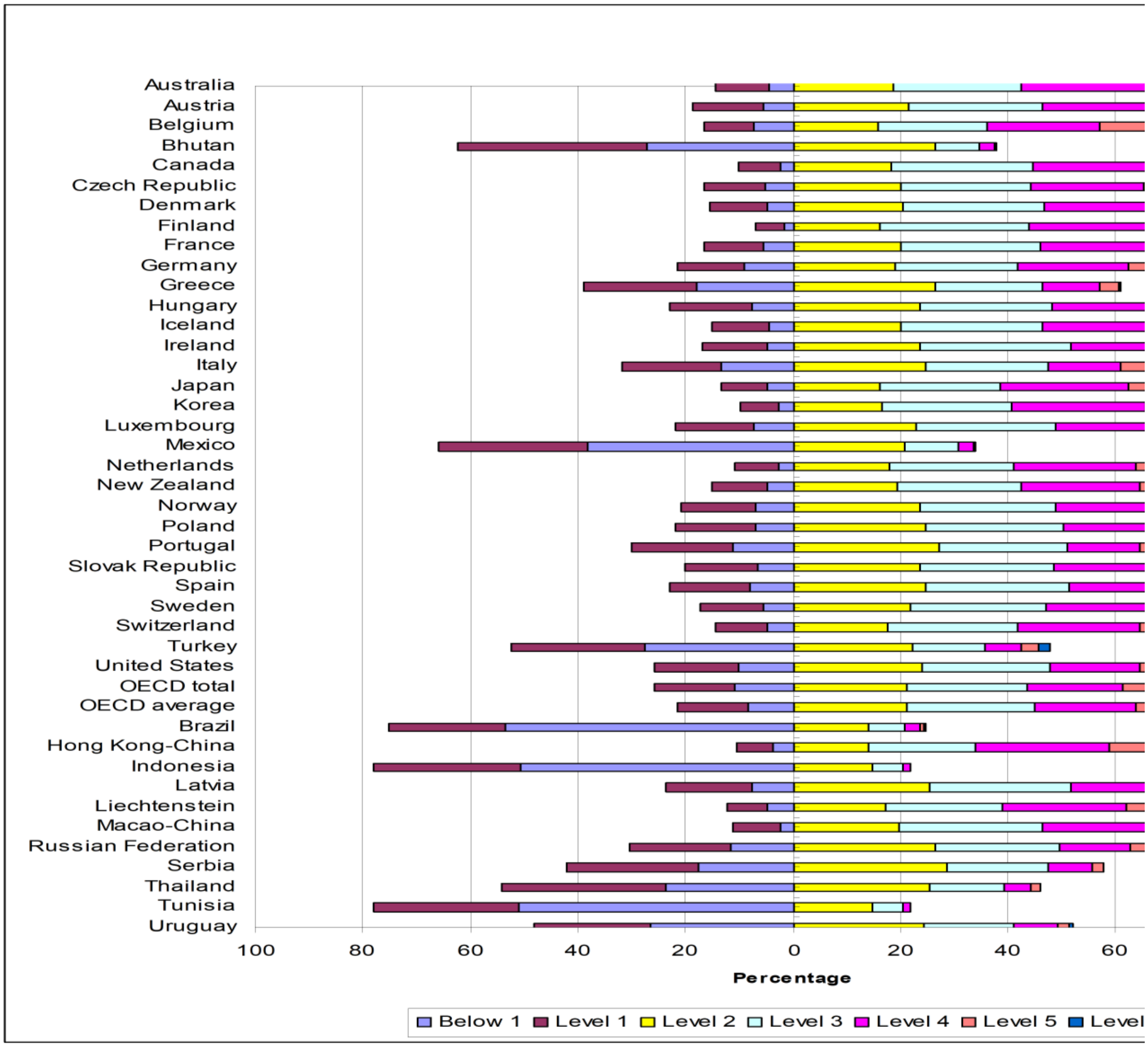
Table 13 shows that Grade 10 Bhutanese students' mean score is greater than the mean scores of Indonesia (360), Tunisia (359), and Brazil (356).

International benchmarks can be derived from the percentages of students at various PISA Mathematics Proficiency Scale levels. Figure 1 visually represents the distribution of Grade 10 Bhutanese students across different levels of the PISA Mathematics Proficiency Scale alongside students from countries participating in PISA 2003. Using data from PISA 2003 (OECD, 2004, p. 354) and the mathematics test, Figure 1 is a comparative illustration of proficiency levels.

On average, only about a third of students across the OECD countries attained levels five and six of the PISA Proficiency Scale, as with Grade 10 Bhutanese students.

Figure 1

Percentage of Students at each Level of the PISA Mathematics Proficiency Scale with Bhutan Included



Discussion

Profiling students' knowledge and skills according to subject content or standards offers valuable insights for parents, teachers, and schools to comprehend strengths and weaknesses in mathematics curriculum content or standards (refer to Tables 6 & 7). This understanding equips parents, teachers, and schools with critical insights into students' subject knowledge and skills, enabling strategic remedial lessons and personalized support (OECD, 2004, 2005, 2007). The examination of students' mathematical knowledge and skills, considering the four domains of the PISA 2003 Mathematical Literacy Framework and the Bhutanese Grade 10 Mathematics Curriculum, reveals that Grade 10 student performance ranges from the maximum of Level 2 to below Level 1 on the PISA Mathematics Proficiency Scale (see Tables 6 & 7). This observation suggests considerable room for improvement in teaching pedagogies and educational interventions. The alignment between teachers' subject content, pedagogical knowledge and skills, student characteristics, and student achievement is widely acknowledged (NCTM, 2000).

Drawing on OECD's assertion that one school year equals an average of 41 score points on the PISA Mathematics Proficiency Scale, it becomes evident that students scoring below Level 1 have limited mathematical knowledge and skills expected of a Grade 10 Bhutanese student. These students need help solving mathematical tasks requiring factual knowledge and routine procedures, putting them at risk of insufficient preparation for societal engagement beyond school (Thomson, 2008). The quarter of Grade 10 students scoring below Level 1 need more mathematical knowledge and skills for successful adaptation to rapid societal and technological change.

Recognizing the significance of higher-order thinking skills for lifelong learning and societal adaptation (OECD, 2005), this study categorizes students' thinking skills into three competency clusters: reproduction, connection, and reflection (see Table 9). The analysis reveals that students excel at lower-order thinking skills (Reproduction cluster) compared to higher-order thinking skills (Connection and reflection clusters). This emphasizes the predominant use of rote learning as the principal learning strategy, aligning with previous studies commissioned by the Royal Education Council of Bhutan (Educational Initiatives, 2009; iDiscoveri Education & Royal Education Council, 2009).The mean performance scores at national, district, and school levels allow schools and districts to exchange expertise and learn from each other's strengths and weaknesses (refer to Table 10). Notably, statistically significant differences in mean performance scores among schools and districts suggest variations that can inform strategic interventions and improvements (Barton, 2002). Schools exhibiting significant differences can offer valuable insights to others, fostering a culture of collaboration and improvement (Barber & Mourshed, 2007). This exchange of ideas can drive meaningful educational interventions, enhance school choice by parents, and increase parental participation, contributing to improved student achievement (Mullis et al., 2004; OECD, 2007).

The significant differences in student performance based on schools' locale (see Table 12) highlight the need for targeted educational interventions in semi-rural and rural schools. The observed pattern, where students in urban schools outperform those in semi-urban, semi-rural, and rural schools, aligns with previous studies commissioned by the Royal Education Council of Bhutan (Educational Initiatives, 2009; iDiscoveri Education & Royal Education Council, 2009).

While mean performance scores are commonly used to measure the effectiveness of an education system, it is essential to consider both quality and equity dimensions (Reynolds et al., 1994; Creemers & Kyriakides, 2008). As defined by these dimensions, an influential school should exhibit a significant positive difference in student achievement compared to other schools and a minimal difference in student achievement within the school (Creemers & Kyriakides, 2008). The study indicates that Bhutan's school education system, characterized by a comprehensive approach, is on track to achieving equity, with 11% of the variance explained by between-school differences and 89% explained by within-school factors (Hattie, 2009; Wang et al., 2023).

While international assessments such as PISA provide a basis for comparing education systems, caution is warranted when comparing test scores. This study positions Bhutan's education system relative to countries participating in PISA 2003, revealing opportunities for cross-fertilization of innovative approaches in school education. The comparative analysis underscores the need for countries with similar characteristics to exchange knowledge and skills related to school effectiveness (Johnson & Owen, 1998).

In conclusion, this study offers a comprehensive examination of Grade 10 Bhutanese students' mathematical knowledge and skills, shedding light on strengths, weaknesses, and opportunities for improvement within the education system. The findings provide a foundation for targeted interventions, collaborative learning among schools and districts, and informed decision-making to enhance student outcomes and adaptability in a rapidly changing societal landscape.

Implications for Instructional Practices

This study's findings and methodological approach carry profound implications for benchmarking student performance, forming the basis for strategic educational interventions to enhance the quality of the Bhutanese education system at various levels—national, school district, school, and classroom.

The exponential growth of the Bhutanese education system, from a student enrollment of around 400 in the 1960s to approximately 168,324 in 2021 (Policy & Planning Division, 2021), underscores the need for continuous efforts in balancing quantity and quality. The strain on educational resources necessitates a focused approach to ensure access and excellence in education. Quality assurance measures, exemplified by end-of-year examinations, are pivotal in monitoring syllabus coverage, student promotions, and teacher evaluations. However, the predominant reliance on recall and procedural knowledge observed in students performing at or below Level 1 of the PISA Mathematics Proficiency Scale raises questions about the alignment of these examinations with the development of higher-order thinking skills. A policy debate on the relationship between end-of-year examinations and student learning strategies is crucial for optimizing the benefits of these assessments.

As indicated by percentages at different PISA Mathematics Proficiency Scale levels, disparities in student performance emphasize the importance of varied teaching pedagogies and adaptable school curricula. It is essential to address diverse learning needs to avoid perpetuating a pyramidal pattern in enrollments, limiting students' knowledge and skills for meaningful participation in the nation's economy or lifelong learning. A policy framework aiming for a rectangular rather than pyramidal enrollment pattern is essential for aligning educational opportunities with students' abilities and fostering success in a globalized economy.

The analysis reveals that 11% of the variance in student performance is attributed to between-school differences. While this could signal achievement in equity and access, the clustering of student achievement at the lower end of the proficiency scale necessitates a focus on reducing performance differences between schools and improving overall student performance. Systemic interventions, such as teacher education, can play a vital role. The influence of schools' locale on student learning, with urban schools outperforming others, underscores the need for interventions ensuring equitable access to educational opportunities for all students and schools.

The methodological aspects of the analyses offer valuable insights for the Bhutan Board of Examinations. The linkage of different tests over time presents an opportunity to study changes in student performances. Additionally, relating students' performance scores to curriculum standards through a proficiency scale provides systemic feedback to stakeholders, aiding informed decision-making. Profiling students' knowledge and skills in different content areas from national examination data offers insights into relative strengths and weaknesses, guiding decisions on professional development programs for teachers and informing policymakers, schools, and parents about student outcomes.

When interpreted in terms of proficiency levels, mean scores from national examinations can serve as benchmarks, facilitate the inter-school exchange of expertise, and inform professional development programs. Generating percentages of students at different proficiency levels aids in planning educational interventions tailored to students' learning needs. Profiling higher-order thinking skills informs the development of pedagogies and curricula demanding complex thinking skills. Disaggregated scores at district and school levels enable benchmarking, fostering an environment where schools learn from each other's strengths and weaknesses. In conclusion, the implications for instructional practices highlight the need for a nuanced, adaptive, and evidence-based approach to education policy and practice in Bhutan. Addressing the identified challenges and leveraging the strengths revealed by the analyses will contribute to the ongoing improvement and evolution of the Bhutanese education system.

The Bhutanese education system could benefit from reassessing examination strategies to promote higher-order thinking skills. Efforts should focus on enhancing pedagogical approaches, encouraging innovative teaching methods, and tailoring interventions based on schools' locales to address disparities effectively. Engaging in strategic policy debates about the alignment between examinations and learning strategies is crucial, alongside a commitment to continuous professional development for teachers. Longitudinal studies can offer insights into the effectiveness of interventions, while international collaboration provides an opportunity to learn from global best practices. Initiatives to foster parental engagement, explore school leadership practices, and integrate technology into education are essential components of a comprehensive strategy for the future. These directions aim to ensure an equitable, innovative, and adaptive education system in Bhutan.

Reference

Barber, M., & Mourshed, M. (2007). *How the world's best-performing school systems come out on top.* http://www.mckinsey.com/clientservice/socialsector/resources/pdf/Worlds_School_systems_final.pdf

Barton, P. E. (2002). *Perspectives on background questions in the national assessment of educational progress*. Washington, D.C: National Assessment Governing Board.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge Taylor & Francis Group.

Education Monitoring & Support Services Division. (2003). *A report on examinations in schools*. http://www.education.gov.bt/Publications/Report%20on%20Exam.pdf.

Educational Initiatives. (2009). *Annual status of student learning*. Educational Initiatives Pvt. Ltd.

Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge Taylor & Francis Group.

iDiscoveri Education & Royal Education Council. (2009). *The quality of school education in bhutan :Reality and opportunities*. http://www.rec.org.bt/formdownload.htm.

Johnson, E. G., & Owen, E. (1998). *Linking the national assessment of educational progress (naep) and the third international mathematics and science study (TIMSS): A technical report* (No. NCES 98-499). National Center for Educational Statistics.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A., & Erberber, E. (2005). TIMSS 2007 assessment frameworks. http://pirls.bc.edu/TIMSS2007/frameworks.html

Mullis, I. V. S., Matin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from iea's trends in international mathematics and science study at the fourth and eight grades*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

NCTM. (2000). *Principles and standards for school mathematics*. NCTM.

OECD. (2004). *Learning for tomorrow's world: first result from PISA 2003*. OECD.

OECD. (2005). *Are students ready for a technology-rich world? What PISA studies tell us*. OECD.

OECD. (2007). *PISA 2006: science competencies for tomorrow's world: Volume 1 analysis*. OECD.

OECD. (2009). *Take the test: Sample questions from OECD's PISA assessments*. OECD publishing.

OECD. (2010). *PISA 2009 results: What students know and can do: student performance in reading, mathematics and science* (Vol. 1). OECD Publishing.

Policy and Planning Division. (2008). *General Statistics 2008*. http://www.education.gov.bt/.

Policy and Planning Division. (2021). *Annual Education Statistics, 2009*. http://www.education.gov.bt/wp-content/uploads/2022/03/AES-2021-Final-Version.pdf

Roediger, H.L., Putnam, A.L., & Smith, M.A. (2011). Ten benefits of testing and their applications to educational practice. *Psychology of Learning and Motivation, 35*. https://mrbartonmaths.com/resourcesnew/8.%20Research/Memory%20and%20Revision/10%20benefits%20to%20testing.pdf

iDiscoveri Education & Royal Education Council. (2009). *Bhutan's annual status of student learning 2010*. Educational Initiatives.

Reynolds, D., Creemers, B. P. M., Nesselrodt, P. S., Schaffer, E. C., Stringfield, S., & Teddlie, C. (1994). *Advances in school effectiveness research and practice*. Elsevier Science Ltd.

SPSS Inc. (2004). SPSS for Windows (Version 18.0). SPSS Inc.

Thomson, S. (2008). *Exploring scientific literacy; how Australia measures up: the PISA 2006 survey of students' scientific, reading and mathematical literacy skills*. In W. Whitham (Eds.), PISA National reports http://www.acer.edu.au/ozpisa/reports.html

Tshering, G. (2012). *Developing a national assessment model to inform educational policy in Bhutan.* https://opal.latrobe.edu.au/articles/thesis/Developing_a_national_assessment_model_to_inform_educational_policy_in_Bhutan/21844797

Tshering, G. (2016). Setting performance standards by using the Bookmark Method. *RABSEL the CERD Educational Journal*, *17*(1), 53-69.

Tshering, G. & Prain, V. (2011). Benchmarking the performance of Bhutanese students with the performance of the students from the OECD's PISA countries. *Educational Research and Evaluation, 17*(4), 263-281.

Wang, X. S., Perry, L. B., Malpique, A., & Ide, T. (2023). Factors predicting mathematics achievement in PISA: a systematic review. *Large-scale Assessments in Education*, *11*(1), 24.

bb, N. L. (1999). *Research monograph No. 18: Alignment of science and mathematics standards and assessments in four states*. National Institute for Science Education.

Webb, N. L. (2006). Identifying content for student achievement tests. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 155-180). Lawrence Erlbaum, Inc.

Weller, L. D. (1996). Benchmarking: a paradigm for change to quality education. *The TQM Magazine*, *8*(6), 24-29.