Utilising a differentiated assessment method in mathematics class: computer adaptive testing for tracking student progress

Baurzhan Yessingeldinov, Zamira Rakhymbayeva, Aikorkem Zhapparova & Laila Tursynova

Centre for Pedagogical Measurements under Autonomous Educational Organisation *Nazarbayev Intellectual Schools*
Astana, Kazakhstan

ABSTRACT: Assessment is an essential component of learning and teaching as it allows teachers, policymakers and other stakeholders to be informed about students' achievements and to make decisions based on the results and evidence. Alongside, assessment helps to harmonise lesson plans and to model activities for students to meet their needs. In the context of mathematics, it helps to develop students' mathematical skills; namely, to prepare them to be accepted by universities. The evolution of technologies accelerates the ways of providing instant feedback and monitoring the quality of students' proficiency. The authors focus on using computer adaptive testing to track students' academic achievements in an organisation of differentiated approach to teaching mathematics. In addition, a survey of teachers was conducted to determine their opinion on computer adaptive testing as a form of assessment. The results of the survey showed that teachers have a positive attitude towards computer adaptive testing, and note its positive aspects, such as taking less time to conduct, accuracy of results and motivation of students.

Keywords: Assessment, differentiated approach, teaching mathematics, computer adaptive testing, technologies in education

INTRODUCTION

Multiple studies have revealed that mathematics is commonly perceived as a challenging subject in education, yet it is a mandatory subject for all students in Kazakhstan. Mathematics serves as a foundation for a multitude of skills, including problem solving and logical reasoning, which are highly sought after by employers in various industries. Hence, teachers are encouraged to utilise a diverse range of instructional resources and activities in the classroom to provide high-quality, evidence-based instruction [1]. Adaptive teaching strategies help to support students in meeting their learning goals. Since students possess unique needs and capabilities, differentiation in the classroom is a critical aspect of the learning process, which is based on the principle that each student learns differently and requires customised instruction to meet their individual needs. This approach enables teachers to create an optimal learning environment for all students in the class.

Assessment holds a significant position in the realm of education, as it plays a crucial role in determining the progression of students' cognitive abilities, including reasoning, decision-making, analysis, synthesis and critical thinking [2]. The provision of formative data on student performance is crucial for teachers, as it enables them to effectively plan the learning process and cater to the individual needs of each student. Formative assessment, which entails ongoing observation, evaluation and feedback to students during the learning process, can be used as an effective tool to inform instruction. To optimise its benefits, teachers should consistently review students' data to detect patterns and trends, and to adjust instruction accordingly, with the aim of meeting the individual needs of students. It is imperative to acknowledge that the scope of assessment should extend beyond conventional pencil-and-paper tests, and encompass the integration of technology to provide efficient and immediate access to student progress data, enabling prompt intervention.

One of the technologically advanced forms of assessment is computer adaptive testing (CAT). CAT is a form of administering assessments in which the difficulty level of questions is adapted based on the student's performance, as determined by computer algorithms. This dynamic process allows for the selection of questions from a pre-determined pool that are tailored to the student's proficiency level [3]. The continuous adjustment of questions during the assessment ensures that students receive questions that are appropriate to their abilities.

The authors present exploratory research to examine the perspective of teachers on computer adaptive testing as a form of differentiated assessment.

LITERATURE REVIEW

A differentiated approach to teaching is not an innovation, it remains a topical issue in the modern educational process, but most teachers are struggling with applying differentiation in the classroom. According to Black and Wiliam, the goal of a differentiated approach to teaching is to ensure that all students have the opportunity to learn and achieve success, regardless of their individual abilities and learning styles. A differentiated approach to teaching would involve tailoring instruction to meet the individual needs of students based on their skill level and learning needs. An effective lesson means that students will work towards the same learning objectives in different ways and approaches [4].

Students learn differently, each of them with specific skills and challenges. It is important for the teacher to identify individual difficulties and use the data for lesson planning, determining the further learning trajectory. It is possible to determine the effectiveness of this process and the result obtained through assessment. Activity in the lesson should be formative, the student should receive constant feedback to improve their own learning. Assessment is an important determinant of future student learning outcomes. Assessment allows teachers and learners to determine what is considered truly valuable in learning. Assessment should be reflected in the activities and teaching methods that the teacher uses in the classroom [5]. Teachers should improve their professional development by applying various methods and forms of assessment in the classroom, involving students in the discussion of assessment criteria [6].

Differentiated assessment is a component of the differentiated approach, and involves assessing students' understanding and progress in a way that is tailored to their individual needs and learning styles. Research on differentiated assessment in mathematics has shown that it can be effective in promoting student learning and engagement. For example, a study by Fuchs found that a differentiated assessment approach, which included the use of multiple measures of student understanding, was effective in improving the mathematics achievement of students with diverse learning needs [7]. However, it is important to note that differentiated assessment can be challenging to implement in practice, and requires a significant amount of planning, organisation and ongoing adaptation on the part of the teacher.

Monitoring students' learning achievements is one example of differentiated assessment. The purpose of monitoring for assessment is to identify areas, where students need additional support and to provide feedback to students, teachers and parents about the student's progress. The data collected through monitoring can be used to make adjustments to improve outcomes. A study by Bryk et al found that formative assessment, which includes monitoring student progress, can improve student achievement and help close achievement gaps [8]. The study found that when teachers regularly check for student understanding and give formative feedback, students perform better on summative assessments [8]. However, many teachers say that they need support to deliver monitoring [9]. Assistance to teachers can be facilitated through the implementation of CAT, leveraging advancements in technology.

CAT is a method of administering tests in which the difficulty level of subsequent test items is determined by the performance of previous items. This approach has been widely adopted in various fields, including mathematics education. By adjusting the difficulty level of test items based on the test taker's performance, CAT can provide a more precise measure of ability with fewer items than traditional fixed-form tests. CAT can provide a more efficient and accurate assessment for students with different levels of ability [10]. For example, different studies have found controversial effectiveness of CAT as compared to linear tests at identifying students with low mathematical abilities and with high mathematical abilities [11][12]. By identifying students' mathematical abilities through CAT, teachers can better tailor instruction to the needs of individual students, leading to improved learning outcomes. This is particularly important in mathematics education, where the ability to accurately assess students' mathematical abilities is crucial for diagnostic purposes, effective instruction and learning [13]. CAT is widely used in many countries and it could be a way to meet the international standards and improve the quality of education in Kazakhstan.

In this study, the teachers' perception of CAT was investigated based on the transition of the Kazakhstani mathematics monitoring examination from a linear testing form to adaptive testing. After several trials on the catR package using the R programming language [14], the expected *a posteriori* was decided to be used as the method of ability estimator, the maximum Fisher information criterion was applied for selecting the next item of the test, the standard error equal to 0.3 and the test length of 50 was set as a stopping criterion of the adaptive monitoring examination. CAT for schools in Kazakhstan was held on the open-source platform Concerto, developed by the Psychometric Centre of Cambridge Business School with the R language as a background for test calculations [15]. CAT is widely used in many countries and it could be a way to meet the international standards and improve the quality of education in Kazakhstan.

In the process of conducting the study, the authors used theoretical and empirical methods of research. The literature review provided a theoretical basis for the study and allowed the authors to determine approaches to the development of computer adaptive testing. The study aims at answering the following research question:

What are teachers' perceptions of CAT as first stand users of CAT developed reports in guiding students towards planning differentiated tutoring?

METHODS

Two different methods were employed in order to answer the research question: a survey and a focus group.

The study was a descriptive research that utilised a survey to gather data from mathematics teachers. The sampling method adopted was convenient sampling, and the survey was sent to 436 mathematics teachers in Kazakhstan through e-mail in a manner that allowed for anonymous responses. The survey was administered through the MS Forms platform, resulting in a response rate of 48.9% (213 participants) from 20 schools in Kazakhstan. The survey was conducted in either the Kazakh or Russian language, at the teacher's discretion, and was anonymous. The questionnaire consisted of two sections, with the first part comprising of multiple-choice questions aimed at collecting general information about the respondents, and the second part consisted of questions aimed at assessing the teachers' opinions on monitoring and CAT as part of the assessment process in mathematics lessons. The responses were collected using a six-point Likert scale, with options ranging from *strongly agree* to *strongly disagree* and *not sure (difficult to answer)*. Individuals who expressed a willingness to participate in focus groups were mandated to report their interest within the confines of the survey.

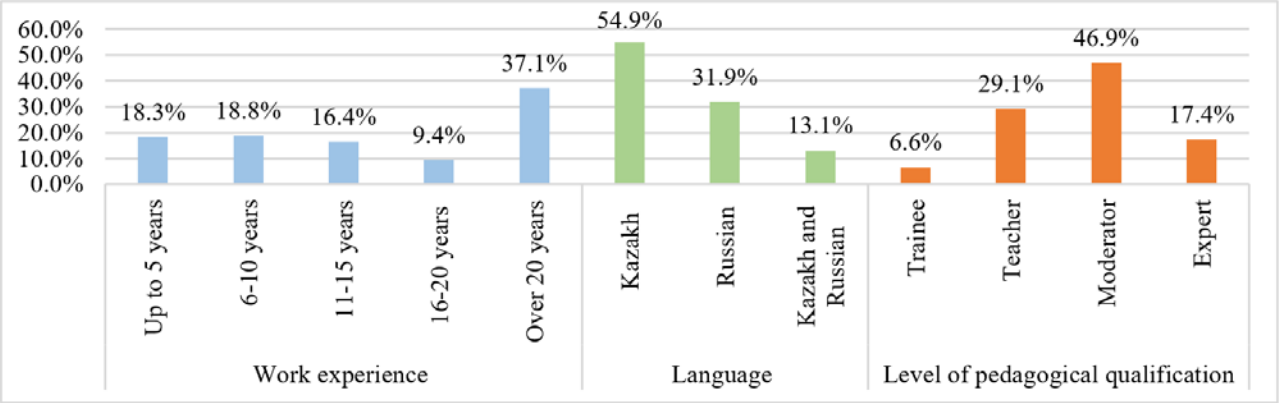

Figure 1: Information about the survey respondents.

The survey sought to elicit the views of teachers with varying backgrounds, including the language of instruction, years of work experience and the level of pedagogical qualification (as depicted in Figure 1). This allowed for the identification of opinions from teachers of diverse categories. The results indicated that over a third (37.1%) of the teachers had more than 20 years of work experience, the majority (54.9%) utilised Kazakh as their language of instruction and 64.3% held a high level of pedagogical qualification.

A subset of eight teachers from six different schools was selected to participate in a focus group discussion which is efficient technique for examining complex and new subjects. All of the participating teachers had at least eight years of experience and held a level of pedagogical qualification of either moderator or expert. The group consisted of two teachers who taught in Kazakh, three teachers who taught in Russian, and three teachers who taught in both Kazakh and Russian. The focus group discussion was conducted via MS Teams.

The focus group discussions proved a valuable technique in examining the impact of technology on individualised classroom activities and homework through personalised assessment. Through this medium, teachers were able to articulate their opinions on monitoring tests and its adaptive computerised version, as well as share their experiences on employing assessment results for differentiated instruction in the classroom. Additionally, the authors were able to gain a deeper understanding of the research question, which is expected to inform the planning of future research and the development of guidelines for teachers to maximise and enhance the utilisation of CAT reports, both at the individual and classroom levels.

The data analysis of the survey was performed using the open-source platform, jamovi [16]. The reliability of the survey was established through a high value of Cronbach's alpha (0.888). The Shapiro-Wilk test revealed that the collected data did not conform to normality, as indicated by the Q-Q plots. To test differences between two groups, the Mann-Whitney U test was utilised, and in cases where more than two independent variables were present, the Kruskal-Wallis test was applied as an alternative to the independent-samples *t*-test and one-way ANOVA test.

RESEARCH FINDINGS AND DISCUSSION

Table 1 presents the responses of teachers to a set of statements regarding monitoring and CAT as part of the assessment process in mathematics lessons. The results indicate that a majority of teachers (81.3%) agree or somewhat agree that monitoring is an effective tool for tracking students' progress in learning and is an example of a differentiated approach to assessment (74.7%).

One of the main questions was about teachers' awareness about CAT in Kazakhstan and the responses showed that more than half of the teachers (56.3%) know about the technology. This is a positive sign as it indicates that teachers are becoming more aware of the benefits of using computer-based testing methods. With a higher level of awareness, there is a good chance that more schools and educational institutions will adopt CAT in the future. CAT has the potential to revolutionise the way assessments are conducted in Kazakhstan. With its ability to provide a more accurate and efficient measurement of students' knowledge and skills, it can help educators make better informed decisions about

their students' learning progress. The results of the survey suggest that there is a growing interest in CAT, which is a positive sign for the future of education in Kazakhstan.

Table 1: The results of the survey.

Statement	Strongly agree	Agree	Somewhat agree	Disagree	Strongly disagree	Not sure
Monitoring of students' academic achievements is an effective tool for tracking student progress in learning	12.7%	31.0%	37.6%	8.5%	7.5%	2.8%
Monitoring is an example of a differentiated approach to the assessment of student academic achievements	6.1%	29.6%	39.0%	13.1%	8.5%	3.8%
Computer adaptive testing (CAT) allows for determining the level of academic achievements of students in a short period of time	9.4%	37.6%	33.8%	5.6%	6.1%	7.5%
CAT can be more effective than traditional tests because it takes less time to assess	8.5%	35.7%	34.7%	6.6%	6.1%	8.5%
The results of students obtained with the help of CAT correspond to the actual level of academic achievements of students (comparable to the current assessment)	6.1%	20.7%	40.4%	16.9%	7.0%	8.9%
I use monitoring results for lesson planning	11.7%	31.9%	36.6%	8.9%	5.2%	5.6%

As a study shows, CAT is reducing testing time for students and is more effective than traditional testing methods in identifying student strengths and weaknesses and providing accurate information on student abilities [3]. Teachers also noted that CAT allows them to determine the level of academic achievements of students (80.8%) and takes less time than traditional tests (78.9%).

Sixty-seven point two percent of respondents noted that the monitoring results are comparable to the current assessment and they use the results for lesson planning (80.2%). Proper use of monitoring results will improve the academic achievement of students. According to Tomlinson, teachers who used assessment results to differentiate instruction were able to effectively meet the individual needs of their students [17]. The study found that teachers who used assessment results to inform instruction were able to create targeted, individualised instruction that helped to improve student learning [17].

In the study outlined in this article, survey items regarding teachers who have heard about CAT before (Mdn = 5, Range = 5) were significantly different from those who have not heard about CAT before (Mdn = 4, Range = 5). The difference is relatively small and seems to be in terms of *agree* and *somewhat agree* and has a small effect size ranging from r = 0.229 to r = 0.370. The Kruskal-Wallis test has shown no statistically significant difference in the qualification level, experience level and the language of instruction

The participants in the focus group concurred that utilising technology in assessments and taking into account each student's unique needs are crucial with the rise of personalised learning and post-pandemic education. The significance of monitoring students' learning progress was acknowledged by educators, though the views of teachers were divided - some emphasised the significance of teacher-led monitoring, while others considered external monitoring results equally important.

The primary and notable benefit of CAT is time efficiency. Compared to the linear form of monitoring examinations taking place in schools of Kazakhstan, which typically required 70 to 175 questions to be answered over a span of two to five days, the configuration of CAT adapted from the previous mathematics monitoring requires a maximum of 50 questions to be answered within 90 minutes at most. At first, two out of eight teachers were aware of this advantage and believed that important sections of mathematics were not being adequately covered. However, after being provided with further explanations, they came to appreciate how CAT could accurately measure student progress, while still reducing testing time. Ultimately, teachers agreed that CAT was a more efficient and effective use of students' time.

Teachers reported utilising the outcomes of monitoring in the design of lessons. Upon recognising areas of difficulty, teachers choose tasks for students and also organise them into subgroups. Some educators indicated that they form groups of students not solely based on shared deficiencies, but also on the opposite principle, where students who have demonstrated mastery in tasks can assist their peers and engage in collaborative learning. Furthermore, teachers acknowledged the provision of individualised homework assignments. One teacher noted that he uses the results of CAT to determine the student's zone of proximal development. During the lesson planning process, this helps to identify topics that students can learn with each other and topics that need to be learned with the teacher. This will help to improve student learning.

Teachers believed that the outcomes of CAT are comparable to those of school assessments. Some educators observed that the majority of students obtain results that are similar to those from current evaluations. Other teachers noted that there were students who displayed either higher or lower results.

I have students who showed results higher than in the current assessment, perhaps this is due to the fact that multiple-choice tasks are used in monitoring, and in the class, we are solving still open tasks.

The views of teachers on the subject of student motivation in taking CAT examinations were consistent. Teachers recognised two crucial elements in CAT. First, improving the testing experience for students means avoiding questions that are either too challenging or too simple, preventing a decrease in their motivation towards the test. Some examples of teacher responses:

In my class, students were motivated to take the test, so their results were comparable to the current assessment.

…when the students were given a task according to their level of preparation, they tried to complete the task. I think it all depends on the first question ...

Teachers observed that students were motivated, when they received automatic results after finishing their tests:

Unlike previous monitoring tests, when students immediately saw the result of the test, they were delighted, they asked their classmates how many points they got ...

This unexpected factor showed the importance of feedback that provides valuable insight into student performance, not only indicates if an answer is correct or not. Therefore, it is crucial to take into account the benefits of automatic feedback in the assessment process. A study by Zimmerman found that students who received timely feedback on their work were more likely to make progress towards achieving their goals [18]. His view is also supported by one of the teachers, who said:

... I think it is right when the student immediately receives the result because the feedback should be timely, and the students remembered the tasks after completing them, so they could immediately draw conclusions about their results and think about why they made a mistake.

According to Henderson, students who received timely feedback on their work were more likely to set goals and focus their efforts on areas that needed the most attention [19].

CONCLUSIONS

It is well established that effective evaluation leads to effective teaching. Combining computer adaptive assessments and teacher observations provides valuable insights and validates a teacher's judgement. The feedback from these assessments supports teachers in making informed decisions to improve student academic outcomes.

The perception of computer adaptive testing among teachers can vary. Some teachers might see it as a useful tool for evaluating student learning and making informed instructional decisions. They might appreciate the ability to quickly and accurately assess student ability and receive detailed results. However, some teachers might question the accuracy of the results and the extent to which they truly reflect student learning.

A literature review, survey and teacher interviews revealed that monitoring student performance is the most efficient method of monitoring student learning progress. These results should inform lesson planning, customisation to meet individual student needs and overcoming challenges in mathematics education. Computer adaptive testing is efficient, as it tailors tasks to the student's abilities, and also motivates students. However, to maximise the impact of CAT, there needs to be the implementation of automatic detailed feedback for students.

REFERENCES

1. Ukobizaba, F., Nizeyimana, G. and Mukuka, A., Assessment strategies for enhancing students' mathematical problem-solving skills: a review of literature. *Eurasia J. of Mathematics, Science and Technol. Educ.*, 17, **3** (2021).
2. Jones, K.O., Harland, J., Reid, J.M. and Bartlett, R., Relationship between examination questions and Bloom's taxonomy. *39th IEEE Frontiers in Educ. Conf.*, 1-6 (2009).
3. Smith, J., Jones, M. and Brown, D., The effectiveness of computer adaptive testing in mathematics instruction. *J. of Educational Technol.*, 22, **4**, 45-55 (2018).
4. Black, P. and Wiliam, D., Inside the black box: raising standards through classroom assessment. *Phi delta kappan*, 92, **1**, 81-90 (2010).
5. Redondo, S.C., Bueno, D.C. and Figuerres, N.R., Analysis of classroom assessment skills and practices across levels in a Catholic educational institution. *CC: The Journal*, 13 (2017).
6. Akiri, E., Tor, H.M. and Dori, Y.J., Teaching and assessment methods: STEM teachers' perceptions and implementation. *Eurasia J. of Mathematics, Science and Technol. Educ.*, 17, **6** (2021).
7. Fuchs, L.S., Assessing intervention responsiveness: conceptual and technical issues. *Learning Disabilities Research & Practice*, 18, **3**, 172-186 (2003).
8. Bryk, A.S., Sebring, P.B., Allensworth, E., Luppescu, S. and Easton, J.Q., Organizing Schools for Improvement: Lessons from Chicago. University of Chicago Press (2010).

9. Yessingeldinov, B.T., Ashirbayev, N.K., Zhumykbayeva, A.K., Sarsekenov, R.M., Ismailova, G.M. and Bibekov K.T., Investigation of teachers' understanding of differentiated approach in teaching mathematics. *Cypriot J. of Educational Sciences*, 17, **5**, 1671-1679 (2022).
10. Weiss, D.J., Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measure. and Eval. in Counseling and Develop*., 37, **2**, 70-84 (2004).
11. Smith, L., Martin, W.G., Wan, A., Duenas, G., Shumack, K. and Beziat, T.L.R., Using reform pedagogy to improve students' application skills in college remedial mathematics courses. *Mathematics and Computer Educ.*, 49, **2**, 124-138 (2015).
12. Wainer, H. (Ed). *Computerized Adaptive Testing: a Primer*. (2nd Edn), Mahwah, NJ: Lawrence Erlbaum (2000).
13. Tay, J., Salazar, A. and Lee, H., Parental perceptions of STEM enrichment for young children. *J. for the Educ. of the Gifted*, 41, **1**, 5-23 (2018).
14. Magis, D. and Barrada, J.R., Computerized adaptive testing with R: recent updates of the package catR. *J. of Statistical Software*, 76, 1-19 (2017).
15. The Psychometrics Centre. Concerto Adaptive Testing Platform. University of Cambridge, 14 February 2023, https://www.psychometrics.cam.ac.uk/newconcerto
16. The jamovi project (2022). *jamovi* (Version 2.3), 04 February 2023, https://www.jamovi.org
17. Tomlinson, C.A. and Moon, T.R., *Assessment and Student Success in a Differentiated Classroom*. Ascd (2013).
18. Zimmerman, B., The effect of timely feedback on student progress. *J. of Educational Psychology*, 111, **3**, 389-400 (2019).
19. Henderson, A., The effect of timely feedback on goal-setting in students. *J. of Educational Psychology*, 110, **4**, 547-558 (2018).

BIOGRAPHIES

Baurzhan Yessingeldinov is a doctoral student at *M. Auezov* South Kazakhstan University, Shymkent, Kazakhstan. He has a Bachelor degree of Science in Mathematics and a Master's degree of Education in Mathematics from *S. Amanzholov* East Kazakhstan University, Ust-Kamenogorsk, Kazakhstan, awarded in 2012 and 2014, respectively. He is currently the Deputy Principal of the Centre of Pedagogical Measurements at the Autonomous Educational Organisation *Nazarbayev Intellectual Schools*.

Zamira Rakhymbayeva is the Deputy Director of the Centre of Pedagogical Measurements at the Autonomous Educational Organisation *Nazarbayev Intellectual Schools*. She is responsible for leading a major change in assessment services, incorporating advanced techniques, methods and technologies. She previously served as a project manager for the student selection system, overseeing all aspects of building the assessment system, including item development and report generation. Her extensive background also includes psychometric and statistical data analysis, using both CTT and IRT methods. Additionally, she has experience in co-ordinating statistical and psychometric analysis and developing IT tools for assessments.

Aikorkem Zhapparova has a Bachelor degree of Science in Mathematics and Master's Degree of Science in Educational Leadership from *Nazarbayev* University, Astana, Kazakhstan, awarded in 2016 and 2018, respectively. She was teaching statistics at the Spectrum International School, Astana, Kazakhstan, between 2016 and 2019. From 2019, she has been employed as an expert at the Statistics Department of the Centre of Pedagogical Measurements at the Autonomous Educational Organisation *Nazarbayev Intellectual Schools*.

Laila Tursynova graduated a Master of Educational Leadership programme from the Graduate School of Education at *Nazarbayev* University, Astana, Kazakhstan. Laila is currently the Director of the Centre of Pedagogical Measurements at the Autonomous Educational Organisation *Nazarbayev Intellectual Schools (NIS)*. Previously, she was the Head of the Research Department, NIS, from 2020, untill recently, where she oversaw implementing research-based approaches into teaching and learning policies of the NIS. Between 2014 and 2020, she was the Deputy Director of the External Summative Assessment and Monitoring of Students Achievements Department, administering all stages of the examination equivalent to A/AS level qualifications in collaboration with Cambridge Assessment International Education. From 2012 to 2014, she was a senior manager for trilingual education policy development. Between 2009 and 2012, she was also a teacher at *Nazarbayev* Intellectual School of Astana (IB DP).