



## Performance-Based Accountability: Examining Turkish Teachers' Perceptions Regarding the Implementation of Large-Scale Assessment

Fatma Cobanoglu\* 

Pamukkale University, TÜRKİYE

Received: May 31, 2024 • Revised: July 21, 2024 • Accepted: August 17, 2024

**Abstract:** This study, designed as a basic qualitative research, aims to evaluate the perspectives of Turkish teachers regarding the nationally implemented large-scale assessments in specific subjects. By employing purposive sampling methods, particularly maximum variation sampling, 14 teachers with different seniorities and branches from various high schools were included in the research. Data were collected using a semi-structured interview form prepared by the researcher. Through the application of descriptive analysis, it was found that teachers viewed these assessments more as a preparation for changes in the higher education transition system and as a tool for achieving standardization across the country rather than as an accountability policy instrument. Despite the low-stakes nature of this assessment, it was observed that teachers exhibited responses like those documented in the literature for high-stakes accountability. Additionally, it was found that teachers prepared students for the exams primarily through practice exercises and experienced anxiety before the exams due to the limitation of their autonomy and the comparison of their students and themselves. Furthermore, most participants believed that the assessment results were not effective in determining students' learning, primarily held the students accountable for the results, and considered such assessments as a source of stress. Based on the findings, recommendations were made for practitioners and researchers.

**Keywords:** *Accountability, national large-scale assessment, performance-based accountability, teacher stress, teacher autonomy.*

**To cite this article:** Cobanoglu, F. (2024). Performance-based accountability: Examining Turkish teachers' perceptions regarding the implementation of large-scale assessment. *European Journal of Educational Management*, 7(3), 155-171. <https://doi.org/10.12973/eujem.7.3.155>

### Introduction

Ensuring that society derives the maximum benefit from publicly funded education is a top priority for governments worldwide. Consequently, increasing concerns about the effectiveness and quality of education systems have driven countries to implement reforms and changes while laying the groundwork for holding schools and teachers accountable for quality. This process began with major education reforms such as the Education Reform Act (1988) in England or the well-known the No Child Left Behind Act (2001) in the United States, introducing new policy tools that changed the way central governments regulate education systems and hold local institutions and actors accountable for their operations and outcomes (Maroy & Voisin, 2017). These reforms have grown in popularity over time. The increasing relevance and authoritative role of the Organization for Economic Co-operation and Development (OECD) in global education governance has also contributed significantly to this outcome. The OECD's focus on the production of new quantitative data sources and its interest in approaching these data from a comparative perspective have amplified its impact on education (Parcerisa et al., 2021).

Accountability—the monitoring and use of student performance data to make judgments about the effectiveness of schools and teachers—is increasingly prevalent in school systems worldwide (Jerrim & Sims, 2022). Governments establish accountability hierarchies to hold educational institutions responsible for delivering high-quality education. Standardized and measurable learning outcomes are considered the most crucial indicators in this regard. Many policymakers and academics have more explicitly linked educational quality to student achievements; this is not due to ideological reasons but because learning outcomes are more concrete, comparable, and measurable than other, more context-sensitive indicators of educational quality (Verger & Parcerisa, 2018). In this context, there has been a significant increase in the use of large-scale assessments (LSAs) in recent years. More students and teachers are subject to the results

#### \* Correspondence:

Fatma Cobanoglu, Pamukkale University, Türkiye. ✉ [fcobanoglu@pau.edu.tr](mailto:fcobanoglu@pau.edu.tr)



of LSAs, and more countries are engaging with LSAs (Emler et al., 2019). As Popham (2000) stated, "It is genuinely difficult to find a large-scale educational assessment that isn't playing some sort of role in a local or regional accountability drama" (p. 283).

Nowadays, the use of large-scale assessment in education systems has become a highly controversial topic. Proponents of these assessments argue that they lead to significant gains in student achievement (Carnoy & Loeb, 2002) and aim to aid learning, assess individual success, and evaluate the quality of education (Hamilton, 2003). They also contend that high stakes standardized exams are strongly linked to improvements in average performance on international standardized assessments (Bergbauer et al., 2018; Yang et al., 2021). Furthermore, advocates suggest that performance-based accountability programs can reduce achievement disparities between students of high socioeconomic status and those of low socioeconomic status (Baidoo-Anu & Baidoo, 2024).

On the other side of the debate, critics argue that standardized tests increase the risk of educational failure, hold teachers accountable with unequal resources, narrow and distort the curriculum, are not suitable indicators of student achievement, and reinforce class and racial disparities. For instance, Berliner (2011) suggests that large-scale testing leads to curriculum narrowing, reducing the perceived talent of many students, restricting creative and enjoyable activities for teachers and students, and limiting critical thinking skills. Nichols and Berliner (2008) also argue that such testing can harm children's education. Many arguments against high-stakes achievement tests stem from opposition to policymakers' desire to quantify and objectify a system that critics see as subjective and unmeasurable, or from the belief that a single test provides a minimalist view of system outcomes (Burger & Krueger, 2003). From another perspective, Klenowski (2011) highlights the unhealthy competition among schools, teaching to the test, increased stress levels for children, parents, and teachers, and high costs as responses to high-stakes testing. Additionally, focusing solely on tested subjects and formats can diminish interest in more complex problem-solving tasks (Baker et al., 2010). Hargreaves (2020) argues that high-stakes tests lead to "teaching to the test, cultural bias, avoidance of innovation, dilemmas over whether to include highly vulnerable students in the testing process, and emotional distress among students and teachers." As Emler et al. (2019) conclude, "The combination of the power of LSAs in dictating education and the narrow scope of skills, knowledge, and human qualities they are able to or choose to measure are the source of their negative impact on education" (p. 284).

In this study, national large-scale assessments (NLSAs) implemented within the context of performance-based accountability were examined from the perspective of teachers, with a particular focus on the emotional impacts perceived by teachers.

## Literature Review

### *Accountability*

In recent years, accountability in education has become central to educational reforms. Growing concerns about the effectiveness and quality of many countries' educational systems have driven the push for reform and change. In this context, accountability has become both a 'means and an end.' Accountability is a complex concept with many meanings and various applications depending on the relevant social field or scientific discipline (such as political science, financial accounting, public administration, or education) (see Maroy & Voisin, 2017). In its simplest form, the term "accountability" describes a relationship in which one party—sometimes interpreted as an individual, sometimes as an institution—has an obligation to explain their actions or performance to another party (Gilbert, 2012).

In contemporary political and academic discourse, "accountability" often serves as a conceptual umbrella encompassing various other concepts such as transparency, equality, democracy, efficiency, responsiveness, responsibility, and integrity. (Bovens, 2007, p. 449) According to Bovens (2007), accountability is "a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgments, and the actor may face consequences" (p. 450). This definition can be used as an analytical tool structured around five operational questions that help understand the characteristics and instruments of accountability (Table 1).

Table 1. Operational Questions That Help Understand the Accountability

1. Who is accountable? / Who should be accountable?	<ul style="list-style-type: none"> <li>- Political institutions (government and government administration)</li> <li>- Public institutions (schools)</li> <li>- Individuals (politicians, senior officials, school principals, teachers...)</li> </ul>
2. To whom (which forum) should accountability be given?	<p>This question reveals different types of accountability:</p> <ul style="list-style-type: none"> <li>- Political accountability: The forum includes political actors or institutions like governments or voters.</li> <li>- Administrative or managerial accountability: The forum is the hierarchy of the public service organization or public administration.</li> <li>- Professional accountability: The forum is a professional organization or audit office.</li> <li>- Social (or market) accountability: The forum includes civil society actors, users of a service, customers, or interest groups (stakeholders).</li> <li>- Legal accountability: The forum consists of courts</li> </ul>
3. What is the nature of the relationship (or type of obligation) between the actor and the forum?	<ul style="list-style-type: none"> <li>- Vertical relationship: The forum has formal authority over the actor due to the hierarchical relationship between the actor and the forum.</li> <li>- Horizontal relationship: Accountability is given to stakeholders without a formal obligation; relationships are based on social or moral obligations.</li> <li>- Diagonal relationship: This is an intermediate form where accountability is given to an intermediary audit or oversight body that does not have direct hierarchical authority over the actor.</li> </ul>
4. What should the accounts be about?	- This can be based on the inputs, processes, outputs, or impacts of the actions taken.
5. On what basis should judgments be made?	These aspects can also be evaluated according to different criteria (e.g., equity, effectiveness, efficiency, propriety, transparency, democracy) and different methods.

Source: Bovens (2007, p. 451)

In education, accountability can be conceptualized and implemented in various ways (Gilbert, 2012), but fundamentally, it means holding key actors within the school system (teachers, principals, schools) responsible for student success (Brill et al., 2018). Generally, accountability is closely related to the concept of responsibility and indicates that teachers are accountable to their students and society. However, this term has a more complex definition because it can also mean “being held to account” and “giving an account”. The former implies a coercive relationship that exerts pressure on the accountable person, while the latter involves taking responsibility and explaining one's actions. Teachers in schools can experience both types of accountabilities in their daily work. For example, they may be held to account by school principals based on their performance, often tied to standardized test results. Simultaneously, they are expected to provide explanations to parents regarding pedagogy, curriculum, or assessments (Lingard et al., 2017). The form of accountability that is more intensively disseminated in educational administration is administrative and outcome-based, viewing teachers and schools as actors who must account to both public administration and society (Verger & Parcerisa, 2018).

However, recently, accountability in education has been understood in an increasingly narrow sense. Accountability now focuses on schools and teachers being held responsible for results and outcomes according to centrally determined goals (quality ratios, external exam performances) and standards (curriculum or assessment). These accounts are given based on various outcome indicators to the chain of command (district or state) or parents, depending on the school and student performance evaluation. In these “new” education accountability or “performance-based accountability” systems, student outcomes are the cornerstone of accountability systems (Maroy & Voisin, 2017). One reason why authorities worldwide have increased such scrutiny within the education system is the association of this monitoring with higher student performance and equal opportunities, supported by influential international organizations like the OECD (2011).

### *Performance-Based Accountability*

Performance-based accountability, the dominant form of accountability spread in the context of the Global Education Reform Movement (GERM), is a metaphorical concept indicating that most education reforms adopted worldwide respond to similar issues and priorities and follow very similar policy logic (Verger et al., 2019). This approach largely refers to the use of data obtained from large-scale student assessments as a mechanism or tool to hold school systems accountable (Popham, 1999). In recent years, there has been a significant increase in the use of large-scale assessments in education. More students and teachers have been subjected to the results of large-scale assessments (LSAs). More countries and regions have been engaged with LSAs (Emler et al., 2019).

National large-scale assessments (NLSAs) are not a new phenomenon in education systems. Initially, they were used for certifying students or diagnosing the problems and challenges faced by education systems. However, within the context of GERM, NLSAs have increasingly been used to monitor the implementation of more standardized curricula and to make schools, principals, and teachers more accountable (Verger et al., 2019). Large-scale national assessments are described

as “the main policy instrument in the hands of governments to ensure that school actors meet national targets and adhere to the teaching and learning standards defined in the national curriculum” and “in a context of increasing decentralization of education systems, NLSAs are a key policy technology to retain the State’s power and to ‘steer at a distance’ the complex network of providers and actors that configure the school system” (Verger et al., 2019, p.8).

The evidence provided by assessment systems can be used for many purposes (O’Neill, 2013). The primary uses in education are educational: the results of student assessments can help students, parents, and teachers judge what has been learned, take appropriate educational steps, and hold students accountable for their own learning or learning deficiencies. However, the same evidence can also be repurposed to hold teachers and schools accountable. Although the existing literature indicates that the primary purpose of large-scale tests is to support student achievement by monitoring those responsible for educating students (Decker & Bolt, 2008; Hanushek & Raymond, 2005; D. Klinger et al., 2008), unfortunately, large-scale tests have become a tool for accountability worldwide. Tests have shifted from being a tool for making important decisions about students to a tool for holding schools and education systems accountable (Baidoo-Anu & Baidoo, 2024; Brill et al., 2018; Chung & Chea, 2015; Kirkpatrick & Zang, 2011; Rahman et al., 2021; Verger et al., 2019). The secondary uses of assessment evidence may appear simple and convenient, but they can lead to many complexities and challenges, resulting in negative consequences. It is important to carefully consider and examine how the possibility of holding schools and teachers accountable for student performance measured by a particular assessment system can affect the actions of the preparatory school and teachers, and what unintended negative side effects may occur.

According to Lingard et al. (2017), who examine accountability from a pharmacological perspective, top-down forms of accountability can easily become toxic and can be exploited as (a) mechanisms for governing education systems and (b) parts of simplistic and populist education policies that excessively blame teachers for poor educational outcomes. Although it is a common belief that there is washback (i.e., the positive or negative effects of a test on teaching and learning), this should not be automatically assumed for any test or exam. Studies suggest that identifying the reflective effects of each test or exam (especially those with high stakes) requires specific research involving various stakeholders, such as teachers and students (Rahman et al., 2021).

In Turkey, one of the most significant factors influencing teachers’ instructional activities is NLSAs. It is stated that teachers design their teaching and learning environments according to the NLSA that students will take and determine the objectives and content with these exams in focus (Güler et al., 2019). In this context, there is very little accessible research aimed at determining teachers’ opinions about NLSAs (Atar & Büyüköztürk, 2017; Atila & Özeken, 2015; Güler et al., 2019; Kahraman, 2014; Sağır & Mart, 2017; Taşkın & Aksoy, 2018), and in these studies, teachers’ emotional experiences are mentioned only between the lines. Teachers feel compelled to conduct their teaching in a test-based manner due to the presence of national exams, and they feel obligated to prepare their students for these exams. This situation creates significant pressure on teachers (Özer Özkan & Acar Güvendir, 2018). Gök (2018), in his research examining the views of stakeholders on large-scale exams, concluded that the common aspect of the descriptions made was that the exam is generally associated with a negative situation, an anxiety-inducing object, or event. Bakırcı and Kırıcı (2018) also pointed out that teachers believe NLSAs increase students’ motivation towards the course and contribute to teacher-student-parent solidarity, but they also noted that these exams lead to the evaluation of school management and teachers, create a competitive environment and stress, and weaken social relationships. Çetin and Ünsal (2019) found that the psychological effects on teachers include creating pressure, making them stressed/tense, and affecting their motivation, while the social effects lead to discrimination, competition, and solidarity among teachers. Additionally, they highlighted that central exams impact the image of teachers; while they have positive effects such as contributing to teachers’ professional development, keeping up with developments in the field, updating their knowledge, becoming more proficient in solving questions, and providing opportunities for self-assessment, they also have negative effects such as making teachers robotic, being exhausting and wearing, and not providing opportunities for socialization.

In Turkey, existing research predominantly focuses on the implementation, outcomes, and effects of large-scale exams on students. This indicates that there is a need for a better understanding of teachers’ experiences and perspectives regarding the exam systems. Examining teachers’ views on the exam systems within the context of accountability can help in understanding the pressures and challenges teachers face and can contribute to designing more effective and supportive assessment policies. This research investigates teachers’ perceptions and opinions regarding the effects of large-scale exams implemented on specific subjects in high schools. In recent years, in addition to NLSAs used to select students for high schools and universities in Turkey, such exams have also begun to be conducted at the middle and high school levels. Although there is no information in the legal regulations prepared by the Ministry of National Education (MoNE) other than the way the exam is conducted, the processes experienced by Turkish teachers compared to their colleagues in other countries seem worthy of investigation. In this context, the following sub-problems are addressed: (a) What are the teachers’ opinions on the purpose of the assessment? (b) What did teachers feel before the implementation of the national exam? What kind of preparations were made in their schools and in their own classes? (c) What did teachers feel after the implementation? What kind of activities were conducted in their schools and in their own classes? (d) Who do they think is responsible for the results of the assessment? (e) According to the teachers’ views, does the administration of the exam create stress?

## Methodology

### Research Design

This research is designed using a basic qualitative research pattern. According to Merriam (2023), basic qualitative research is conducted to determine how participants interpret the subject and what meaning they attribute to their experiences.

### Sample and Data Collection

In this study, a working group was formed using a purposeful sampling method, specifically maximum variation sampling, from teachers working in high schools. Purposeful sampling involves "identifying and selecting individuals or groups of individuals that are especially knowledgeable about or experienced with a phenomenon of interest," while maximum variation sampling aims to "capture the diverse experiences of various stakeholders in different contexts" (Yağar & Dökme, 2018, p. 4). Participants were selected based on specific criteria to ensure maximum variation. These criteria included teaching the subjects where NLSA was implemented (9th grade, English, Turkish Language and Literature, and Mathematics), working in different types of schools (vocational high schools, academic high schools admitting students through centralized exams, and high schools admitting students based on their residential address), having different levels of professional experience, and being either male or female. Efforts were made to ensure diversity among the represented high schools. Vocational high schools offer specific vocational training. Academic high schools admitting students through centralized exams are generally focused on academic achievement. High schools admitting students based on their residential address enroll students according to their place of residence. This diversity enables the understanding of different educational contexts and the experiences of teachers within these contexts.

The process for selecting participants was as follows: First, volunteer teachers who were knowledgeable about the study's purpose and scope and had experience in these areas were identified. Then, among these teachers, those from various types of schools and with different professional experiences were selected to ensure maximum variation. Information about the study group is presented in Table 2.

Table 2. Information About the Study Group

		f
Gender	Female	8
	Male	6
Seniority	5-10 years	2
	21-25 years	3
	26-29 years	4
	30 years +	5
Education	Bachelor's	13
	Master's	1
Branch	Foreign language	4
	Mathematic	5
	Literature	5

In the research, a semi-structured interview form developed by the researcher was used as the data collection tool. During the development of this form, relevant literature was first reviewed, draft questions were prepared, and expert opinions were sought. Additionally, pilot interviews were conducted to ensure the clarity of the questions. The interview form includes a total of 10 questions, some of which are: "What do you think about your school's success in national exams?", "Who do you think are responsible for these results?", "Do you think these exams are a source of stress for teachers?". The interviews were conducted individually with the participants at designated times and places. With the participants' consent, the interviews were recorded. In the study, to maintain the confidentiality of the teachers included in the interviews, their names were replaced with codes such as T1, T2, and T3. Also, vocational high school is abbreviated as "Voc.", address-based school as "Add.", and academic high school as "acad."

### Analyzing of Data

In this research, the data obtained from face-to-face interviews with teachers were summarized and interpreted according to predetermined themes (Yıldırım & Şimşek, 2005), and thus, a "descriptive analysis" process was applied in the analyses. The researcher transcribed the interviews and organized them into a Word document. Each sub-problem was treated as a distinct theme during the analysis. In the coding phase, the researcher identified and coded significant segments of the respondents' answers within each theme, using words or sentences. This process allowed for a clear summary of the interview data and the creation of codes for each theme. This analysis was supported by direct quotations from the participants. No software was used during this process.

### *Validity and Reliability*

In this research, to ensure the validity of the data, the interview questions and the researcher's judgments were presented based on the literature (Merriam, 2023; Yıldırım & Şimşek, 2005). Additionally, to ensure external validity, which allows the research results to be generalized, detailed information about the research process, stages, analysis, and results was provided, and all testable details were shared with the readers.

Various strategies have been employed to ensure the reliability of the data (Merriam, 2023; Yıldırım & Şimşek, 2005). A subset of the research data was independently coded by another expert in the field, and the results were compared and evaluated. Discrepancies were discussed and resolved to reach a consensus. Additionally, the research process and findings were discussed with researchers who were not part of the study but were knowledgeable about the topic. This approach helped to avoid biases and enabled in-depth analyses. Furthermore, to allow readers to evaluate the validity and transferability of the findings, detailed descriptions of the research context, participants, and processes were provided. This comprehensive explanation of the study's context helps readers to adapt the findings to their own contexts.

### *Researcher's Role and Reflexivity*

The researcher has over twenty years of experience in higher education in the field of educational management, as well as experience in human resources management and organizational behavior in educational organizations, and research methods. This background has supported efforts to maintain objectivity during the data collection and analysis process.

## **Findings/Results**

Based on the interviews conducted with teachers, the findings have been analyzed under the themes of NLSA's "purpose," "preparations" and "feelings" beforehand; "the activities" and "feelings" afterwards; and whether there were "the responsible" and "sources of stress" regarding the results obtained. The findings are presented in tables in the following paragraphs.

### *Thema 1. The Purpose of the NLSA*

The codes under this theme are presented in Table 3.

*Table 3. The Purpose of the NLSA*

<b>Theme</b>	<b>Code</b>	<b>f</b>
Purpose	Change in system	7
	Determining levels	7
	Standardized assessment	3
	Insecurity / Objective measurement	4
	The implementation of the curriculum	5
	Determining student success	2
	Determining teacher competence	2
	Unknown	2

According to the views of some participants, the conducted NLSA is considered a preparation for changing the current system of student selection for higher education ( $f=7$ ). Some of the participants' opinions on this matter are as follows:

*"I think it is to abolish the university entrance exam. That's what I think. It seems more logical to me for students to enter universities based on their average rather than a single exam (T 6)."*

*"I actually think they are trying to move away from the exam system (T 12)."*

*"Probably in the future, the transition to higher education will be through school exams (T 13)."*

It has been observed that many teachers believe that exams are conducted to determine the students' level. Additionally, they think that this exam is conducted to implement a standardized assessment across the entire country. However, some participants mentioned that the central administration conducted this assessment because they had not trusted them in terms of measurement and evaluation and to ensure an objective evaluation. For instance:

*"What I'm saying is that all teachers should teach the same thing, implement the curriculum, and then there should be a common exam. I think that's the Ministry's goal (T 1, English, Voc.)"*

*"I think such a practice is implemented to ensure uniformity in teaching the topic across all schools, aiming to ensure that the same subject matter is covered in all schools for the sake of coherence (T 2, Literature, Voc.)"*

*"There is the idea of ensuring uniformity in practice (T 13, Literature, Add.)."*

*"I mean, I think it is to achieve unity and solidarity or a commonality in those achievements. But ultimately, we have an annual plan. Every teacher gets approval based on the annual plan, and they conduct their lessons according to that approval. They don't teach randomly. We have a textbook, a specific textbook. Despite this, I think they believe that unity and solidarity are not being achieved (T 3, Voc., Literature)."*

*"Teachers give certain grades in every school, they teach something, and as a result, they conduct an assessment. But since they think there isn't a standardized assessment, I believe they want to ensure a common assessment (T 11, Literature, Acad.)."*

*"It could be to measure the education levels in schools. It could be to understand whether students grasp the achievements better. It could be to measure the class levels. It could be to assess whether the teacher is competent or not (T 10, Math, Acad.)."*

One of the teachers who was confused about the purpose of this exam also expressed his views as follows:

*"There is an intention, okay, we're doing a standardized exam... But it has a few drawbacks. Firstly, the first term exam was multiple-choice. People objected to this. Then it was changed to an open-ended exam in the second term. Therefore, the Ministry needs to decide. What do I want? They need to decide whether to use multiple-choice or open-ended questions based on their goal... Then the teaching community said, 'If you want open-ended questions from us, why are you giving a test?' They objected. The second exam came as open-ended. Well, is this something that can change based on feedback? Or does it have a specific goal, and are they doing something towards that? They need to decide on this. I see an ambiguous situation here." (T 11, Literature, Acad.)*

#### *Theme 2. "Preparations Made by Teachers and Their Feelings" Before Evaluation*

The codes for the views of teachers regarding the preparations made in their schools and classrooms before the NLSA, as well as their feelings, are provided in Table 4 and Table 5.

*Table 4. Preparations Made by Teachers for Evaluation*

<b>Theme</b>	<b>Code</b>	<b>f</b>
Preparations	Practicing	9
	The implementation of the curriculum	6
	Textbook of Ministry of Education	5
	Assignment	5
	Material support	4
	Speaking skills	3
	Guidance	4
	Subject review	2
	Departmental studies	2
	Ensuring class participation	2
	Administrative support	1

It was observed that teachers mostly prepared their students for the NLSA by practicing about the subject in their lessons (f=9) and implementing the curriculum according to the annual planning (f=6). In addition to the teachers who stated that they used the Ministry's textbook for instruction since the exam was administered by the Ministry, some teachers assigned homework (f=5) and used different materials (f=4). Furthermore, especially foreign language and literature teachers mentioned that they also conducted activities related to speaking and comprehension skills tested in their respective exams. Teacher opinions on the subject can be exemplified as follows:

*"Now, after covering the regular curriculum, we did a simulation exam with the students. That is, we prepared the exam questions ourselves (T 11, Literature, Acad.)."*

*"Since you don't know what will come up, we prepared by assigning students as much as possible and occasionally making revisions within the class (T 9, Math, Add.)."*

*"Within the framework of the annual plan, using the experience of years, I already predict questions that could come more or less accurately. When practicing on the board, if there is missing information in the questions I predicted, I complete it before the exam (T 2, Literature, Voc.)."*

In addition to all these instructional preparations, it has been observed that some teachers also guided and motivated their students (f=4) regarding exam rules, doing more reading, and studying regularly. For example:

*"In preparation for these exams, problem solving, and assignments are made according to the questions that may be asked... For this, I prioritize speaking and listening activities in my classes. Instead of teaching*

*grammar, I give assignments that will make the students more active in class, such as giving presentations and discussing a topic (T 12, F. Language, Add.)."*

*Table 5. Teachers' Feelings Before Evaluation*

Theme	Sub-themes	f	Codes	f
Feelings	Anxiety/Stress	5	Evaluation of students	4
			Time pressure	3
			Comparison	3
			Exam questions	2
			School image	1
	Bad	5	Decreased autonomy	8
			Favoritism	3
			Untrusted teacher	2
			Implementation-related	1
	Happy/Stress-free	6	Trust in school/school image	3
			Sense of duty	2
			Evaluation of the teacher	1
	Excitement	1		

Teachers who reported experiencing stress/anxiety ( $f=5$ ) before the exam cited the following reasons: their students being evaluated through this exam, the concern about not being able to implement the curriculum on time, the possibility of being compared with others based on exam results, the uncertainty of the exam questions, and the potential damage to their school's image. Teachers who said they felt bad ( $f=5$ ) indicated that the reasons for their feelings were the reduction of their autonomy, the lack of trust in them, and concerns about potential favoritism during the implementation. Examples include the following:

*"The idea of being compared to all schools puts one under stress. After all, you're in a school where a certain level of success is expected (T11, Literature, Acad.)."*

*"I think this is to mechanize the teacher. It's putting them into a mold, not trusting them. Because ultimately, if we're given a curriculum, when I create written exam questions, I'm not going outside the curriculum, I'm not going outside the textbook, so I should have a say (T3, Literature, Voc.)."*

*"Anxiety...Did any unexpected questions come up, or did we go beyond the curriculum? Or was there a question on a point we might have overlooked?" (T9, Math, Add.)."*

*"I feel like we don't really get to use our initiative much. After all, we're the ones teaching the subject, we're the ones who know the reality of the students, we're the ones who know the reality of our school... I feel bad. I feel like my opinion isn't consulted or I'm excluded (T2, Literature, Voc.)."*

*"I worry if I won't be able to cover the topics, my students will suffer. Also, I feel anxious if my students fail (T14, Math, Add.)."*

Although most teachers reported experiencing these negative feelings, some also mentioned feeling happy due to their confidence in their school's success and the satisfaction of having fulfilled their responsibilities. For example:

*"We don't know anything, but this is one thing for sure, because I'm very comfortable knowing that I've done my job properly (T 7, Math, Add.)."*

### *Theme 3. "The Activities Teachers Conducted and Their Feelings" After Evaluation*

The codes for the views of teachers regarding the activities they conducted in their classrooms and their feelings after the NLSA are provided in Table 6 and Table 7



Table 6. The Activities Teachers Conducted in Their Classrooms After The NLSA

Theme	Codes	f
Afterwards	Questioning mistakes	8
	Practicing	7
	Guidance	6
	Individual support	5
	Topic review	5
	Exam analysis	3
	Lack of individual support	2
	Assignments	2
	The implementation of the curriculum	2
	Parent meetings	1

After the NLSA, it was observed that some teachers discussed with their students which questions they could not solve (f=8), conducted exercises related to these subjects (f=7), provided guidance for future exams (f=6), gave individual attention to students who needed it (f=5), and re-taught the concepts that were not fully understood (f=5). A few other teachers mentioned that they continued to follow the routine teaching program as they did before the evaluation (f=2), assigned homework to address deficiencies (f=2), and held meetings with parents (f=1). Some teachers' views on the subject are as follows:

*"I ask about the topics my students didn't understand or why they couldn't complete certain tasks. If necessary, we try to address these gaps with small explanations or remedial actions (T5, Literature, Add.)".*

*"By solving examples, assigning homework, and reviewing the material in class, I try to bring my students who have gaps in their knowledge to a certain level. I tell them, 'I understand that you don't know this topic, or you have a deficiency in this area. Let's address this deficiency under my guidance so that you won't face this fear and anxiety in future educational stages (T2, Literature, Voc.)".*

*"I conduct exam analysis each time, publishing the questions and going over them. I explain what is expected from them, teaching them how to develop appropriate methods and techniques for the next exam (T13, Literature, Add.)".*

*"Announcing the results in the class leads to a satisfying response from the students. However, I remind them that our target is at a higher level and motivate them to continue working (T15, F. Language, Acad.)".*

Table 7. The Teachers' Feelings After The NLSA

Theme	Sub-themes	Codes	f
Feelings	Happy/Relieved	9 Student success	4
		Sense of duty	2
	Sad / Unhappy	3 Unrealistic results - easy questions	12
		Discipline problems	5
		High expectations	2
		Fatigue from change	2
		Poor results	1
		Favoritism	1
		Comparison with other teachers	1
	Unaffected	2 Expected results	2

When examining the teachers' responses, it is evident that most of them felt happy and relieved (f=9) due to their students' success after the NLSA. For instance,

*"There isn't a huge gap or difference between the classes. Seeing this, we think we did our job well. We felt good (T 5, Literature, Add.)".*

On the other hand, a few teachers expressed feeling sad and unhappy (f=3). Although the teachers were pleased with their students' performance, they noted that the exam questions were too easy and did not reflect true success (f=12) and some discipline problems occurred:

*"I can't be happy about the success because the questions were easy. I can't see it as my success or failure (T 14, Math, Add.)".*

*"It would be better if we measured all schools separately for the purpose of assessment, such as science high schools, Anatolian high schools, and vocational high schools. Because it tried to take the middle ground. It asked questions according to the Anatolian High School group. But this time, Science High Schools, of course, had a huge advantage here. And this automatically pushes the children towards laziness. They're not studying (T 10, Math, Acad.)"*

*"...I believe that a teacher who does not conduct assessment and evaluation does not efficiently carry out the teaching process. The strangest thing is asking questions that are different from the shared achievements before the exam. When teachers say, 'you are only responsible for these' and different questions come, I think that the relationships between the teacher and the student are also damaged (T 13, Literature, Add.)"*

Additionally, a few teachers (f=2) mentioned that the results did not affect them as their students performed as expected:

*"I don't feel much. The results didn't surprise me. I didn't think the questions were seriously evaluative (T2, Literature, Voc.)"*

*"When the results are announced, I only feel a sense of satisfaction regarding the success. The exam results don't significantly affect us (T15, F. Language, Acad.)"*

#### *Theme 4. The Responsible for The Results*

According to the teachers' views, the results of the evaluation have both direct and indirect responsible parties (Table 8):

*Table 8. Responsible Parties of Evaluation Results*

<b>Theme</b>	<b>Sub-themes</b>	<b>Codes</b>	<b>f</b>
Responsible	Direct	Student	12
		System	5
		Teacher	4
		Teacher-student-family	4
		Previous teachers	3
		Administrators	2
		All stakeholders in education	1
		Curriculum-teacher-student-family	1
	Indirect	Social Economic Status	8
		Supplementary material	2
		Crowded classroom	1
		Teacher experience	1
		Extra curriculum learning	1

As can be seen, teachers believe that much of the responsibility for the results lies with the students (f=12). Additionally, it is observed that some believe the teachers (f=4) is the architect of these results, while others argue that the responsibility lies with the education system (f=5):

*"Teachers provide education to students, and the quality of the student helps ensure they get high marks (T 8, Math, Add.)"*

*"Because the incoming student has finished at least 7-8 books for the high school entrance exam. They are like a runner ready to run (T 10, Math, Acad.)"*

*"A large part of the responsibility lies with the system. Because, ultimately, the teacher conveys their knowledge, experience, and education in the classroom. The system not being well-established and constantly changing is where the responsibility lies, in my opinion. So, I blame the system (T 3, Literature, Voc.)"*

*"Actually, ...the system, kids who should be in a Vocational High School academically are in our school instead of an Anatolian High School (T 5, Literature, Add.)"*

*"Because our school's success is quite high, the ones responsible for the exam results are the teachers in our school who do their job properly. (T 15, F. Language, Acad.)"*

Some teachers emphasized that the educational process is a team effort, rather than solely blaming the teacher or the student. For instance:

*"Curriculum, student, teacher, and parent. Well, if the teacher doesn't teach, the student can't succeed. If the student doesn't work, no matter how much the teacher tries, they can't succeed. Therefore, I believe all four components are necessary (T 11, Literature, Acad.)"*

*"It's like a three-legged stool. Parent, teacher, and student, in my opinion. The teacher alone isn't to blame. The student alone isn't to blame, and the parent alone isn't to blame either. ... The parent is also unaware. They don't know, they're not aware. For example, how can they study English with their son or daughter? That's why I think of the three legs of the stool. If the student is willing, the teacher is willing, but the parent isn't, that stool leg has always been classic, or if the student isn't ready, the teacher is ready, and the parents are ready to get private lessons, but the child isn't motivated. Or my teacher friend isn't willing... (T 1, F. Language, Voc.)"*

*"I can't say this success belongs to us. This success is due to the teachers from middle school, even elementary school. Because our subject requires continuity. If the child doesn't come equipped from the earlier grade, they face difficulties here. So, this success is entirely the teachers' and the child's individual success. If they failed, it is also their own failure. (T 6, F. Language, Add.)"*

Some teachers mentioned that, indirectly, the students' socioeconomic status (f=8), the ministry's restriction on the use of supplementary materials (f=2), and crowded classrooms (f=1) also have an impact on the assessment results:

*"Of course, we are sad. Our goal is to increase success, but most of our students work in the afternoons or on weekends. As I mentioned before, due to their poor financial situations, they don't have the chance to review their lessons at home. They don't listen to foreign songs or watch foreign TV series. If they did, speaking from my own field, it would be much better. For example, I have some technical students who listen to foreign songs and share them with me, and they watch TV series. Their English is excellent. They learn vocabulary from those sources." (T 1, F. Language, Voc.)"*

Table 9. The Responsibility of Teachers

Theme	Codes	f
Responsibility of the Teacher	Taking Responsibility	6
	Feeling Partially Responsible	3
	Not Feeling Responsible	5

When teachers were asked whether they felt responsible for the evaluation results, some indicated that they felt responsible (f=6), while others reported feeling no responsibility at all (f=5) or only partially feeling responsible (f=3):

*"Responsibility? I want to take it. But, you know, when I compare the results and if a student has a low grade, I don't feel entirely at fault. Because we're presenting a system that's being imposed on the students. So, I don't feel like the success or failure of a student is 100% on me. (T 3, Literature, Voc.)"*

*"In an uncertain environment, honestly, there isn't much of a sense of taking responsibility. Just imagine, what was promised to be asked in the exam wasn't asked, you're left embarrassed in front of the students, and you don't even know what will happen next time. (T 13, Literature, Add.)"*

*"We certainly supported them, but I think the most important thing is the student's own effort. (T 10, Math, Acad.)."*

*"Absolutely, sir. Because those kids are our mirror (T 5, Literature, Add.)."*

#### Theme 5. The Source of Stress

The sub-themes and codes that emerged when teachers were asked whether this national assessment was a source of stress are presented in Table 10.

Table 10. The Sources of The Stress

Theme	Sub-themes	f	Codes	f
Source of stress	Yes	13	Decreased autonomy	7
			Uncertainty/ Change	5
			Exam process	4
			Time pressure	2
			Teacher inadequacy	1
			School type-dependent	1
	No	5	Exam-oriented school	2
			Conscious teacher	1
			Familiar content	1

It is observed that many teachers (f=13) agree that the assessment generally creates stress. The reasons for this stress are highlighted as decreased autonomy (f=7), uncertainty and changes in implementation (f=5), and the examination process itself (f=4).

*"...The state has really restricted us a lot. My biggest issue is this. They say, 'Here, I'm sending you a book, and you can't go beyond this book.' They've even restricted our exams. They say, 'You will ask these specific questions (T 6, F. Language, Add.).'"*

*"I think these exams can significantly impact teachers' stress levels, but this depends on the type of school. In schools with high-achieving students, I don't think the stress level is as high for teachers (T 15, F. Language, Acad.)."*

*"I want to have a say in the students I train. I want to be in a more respected position professionally. I want to have more of a say, and for my classroom to be left to me (T 3, Literature, Voc.)"*

*"We go quickly due to the anxiety of not being able to cover the topics in time. This prevents us from processing the topics thoroughly. (T 14, Math, Add.)"*

*"Because it's a practice outside the system we've been accustomed to for years, and teachers are inadequate in instructing students on this matter, it causes stress. (T 12, F. Language, Add.)"*

On the other hand, some teachers mentioned that they do not experience stress because they are already accustomed to the exam system, they fulfill their duties regularly, and the exam content is predetermined. For instance,

*"Our school has always been an exam-oriented school. Both students and teachers are very aware of this. We didn't have much stress because we know the quality of our students, our own quality, and the quality of the books we've solved. Because we were prepared, we didn't experience much stress (T 10, Math, Acad.)"*

### Discussion

The purpose of assessment and evaluation in education is to make judgments about what students know or can do and, increasingly, how schools perform (D. A. Klinger & Rogers, 2011). Therefore, large-scale assessments have taken on a primary role in educational accountability because they provide a common and seemingly fair criterion for tracking student achievement over time and making comparisons. They are relatively efficient and, most importantly, yield visible results (Linn, 2000). In Turkey, in addition to the large-scale assessments currently implemented to select students, different assessments have also been conducted on a subject basis. The perspectives of teachers regarding this new assessment, which is new to teachers, have been examined in this research. The results obtained are provided below.

First, it should be noted that although the participants work in different fields and schools, there are no significant differences in their opinions because they are subject to similar rules and procedures. Although teachers seem to have mixed feelings about the purpose of the administered NLSA, it has generally been interpreted as a sign of an impending change in the system for selecting students for higher education. Additionally, it has been perceived that the aim of the national exam is to determine the levels of student achievement nationally or to ensure coordination in the implementation of the curriculum, while some believe that the government administers the exam itself due to a lack of trust in teachers. However, there is not much confidence among teachers that the exam may have been conducted to evaluate them. The main reason for this confusion could be attributed to the Ministry's failure to provide a detailed explanation of the purpose of the exam. It is evident that the concern and pressure for quality in education systems compel educational organizations to work harder, produce more, keep up with innovations, and implement them. This situation necessitates the restructuring, organization, supervision, and replanning of policies and processes, which can create stress for employees due to uncertainties (Gümüştekin & Öztemiz, 2005). Therefore, not knowing the purpose of the exam or having concerns about it is a negative situation for teachers. Looking at this result from another perspective, LSAs have been used for a variety of purposes. Chief among them is three: accountability, selection, and comparison (Emler et al., 2019). While teachers agree that NLSAs are conducted for the purpose of comparison and preparation for future selection exams, they do not seem to attribute the purpose of holding education actors accountable for their actions and outcomes.

Most teachers were found to feel bad, anxious, and stressed before the exam. The factors causing these negative feelings include the evaluation of their students, the comparison with other schools and teachers, and especially the perceived decrease in their autonomy. Data-driven accountability inherently requires closer monitoring of teachers' performance using specific performance standards. If the established standards are excessively high, monitoring can increase the fear of not reaching the necessary standards among employees or lead to pressure on employees to outperform their peers. Managers may see this kind of work pressure as a desired effect of accountability because it creates pressure on teachers to raise performance standards (Jerrim & Sims, 2022). However, as seen in this study, such work pressure can also lead to stress and anxiety. Another reason for the stress experienced by teachers is their perception of reduced autonomy. Autonomy is "direction over one's own decisions and actions" (Worth & Van den Brande, 2020, p. 6) and data-driven accountability can lead to a loss (Jerrim & Sims, 2022). Studies suggest that performance-based tools tend to limit teachers' decision-making powers – that is, teachers' decisions and actions are increasingly conditioned by external evaluations, performance measurements, and related policies (Parcerisa et al., 2022). School leaders and teachers may not feel free to teach what they consider important; instead, they may have to focus on measured outcomes. Moreover, they may feel compelled to teach students in a certain way (i.e., a method that they or the principal believes will maximize

student performance). The reality is that greater teacher autonomy is associated with lower job-related stress and higher teacher job satisfaction (Worth & Van den Brande, 2020, p. 14). As a result, in the culture of performative accountability in twenty-first century education, efficiency is regarded as a "good thing" regardless of its cost to individuals (such as increased pressure, loss of autonomy, surveillance, and evaluation, limited participation in decision-making processes, and lack of personal knowledge) (Perryman & Calvert, 2020). On the other hand, it was found that teachers who reported not experiencing assessment-related stress also trusted in their school's success and themselves.

The participants were found to prepare their classes for the NLSA mostly. Teachers are seen to engage in exercises, assignments, and supplementary materials focusing on the topics likely to appear in the exam. As mentioned earlier, no information was provided about the purpose and outcomes of the government-conducted exam; in other words, this administered exam can be considered a low-stakes test. Nevertheless, as seen in high-stakes assessments (Berliner, 2011; Klenowski, 2011; D. A. Klinger & Rogers, 2011; Nichols & Berliner, 2008), it can be said that teachers conveyed to their students during exam preparations that the primary goal of learning is to achieve good results. LSAs distort education by distorting instruction, resulting in the "phenomenon of teaching to the test" (Abu-Alhija, 2007; Emler et al., 2019). Teaching to the test refers to "instruction that is solely focused on and often limited to what is to be included in the test" and "spending class time teaching students test-taking skills" (p. 285). These tests typically lead teachers to teach the test and teach towards the test. The former involves gathering ideas about what the test questions will be and teaching students the necessary responses accordingly. Such practices cannot foster learning. The latter, teaching towards the test, encourages identifying the concepts to be assessed and teaching these concepts to students (Rahman et al., 2021). Hargreaves (2020) has also argued that teaching to the test leads to "teaching the test and emotional discomfort among students and teachers." Similarly, existing literature shows that high-stakes tests significantly reduce the time spent on vital creative skills such as fundamental learning and self-monitoring skills, organizing knowledge, problem-solving, etc. (Ananda & Ashadi, 2021; Atila & Özeken, 2015; Berliner, 2011; Gonzalez et al., 2017; Klenowski, 2011). Darling-Hammond et al. (1995), Abu-Alhija (2007), and Erskine (2014) have also stated, as in the current study, that teachers narrow the curriculum, shorten the instructional time in favor of test preparation activities, and use more test-like instructional practices to prepare for large-scale assessments. Baker et al. (2010) draw attention to similar results, suggesting that focusing excessively on basic math and reading scores can lead to curriculum narrowing and simplification, and focusing only on tested topics and formats may reduce interest in more complex problem-solving tasks.

Teachers, who reported conducting evaluations of the exams in their classes post-examination, have also engaged in test-focused activities and guidance. It has been concluded that they experienced a sense of relief after the evaluation, following the anxiety and stress they experienced before. However, they also expressed sadness about the perceived lack of validity and reliability of the exams, as they did not find the results realistic or discriminatory. As Abu-Alhija stated, "test scores may go up, but learning doesn't change" (p. 56). According to Emler et al. (2019), the most damaging side effect of large-scale assessments (LSAs) is the distortion of education: LSAs distort the purpose of education by misleading the public to believe that performance on LSAs accurately reflects the quality of education, albeit not necessarily intentionally. Concerns have also been raised in the literature about increasing test scores not being indicative of higher levels of education (D. A. Klinger & Rogers, 2011). Ravitch (2011) addressed these concerns by stating that curriculum and instruction are far more important than accountability and testing, arguing that accountability and testing should be tools, not goals, in education, but they have currently become goals in themselves. Contrary to the "negative" findings, large-scale assessments can positively influence teaching and learning. Educators increasingly use the results of large-scale assessments, along with school and class data, to review teaching practices with an effort to enhance students' learning (Abu-Alhija, 2007; D. A. Klinger & Rogers, 2011). Indeed, the current study also found that teachers provide guidance to facilitate students' learning, review teaching methods, and individually assist students in need.

One of the most striking results of the study is that teachers mostly pointed to students as the ones responsible for the results achieved in the exam. Additionally, although to a lesser extent, they also saw their families and themselves, the current system, previous teachers, and administrators as responsible. Moreover, it has been concluded that teachers indirectly believed that children's socioeconomic status also influenced these results. Furthermore, while some teachers felt responsible for the outcomes, others did not feel responsible at all. A similar situation was also reported by Berryhill et al. (2009). According to the researchers, concerning self-efficacy, although teachers had little concern that student achievement could be affected by personal factors such as their own pedagogical abilities, they believed that situational and student-related factors (e.g., home conditions) played a more significant role. In the context of accountability policies, competence has gained a new dimension since teacher performance is indirectly evaluated not by their direct performance but by their students' test scores. Teachers' insufficient self-efficacy might manifest as a way for them to alleviate stress by avoiding taking responsibility for student outcomes. Additionally, considering that teachers may partially be justified in their assessments of their responsibilities by considering that school outcomes in standard tests are largely influenced by the student's socioeconomic status (Berryhill et al., 2009). Holding only teachers responsible for students' performance is inherently problematic. A range of factors has been found to affect student learning outcomes (Baker et al., 2010): the effects of other teachers (both previous and current teachers, special educators or instructional specialists teaching students), the effects of school conditions (quality of curriculum materials, specialist or

remedial support, class size, and other factors affecting learning), out-of-school learning experiences (school attendance and learning experiences at home, among friends, in museums and libraries, in summer programs, online, and in community interactions), well-educated and supportive parents who help their children with homework and provide them with various advantages, family resources, student health, family relocation status, and peer influence. Student achievement cannot be attributed solely to any single teacher.

Essentially, it was found that the NLSA implemented in the current study was almost unanimously considered a direct source of stress for nearly all teachers. These results were also reported similarly by Gonzalez et al. (2017). According to Stecher et al. (2003), accountability is "rewards or punishes schools and school staff on the basis of student scores on achievement tests" (p. 3). While this approach aims to improve the quality of education, it can create stress and pressure on teachers and school staff because their performance is directly evaluated based on student test results. The term "accountability" is perceived as a form of pressure on schools and teachers (Gilbert, 2012). Berryhill et al. (2009) aimed to explore the effects of accountability policies on the well-being of elementary school teachers in their study and reported a range of negative outcomes such as lack of policy support, role conflicts, and low self-efficacy leading to emotional exhaustion among teachers. According to Baker et al. (2010), accountability pressures are particularly associated with higher rates of turnover and decreased morale among teachers in high-need schools. Jerrim and Sims (2022) reported strong evidence that stress "emotionally spreads" among colleagues in schools: if someone experiences stress, other teachers are more likely to feel stressed due to the responsibility.

### Conclusion

The study, which evaluates the NLSA on a subject-specific basis from the perspective of Turkish teachers, concludes that teachers perceive this evaluation not as a tool for accountability policies but rather as preparation for changes in the transition system to higher education and standardization across the country. Although this evaluation is considered low-stake, teachers exhibit responses akin to those associated with high-stake accountability, as found in the literature. It was observed that teachers prepare students for the assessment primarily through practice and experience anxiety before the exam due to the limitation of their autonomy and the prospect of comparison with both students and themselves. Furthermore, most participants believed that assessment results do not effectively determine students' learning, hold students accountable for assessment outcomes, and perceive such evaluations as a source of stress for teachers.

### Recommendations

The results of the current research indicate that accountability policies may negatively impact teachers' perceptions of autonomy, job satisfaction, and mental health. To address these issues, the findings suggest that the design and implementation of more effective and supportive assessment policies could be guided as follows: (a) Allowing Flexibility for Teachers: Providing teachers with more flexibility to meet assessment standards enables them to use their professional judgment and tailor teaching methods to students' needs. Therefore, it would be beneficial to provide teachers with the necessary leeway. (b) Encouraging Collaboration: To mitigate the negative effects of accountability pressures, collaboration can be encouraged by sharing best practices among teachers and schools. (c) Providing Support Systems: Considering that such evaluations may cause stress among teachers, it is important to provide necessary support. Resources and guidance should be provided to teachers to support them during this process. Offering support systems such as professional development opportunities and a more supportive school culture will be beneficial in reducing teachers' stress and burnout. (d) Aligning with Contextual Appropriateness: It is important to ensure that assessment policies focus not only on student outcomes but also on the conditions that support effective teaching and learning, thereby aligning with contextual appropriateness. These recommendations could help create assessment policies that have a more positive impact on teachers and support both teacher and student well-being.

Although there are limitations, based on the results, the following recommendations also can be made: (a) Sharing objectives and outcomes: It is important to share the objectives and outcomes of the evaluation conducted by the ministry with the teachers. This can reduce the sense of uncertainty among teachers and make such practices more acceptable. (b) Using Different Assessment Methods: Especially in cases where significant effects such as program narrowing are thought to occur, it is recommended to support comprehensive assessments with different measurement methods. This approach may be more effective in achieving program objectives and can prevent teachers from feeling a decrease in their autonomy. (c) Involving teachers in decision-making: It is important to consider the views and suggestions of teachers regarding the implementation of the evaluation process. This can contribute to making the evaluation process fairer and more effective.

### Limitations

Although the data obtained from a group of teachers who participated in the NLSA deeply informed this research, the non-generalizability of the results remains a significant limitation. In the future, conducting similar studies using quantitative methods in high schools and middle schools where standardized exams are administered, and across different geographical regions, is expected to significantly contribute to the understanding of the issue. Additionally, examining the perspectives of administrators, who are also key actors in performative accountability, would be beneficial both in contributing to the literature and in addressing the issues related to implementation.

## References

- Abu-Alhija, F. N. (2007). Large-scale testing: Benefits and pitfalls. *Studies in educational evaluation*, 33(1), 50-68. <https://doi.org/10.1016/j.stueduc.2007.01.005>
- Ananda, E., & Ashadi, A. (2021). High-stakes testing and English teachers' role as materials developers: Insights from vocational high schools. *Studies in English Language and Education*, 8(1), 115-130. <https://doi.org/10.24815/siele.v8i1.17518>
- Atar, H. Y., & Büyüköztürk, Ş. (2017). Temel eğitimden ortaöğretime geçiş: Uygulamalar ve sonuçları [Transition from primary to secondary education: Practices and consequences]. *Eğitime Bakış Dergisi*, (40), 15-25. <https://bit.ly/3YQmBnC>
- Atila, M. E., & Özeken, Ö. F. (2015). Temel eğitimden ortaöğretime geçiş sınavı: fen bilimleri öğretmenleri ne düşünüyor? [Examination for Transition from Primary Education to Secondary Education (TEOG): What Do Science Teachers Think About It?]. *Ondokuz Mayıs University Journal of Education Faculty*, 34(1), 124-140. <https://dergipark.org.tr/tr/download/article-file/188142>
- Baidoo-Anu, D., & Baidoo, I. E. (2024). Performance-based accountability: exploring Ghanaian teachers perception of the influence of large-scale testing on teaching and learning. *Education Inquiry*, 15(3), 333-350. <https://doi.org/10.1080/20004508.2022.2110673>
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010, August 27). *Problems with the Use of Student Test Scores to Evaluate Teachers*. Economic Policy Institute. <https://www.epi.org/publication/bp278/>
- Bakırcı, H., & Kırıcı, M. G. (2018). Temel eğitimden ortaöğretime geçiş sınavına ve bu sınavın kaldırılmasına yönelik fen bilimleri öğretmenlerinin görüşleri [Science teachers' views on the removal of the transition from primary to secondary education exam]. *Van Yüzüncü Yıl University Journal of Education/Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 15(1), 383-416. <https://doi.org/10.23891/efdyyu.2018.73>
- Bergbauer, A. B., Hanushek, E., & Woessmann, L. (2018). *Testing with accountability improves student achievement*. VOXEU CEPR. <https://bit.ly/3ytluj0>
- Berliner, D. (2011). Rational responses to high stakes testing: the case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education*, 41(3), 287-302. <https://doi.org/10.1080/0305764X.2011.607151>
- Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The effects of education accountability on teachers: are policies too-stress provoking for their own good? *International Journal of Education Policy and Leadership*, 4(5), 1-14. <https://doi.org/10.22230/ijep.2009v4n5a99>
- Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4), 447-468. <https://doi.org/10.1111/j.1468-0386.2007.00378.x>
- Brill, F., Grayson, H., Kuhn, L., & O'Donnell, S. (2018). *What impact does accountability have on curriculum, standards and engagement in education? A literature review*. National Foundation for Educational Research (NFER).
- Burger, J. M., & Krueger, M. (2003). A balanced approach to high-stakes achievement testing: An analysis of the literature with policy implications. *International Electronic Journal for Leadership in Learning*, 7, Article 4. <https://bit.ly/4dNxsTN>
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-331. <https://doi.org/10.3102/01623737024004305>
- Chung, P. J., & Chea, H. (2015). South Korea's accountability policy system and national achievement test. In W. C. Smith (Ed.), *The global testing culture: Shaping education policy, perceptions, and practice* (p.249-260). Symposium Book.
- Çetin, A., & Ünsal, S. (2019). Merkezi sınavların öğretmenler üzerinde sosyal, psikolojik etkisi ve öğretmenlerin öğretim programı uygulamalarına yansımaları [Social, psychological effects of central examinations on teachers and their reflections on teachers' curriculum implementations]. *Hacettepe University Journal of Education/Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34(2), 304-323. <https://doi.org/10.16986/HUJE.2018040672>
- Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. Teachers College Press.
- Decker, D. M., & Bolt, S. E. (2008). Challenges and opportunities for promoting student achievement through large-scale assessment results: Research, reflections, and future directions. *Assessment for Effective Intervention*, 34(1), 43-51. <https://doi.org/10.1177/1534508408314173>
- Education Reform Act. (1988). <https://www.legislation.gov.uk/ukpga/1988/40/contents/enacted>

- Emler, T. E., Zhao, Y., Deng, J., Yin, D., & Wang, Y. (2019). Side effects of large-scale assessments in education. *ECNU Review of Education*, 2(3), 279-296. <https://doi.org/10.1177/2096531119878964>
- Erskine, J. L. (2014). It changes how teachers teach: How testing is corrupting our classrooms and student learning. *Multicultural Education*, 21, 38-40.
- Gilbert, C. (2012). *Towards a self-improving system: The role of school accountability*. National College for School Leadership. <https://bit.ly/3SC4fD0>
- Gonzalez, A., Peters, M. L., Orange, A., & Grigsby, B. (2017). The influence of high-stakes testing on teacher self-efficacy and job-related stress. *Cambridge Journal of Education*, 47(4), 513-531. <https://doi.org/10.1080/0305764X.2016.1214237>
- Gök, G. (2018). *Geniş ölçekli sınavların sosyal ve psikolojik etkilerine ilişkin paydaş görüşleri* [The opinions of stakeholders concerning the social and psychological effects of large scale examinations ] [Master's thesis, Akdeniz University]. Akdeniz University Dspace. <https://bit.ly/3MaHAtM>
- Güler, M., Arslan, Z., & Çelik, D. (2019). 2018 Liselere giriş sınavına ilişkin matematik öğretmenlerinin görüşleri [Mathematics teachers' views on the 2018 entrance exam for high schools]. *Van Yüzüncü Yıl University Journal of Education/Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 16(1), 337-363. <https://doi.org/10.23891/efdyyu.2019.128>
- Gümüştekin, G. E., & Öztemiz, A. B. (2005). Örgütlerde stresin verimlilik ve performansla etkileşimi [Productivity and performance interaction with stress of organization]. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 14(1), 271-288. <https://bit.ly/3WA5gwz>
- Hamilton, L. (2003). Chapter 2: Assessment as a policy tool. *Review of Research in Education*, 27(1), 25-68. <https://doi.org/10.3102/0091732X027001025>
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297-327. <https://doi.org/10.1002/pam.20091>
- Hargreaves, A. (2020). Large-scale assessments and their effects: The case of mid-stakes tests in Ontario. *Journal of Educational Change*, 21, 393-420. <https://doi.org/10.1007/s10833-020-09380-5>
- Jerrim, J., & Sims, S. (2022). School accountability and teacher stress: international evidence from the OECD TALIS study. *Educational Assessment, Evaluation and Accountability*, 34, 5-32. <https://doi.org/10.1007/s11092-021-09360-0>
- Kahraman, İ. (2014). Merkezi ortak sınav uygulamasının etkilerine ilişkin öğretmen görüşleri [The effect of common implementation that related to teachers' opinion]. *Munzur Üniversitesi Sosyal Bilimler Dergisi*, 2(4), 53-73. <https://bit.ly/3MBLjkC>
- Kirkpatrick, R., & Zang, Y. (2011). The negative influences of exam-oriented education on Chinese high school students: Backwash from classroom to child. *Language testing in Asia*, 1, Article 36. <https://doi.org/10.1186/2229-0443-1-3-36>
- Klenowski, V. (2011). Assessment for learning in the accountability era: Queensland, Australia, *Studies in Educational Evaluation*, 37(1), 78-83. <https://doi.org/10.1016/j.stueduc.2011.03.003>
- Klinger, D., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, (76), 1-34. <https://bit.ly/46QPKBI>
- Klinger, D. A., & Rogers, W. T. (2011). Teachers' perceptions of large scale assessment programs within low-stakes accountability frameworks. *International Journal of Testing*, 11(2), 122-143. <https://doi.org/10.1080/15305058.2011.552748>
- Lingard, B., Sellar, S., & Lewis, S. (2017). *Accountabilities in schools and school systems*. Oxford research encyclopedia of education. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190264093.013.74>
- Linn, R. L. (2000). Assessments and accountability. *Educational researcher*, 29(2), 4-16. <https://doi.org/10.3102/0013189X029002004>
- Maroy, C., & Voisin, A. (2017). *Think piece on accountability* [Research Report]. HAL open science. <https://bit.ly/46IdQ16>
- Merriam, S. B. (2023). *Nitel araştırma: Desen ve uygulama için bir rehber* [Qualitative research: A Guide to Design and Implementation] (Trans. S. Turan, F. K. Canbaz, M. Öz, E. Karadağ, D. Yılmaz, H. Özen, M. Yalçın, S. İşçi, Ö. Öztekin, E. Dinç, & S. Yüksel). Nobel. (Original work published 2009).
- Nichols, S. L., & Berliner, D. C. (2008). Why has high-stakes testing so easily slipped into contemporary American life? *Phi Delta Kappan*, 89(9), 672-676. <https://doi.org/10.1177/003172170808900913>
- No Child Left Behind Act. (2001). <https://www.congress.gov/bill/107th-congress/house-bill/1>



- O'Neill, O. (2013). Intelligent accountability in education. *Oxford Review of Education*, 39(1), 4-16, <https://doi.org/10.1080/03054985.2013.764761>
- Organisation for Economic Co-operation and Development. (2011). *School autonomy and accountability: Are they related to student performance?* <https://doi.org/10.1787/5k9h362kcx9w-en>
- Özer Özkan, Y., & Acar Güvendir, M. (2018). Merkezi sınavların öğretmenler üzerindeki öğretimsel ve duyuşsal etkilerini belirlemeye yönelik öğretmen ölçeğinin geliştirilmesi [Development of the teacher's scale that measures the instructional and affective influences of the central examinations on the teachers]. *Inonu University Journal of the Faculty of Education/İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 19(3), 189-204. <http://doi.org/10.17679/inuefd.394383>
- Parcerisa, L., Fontdevila, C., & Verger, A. (2021). Understanding the PISA Influence on National Education Policies: A Focus on Policy Transfer Mechanisms. In S. Jornitz & A. Wilmers (Eds.), *International perspectives on school settings, education policy and digital strategies. a transatlantic discourse in education research* (pp. 182-194). Barbara Budrich Verlag. <https://doi.org/10.2307/j.ctv1gbrzf4.14>
- Parcerisa, L., Verger, A., Pagès, M., & Browes, N. (2022). Teacher autonomy in the age of performance-based accountability: A review based on teaching profession regulatory models (2017-2020). *Education Policy Analysis Archives*, 30(100), 1-29. <https://doi.org/10.14507/epaa.30.6204>
- Perryman, J., & Calvert, G. (2020). What motivates people to teach, and why do they leave? Accountability, performativity and teacher retention. *British Journal of Educational Studies*, 68(1), 3-23, <https://doi.org/10.1080/00071005.2019.1589417>
- Popham, W. J. (1999). Where large scale educational assessment is heading and why it shouldn't. *Educational Measurement: Issues and Practice*, 18(3), 13-17. <https://doi.org/10.1111/j.1745-3992.1999.tb00268.x>
- Popham, W. J. (2000). Big change questions. "Should large-scale assessment be used for accountability?" --answer: depends on the assessment, Silly! *Journal of Educational Change*, 1, 283-289. <https://doi.org/10.1023/A:1010054525759>
- Rahman, K. A., Seraj, P. M. I., Hasan, M. K., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Lang Test Asia*, 11, Article 12. <https://doi.org/10.1186/s40468-021-00129-2>
- Ravitch, D. (2011). *The death and life of the great American school system: How testing and choice are undermining education*. Basic Books.
- Sağır, M., & Mart, M. T. (2017). Temel eğitimden orta öğretime geçiş (TEOG) sistemine ilişkin öğretmen görüşleri [Teachers' views on the transition from primary to secondary education (TEOG) system]. *Eğitime Bakış*, (40), 56-59. <https://bit.ly/3SV0hWi>
- Stecher, B., Hamilton, L. S., & Gonzalez, G. (2003). *Working smarter to leave no child behind*. Rand Corporation.
- Taşkın, G., & Aksoy, G. (2018). Öğrencilerin ve öğretmenlerinin teog sistemi görüşleri ışığında ortaöğretime geçiş sisteminden beklentileri [Students' and teachers' expectations from the transition to secondary education system in the light of TEOG system views of them]. *International Journal of Active Learning*, 3(1), 19-43. <https://bit.ly/3M7nn8v>
- Verger, A., & Parcerisa, L. (2018). Test-based accountability and the rise of regulatory governance in education: A review of global drivers. In A. Wilkins & A. Olmedo (Eds), *Education governance and social theory: Interdisciplinary approaches to research* (pp. 139-158). Bloomsbury.
- Verger, A., Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: a political sociology of global education reforms. *Educational Review*, 71(1), 5-30, <https://doi.org/10.1080/00131911.2019.1522045>
- Worth, J., & Van den Brande, J. (2020, January 29). *Teacher autonomy: How does it relate to job satisfaction and retention?* National Foundation for Educational Research (NFER). <https://bit.ly/3YStpRH>
- Yağar, F., & Dökme, S. (2018). Niteliksel araştırmaların planlanması: araştırma soruları, örneklem seçimi, geçerlik ve güvenirlik [Planning of qualitative researches: research questions, samples, validity and reliability]. *Gazi Sağlık Bilimleri Dergisi*, 3(3), 1-9. <https://bit.ly/3WCOrBn>
- Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, 147(4), 399-435. <https://doi.org/10.1037/bul0000309>
- Yıldırım, A., & Şimşek, H. (2005). *Sosyal bilimlerde nitel araştırma yöntemleri* [Qualitative research methods in the social sciences] (5th ed.). Seckin Publishing.