International Journal of Assessment Tools in Education
2026, Vol. 13, No. 1, 45–65
https://doi.org/10.21449/ijate.1580050
journal homepage: https://dergipark.org.tr/en/pub/ijate *Research Article*

Determining factors influencing middle school teachers in demonstrating test score inflation

Deniz Gökçe Güngör[1*], **Ömer Kutlu**[1]

[1]Ankara University, Faculty of Education, Department of Educational Sciences, Measurement and Evaluation in Education

ARTICLE HISTORY
Received: Nov. 5, 2024
Accepted: Sep. 25, 2025

Keywords:
Test score inflation, Student achievement, High-stake tests, Grading behavior, Score inflation.

Abstract: This study examines the determinants of test score inflation among middle school teachers within Türkiye's exam-driven education system. Test score inflation - defined as increasing students' scores beyond actual performance due to internal or external pressures- poses challenges for fairness, validity, and instructional integrity. To investigate its underlying factors, the researchers developed and validated the Test Score Inflation Scale following DeVellis's (2012) framework. The 20-item, three-factor instrument (Avoidance of Pressure, Student Support, Fairness) demonstrated strong psychometric properties (α = .81–.89; ω = .97; KMO = .855). Data were collected from 507 teachers of Turkish, Mathematics, Science, and Social Studies across public and private schools. Multivariate analyses (MANOVA, multiple regression) revealed significant effects of gender and school type: female teachers and those in private institutions exhibited higher inflation tendencies across all subscales. Conversely, seniority, postgraduate education, managerial role, and subject area showed no significant influence. Regression models indicated that individual (gender, school type) and contextual (physical and sociological characteristics of schools) factors jointly explained 38% of total variance (R^2 = .25-.31 across subscales). Findings suggest that inflation behaviors may reflect adaptive responses to institutional pressures rather than deliberate misconduct. The results underscore the need for policies that promote process-oriented assessment, ethical grading dialogue among teachers, and improved resource conditions that reduce pressure to inflate scores. Future research should validate the scale across contexts and employ longitudinal or mixed-method designs to explore how institutional dynamics and professional norms shape teachers' assessment integrity.

INTRODUCTION

In educational systems shaped by high-stakes accountability, a growing concern has emerged regarding discrepancies between students' actual learning and the scores reported in assessments. This phenomenon -often described as test score inflation- raises critical questions about fairness, measurement validity, and instructional integrity. While *grade inflation* broadly refers to the increase in academic grades without corresponding improvements in learning (Caruth & Caruth, 2013; Woodruff & Ziomek, 2004; Ziomek & Svec, 1997), *test score inflation* more specifically captures the rise in standardized test results independent of actual gains in competence (Jewell *et al*., 2013; Koretz, 2005). These distinctions are not merely semantic;

*Corresponding author at: Deniz Gökçe GÜNGÖR ankugokce@gmail.com Ankara University, Faculty of Education, Department of Educational Sciences, Measurement and Evaluation in Education

e-ISSN: 2148-7456

they reflect different mechanisms and implications for educational equity and policy. Test score inflation often stems from a combination of institutional pressures and instructional practices. Practices such as test coaching, peer assessments with inflated scores, and overuse of multiple-choice formats are increasingly common in competitive schooling contexts (Lipnevich *et al*., 2020; Popham, 1991). While these strategies may serve short-term goals, they ultimately distort assessment data and obscure meaningful feedback. In Turkiye, where national high-stakes exams (e.g., LGS, YKS) play a decisive role in student placement and school rankings, teachers may face implicit or explicit pressure to inflate scores. This context necessitates a focused investigation into the individual and institutional factors shaping such behaviors.

Despite a growing international literature on test score inflation, studies that focus specifically on test score inflation in middle school settings remain limited. Much of the existing research either investigates GPA trends (Jaschik, 2011; Rojstaczer & Healy, 2012) or addresses higher education contexts (Babcock, 2010; Bar *et al*., 2009), which differ structurally and motivationally from the K-12 setting. Empirical studies exploring teachers' motivations and behaviors in test score manipulation at the classroom level - particularly through psychometric instruments- are rare. This study aims to address this gap by developing a multidimensional scale grounded in observed behaviors and motivational patterns in middle school assessment contexts. It differs from previous work by providing a quantitative, teacher-focused tool to detect inflationary tendencies, rather than relying on indirect or archival metrics such as GPA increases over time. Furthermore, teacher characteristics such as gender, experience, and subject area have been shown to influence assessment behavior (McMillan, 2001; Sun, 2019), yet their specific role in score inflation remains underexplored. By incorporating these variables, the present study provides a more nuanced understanding of the contextual and individual factors contributing to inflation practices. Unlike previous research that has primarily examined GPA trends in higher education (e.g., Babcock, 2010; Rojstaczer & Healy, 2012), this study focuses on middle school contexts and employs a multi-variable, scale-based approach. By incorporating teacher-level variables (e.g., gender, experience, school type), the study provides a novel contribution to understanding the contextual and individual determinants of score inflation.

This study addresses this gap by developing and validating the Test Score Inflation Scale, a multidimensional instrument based on DeVellis's (2012) scale development model. Using data from 507 teachers across public and private middle schools, the study examines how individual characteristics, institutional settings, and professional practices relate to score inflation tendencies. The findings offer a basis for evidence-informed policy reforms aimed at improving assessment integrity in the Turkish education system.

Methods of Test Score Inflation

Methods of giving extra points

Test score inflation is a concept that has been particularly emphasized since the early 2000s. It is known that certain factors affecting the teaching-learning process contribute more to this phenomenon. Some of these factors are discussed below.

Peer Assessment. A frequently used method is giving extra points to students; however, students do not find this evaluation reliable and doubt the value of their own learning. Nevertheless, numerous studies have confirmed its effectiveness when applied appropriately. For instance, Xiao and Lucking (2008) demonstrated that combining quantitative and qualitative feedback methods in peer assessment improved both reliability and student satisfaction. Similarly, Gehringer *et al*. (2007) showed that structured peer review using the Expertiza platform led to high-quality, evidence-based feedback comparable to instructor evaluations. As Xiao and Lucking (2008) emphasize, peer assessment is generally conducted in two ways: (i) one student evaluates another student's work, or (ii) all students in the class evaluate a peer's work. However, when this process is not well-managed, artificial score

increases may occur, leading to test score inflation.

Quiz. Unannounced short quizzes that offer extra points function as external motivators intended to boost and sustain student effort (Deci *et al*., 1999; Pintrich, 2003). Such brief, low-stakes assessments -often administered at the end of a lesson- serve as formative checks of students' understanding (Black & Wiliam, 1998). Bangert-Drowns et al. (1991) demonstrated that incorporating such quizzes with timely feedback significantly improves performance on subsequent assessments. However, McDaniel et al. (2007) caution that overreliance on unannounced quizzes may promote surface-level learning strategies -encouraging students to habitually solve excessive practice items- which can contribute to test score inflation. Morrell (1945) argues that conducting short quizzes at the beginning of each class could lead to spending approximately fifteen minutes of class time, but is highly beneficial in the instructional process. The regular use of short quizzes can make students aware of when these quizzes will occur, encouraging them to solve too many tests, which can contribute to test score inflation (Frisch, 2018).

Extra credit assignments. It is observed that teachers allocate space for extra credit assignments when preparing their course outlines. Students are asked to write a report using the materials provided in the course and can receive extra credit in this way. Many studies argue that assignments for extra credit should include application-level questions and support students in making plans to facilitate their self-learning. However, teachers often ask questions that are below the student's learning level in these assignments and only include questions that reflect their existing knowledge (Shepard, 2010). Students meeting the requirement of writing a necessary number of pages in these assignments to receive extra credits meet the definition of score inflation. Additionally, Pynes (2014) mentions that teachers give extra credit assignments to students so they can learn the topics of unanswered questions during exams. Alderson & Hamp-Lyons (1996) point out that in this case, students whose scores are inflated cannot complete their incomplete learning because they do not receive effective feedback, which means teachers will inflate the next measurement results again.

Preparation for tests. Focusing on test content rather than the content and learning outcomes of the course is considered inappropriate behavior in the teaching-learning process. This inappropriate behavior, although there is no real increase in student learning, leads to an increase reflected in scores. Furthermore, structuring instruction in this way also unfairly disadvantages students who do not have the opportunity to learn a broader content area (Haladyna *et al*., 1991; Moore, 1993). Popham (1991) noted that while there is no improvement in students' achievements in the content being tested, test scores are increasingly being prepared. Teachers conduct instruction in the classroom as practice to increase their students' scores on high-stakes tests by including actual items from these tests. These practices include exercises that closely resemble the actual items of the test, thereby teaching the test itself (Popham, 2001). Posner (2004) refers to the use of time and resources allocated to teaching for test preparation as "teaching to the test." One of the consequences of inappropriate test preparations is that teachers avoid preparing exams consisting of open-ended items. This approach, followed by teachers to familiarize students with multiple-choice items, not only artificially inflates test scores but also deviates from good teaching practices. It also deprives students of opportunities to engage in writing and organizing thoughts in different cognitive processes, which are important learning outcomes (Darling-Hammond & Wise, 1985).

Test Consumption. In their study, Kulik, Kulik, and Bangert (1984) found that students would gain much more if they practiced on parallel forms of a test. Pfeifer, Brown, and Juvonen (2007) stated that students practice by solving similar tests to achieve high scores on high-stakes exams. Students believe that the more tests they solve, the more accustomed they will become to high-stakes exams, as these exam results determine their future. Marshall (2003) noted that many educators fall into the misconception of assigning numerous test-solving tasks to teach

students test-taking techniques. Furthermore, it is believed that teachers who think that the more multiple-choice items students encounter, the more prepared they will be for high-stakes exams, contribute to test score inflation by preparing test materials in a manner similar to the content of these tests for their in-class assessments. Nichols and Berliner (2007) indicated the commercialization of this phenomenon, stating that companies established for exam preparation offer test packages tailored to the content of the exam to schools and districts, in addition to providing individual tutoring to students. Thus, the opportunities provided by exam preparation companies encourage students to consume more tests. By using materials with the same structure as the items and format of standardized tests, which lack real-life connections and measure high-level skill development, test companies significantly contribute to test score inflation.

Given that inflated test scores can obstruct the provision of high-quality feedback and impede the remediation of learning gaps, the purpose of this study is to determine the factors that influence middle school teachers' tendencies to engage in test score inflation. By systematically investigating individual characteristics, school context, and measurement-evaluation interactions, the research seeks to inform educators and policymakers about the root causes of score inflation, teachers' tendencies and support the design of targeted interventions. By doing so, the research seeks to guide efforts to address the problem by targeting the underlying causes and the core tendencies of teachers.

Justification of research variables

The variables selected for this study are grounded in theoretical frameworks and supported by international empirical literature. Gender has been found to influence ethical sensitivity and decision-making in assessment contexts, with some studies reporting that female teachers may exhibit greater caution or adherence to guidelines in grading practices (Glover *et al*., 2002; Stajkovic & Luthans, 1997). The type of school -specifically the distinction between public and private institutions- can affect teachers' assessment autonomy, administrative expectations, and perceived performance pressure, which may all relate to score inflation behaviors (Lavy, 2008). The class size or overall student population is also relevant, as larger class sizes often limit individualized feedback and promote more mechanistic scoring strategies (Blatchford *et al*., 2003). Finally, the socioeconomic status (SES) of the school has been linked to external pressures from parents, school reputation concerns, and efforts to maintain favorable public perception- all of which may shape teachers' decisions regarding test scoring (Reardon, 2011; Sirin, 2005). These variables were therefore included to capture contextual and demographic factors that may contribute to teachers' tendencies to inflate test scores.

Purpose of the Research and Research Questions

The aim of this research is to determine the test score inflation tendencies of middle school Turkish, mathematics, science, and social studies teachers in 5th, 6th, 7th, and 8th grades. In line with this aim, the following questions have been addressed:

(1) Based on the results obtained from the Test Score Inflation Scale, what are the tendencies of teachers towards test score inflation in terms of the following sub-dimensions:

a. Avoidance of Pressure,

b. Student Support, and

c. Fairness?

(2) Is there a significant difference in the tendencies of teachers based on their gender, seniority, type of school they work in, status of pursuing postgraduate education, status of holding administrative positions at school, fields of expertise, and whether they have received training on measurement and evaluation?

Which variables predict the scores obtained from the sub-dimensions of the Test Score Inflation Scale

a. Avoidance of Pressure,
b. Student Support,
c. Fairness

separately for each dimension, in the following order: individual characteristics (gender, seniority, type of school, status of pursuing postgraduate education, status of holding administrative positions at school, field of expertise, whether they have received training on measurement and evaluation), measurement and evaluation competencies (providing effective feedback, monitoring student achievement, writing achievement-based questions, implementing process-oriented assessment tools, preparing grading rubrics, grading), personality traits (job satisfaction, ability to communicate openly and sincerely with administrators, positive relationships with students, understanding of the curriculum, communication with parents), and the physical and sociological structure of the school (value given to education, approach of the school administrator to grading, profile of parents, exam-oriented teaching, heating of the school, etc.)?

METHOD

Within the scope of this research, the test score inflation tendencies of teachers have been examined with numerous variables. This research adopts a correlational research model, which investigates the relationship between multiple variables. Correlational research aims to determine the presence and degree of co-variation between two or more variables (Creswell, 2012; Fraenkel & Wallen, 2006). Therefore, correlational research also falls under descriptive research (Fraenkel & Wallen, 2006).

The Participants

The participants comprised 507 teachers of Turkish, Mathematics, Science, and Social Studies in grades 5-8 from both public and private middle schools. They were selected because they constituted the core domains of Türkiye's high-stakes examinations, and the specified grade levels represent a period in which exam-driven practices intensify. A purposive sampling strategy was used to include teachers who taught one of the four subjects, had at least one semester of experience in classroom assessment practices, and represented both public and private school types. Table 1 presents the demographic and professional profiles of the participants.

Table 1. *The individual features of the study group.*

Variable	Group	*N*	%
Gender	Female	265	52.26
	Male	242	47.74
Professional seniority	1-5	218	43
	6-10	66	13.01
	11-15	28	5.52
	16-20	92	18.15
	21 and over	103	20.32
School type	Public School	248	48.92
	Private School	259	51.08
Branch	Turkish	212	41.82
	Mathematics	175	34.51
	Science	58	11.44
	Social Studies	62	12.23
Postgraduate education status	Postgraduate	312	61.54

Variable	Group	*N*	%
	Bachelor	195	38.46
The status of being a principal in the school	Experienced	191	37.67
	Inexperienced	316	62.33
Status of receiving education on the subject of measurement and evaluation	Yes	62	12.23
	No	445	87.77

As seen in Table 1, 52.2% of the 307 middle school teachers participating in the research are female. While 43% of the teachers have 1-5 years of seniority, 41.8% teach Turkish, 34.5% teach mathematics, 12.3% teach social studies, and 11.4% teach science. Furthermore, 51.1% of the teachers work in private schools, while 62.4% do not hold administrative positions at their schools. It is observed that 61.5% of the participants have a postgraduate education level, and 87.8% have received training on measurement and evaluation.

Data Collection Instruments

Test Score Inflation Scale development

The development of the Test Score Inflation Scale was conducted in accordance with the eight-step model proposed by DeVellis (2012), ensuring theoretical grounding and psychometric robustness. The scale was developed following DeVellis's (2012) framework in eight steps: (1) defining the construct based on literature, (2) generating an initial pool of 80 items, (3) selecting a 4-point Likert response format, (4) expert review by measurement and language specialists, (5) pilot administration to 180 teachers, (6) item analysis and EFA with Promax rotation, (7) reliability estimation (Cronbach's α and McDonald's ω), and (8) CFA.

Step 1 – Defining the Construct. The construct of "test score inflation" was defined as a teacher's tendency to increase students' test scores beyond what their actual performance warrants, often as a result of internal motivations or external pressures. It was conceptualized as a multidimensional structure, comprising ethical compromises, student-centered justifications, and situational pressures, grounded in extensive literature on assessment and measurement (e.g., Koretz, 2005; Lipnevich *et al*., 2020; Popham, 1991). This step aimed to ensure strong construct validity by clearly delineating the boundaries and dimensions of the phenomenon. The literature suggests that score inflation is driven not by a single cause but a constellation of internal and external motivators, including ethical compromises (Koretz, 2005), contextual pressures (Popham, 1991), and attempts to protect students (Lipnevich *et al*., 2020). Therefore, the construct was conceptualized as multidimensional, incorporating three latent domains identified during theoretical analysis: Avoidance of Pressure, Student Support, and Fairness. This multidimensionality aligns with the framework of defensible test use in school-based assessment contexts (McMillan, 2001).

Step 2 – Generating an Item Pool. An initial item pool of 80 items was generated based on a thorough literature review and a conceptual model reflecting multiple facets of the construct. These items aimed to capture teacher behaviors and beliefs regarding test score modification across various classroom and institutional contexts. The item generation process was theory-driven. Items were derived based on the three proposed dimensions and their operational definitions. Ethical compromise items reflected justifications for score inflation that violated fairness principles (e.g., giving points to help students pass). Contextual pressure items reflected systemic or administrative expectations (e.g., avoiding conflict with parents or principals). Student support items reflected protective intentions (e.g., compensating for socioeconomic disadvantage). Behavioral indicators and prior qualitative studies (e.g., Cizek, 1999; McMillan, 2001) also guided item phrasing.

Step 3 – Choosing the Format of Measurement. A 4-point Likert-type response format was selected, comprising: (1) *Strongly Disagree*, (2) *Disagree*, (3) *Agree*, and (4) *Strongly*

Agree. The absence of a neutral midpoint was deliberate in order to reduce indecisiveness and prompt more reflective responses from participants. This format is commonly recommended in measuring ethically sensitive constructs where social desirability may play a role (Kulas *et al*., 2008).

Step 4 – Expert Review for Content Validity. The initial items were subjected to expert review by three specialists in educational measurement and two language experts. Items were evaluated in terms of content relevance, clarity, and linguistic accuracy. As a result of this review, 23 items were eliminated or revised, yielding a 57-item draft. A further refinement reduced this to 43 items.

Step 5 – Administering Items to a Development Sample. The 43-item draft scale was administered to 180 middle school teachers in Türkiye, selected using convenience sampling. Data were screened for missing values, outliers, and assumptions of normality.

Step 6 – Item Analysis and Exploratory Factor Analysis (EFA). An EFA was conducted on the pilot sample (N = 180) using Principal Axis Factoring with Promax rotation, which was deemed appropriate given the anticipated correlations among latent constructs. Prior to extraction, the suitability of the data was assessed. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .855, and Bartlett's Test of Sphericity yielded a statistically significant result (χ^2 = 1854.55, df = 210, p < .001), indicating the data were factorable.

A combination of Kaiser's criterion (eigenvalues > 1), the scree plot, and parallel analysis was used to determine the number of factors to retain. Three factors emerged, accounting for 49.41% of the total variance. Items with factor loadings below .40 or with high cross-loadings (> .30) were eliminated. The final structure consisted of 20 items distributed across three interpretable and theoretically coherent dimensions:

- Avoidance of Pressure (8 items),
- Student Support (4 items),
- Fairness (8 items).

All factor loadings exceeded .50, and the internal consistency values (Cronbach's α) ranged from .814 to .894, with .931 for the total scale.

Using multiple, complementary retention criteria (eigenvalues, scree, and parallel analysis) reduces the risk of factor over-extraction and supports the stability of the retained solution (see Henson & Roberts, 2006).

The naming of the three factors was guided by both the item content and theoretical underpinnings. Items in Factor 1 consistently referenced teachers' resistance to external pressures (e.g., from parents or administrators), aligning with "Avoidance of Pressure." Factor 2 contained items reflecting affective concern and advocacy for students, thus labeled "Student Support." Factor 3 captured ethically motivated behavior opposing unfair inflation, forming the "Fairness" factor. This interpretation is consistent with prior conceptualizations of ethical assessment behavior (Brookhart, 2004; McMillan, 2001).

Step 7 – Reliability Analysis. Cronbach's alpha coefficients for the three subdimensions - Avoidance of Pressure (8 items, α = .894), Student Support (4 items, α = .814), and Fairness (8 items, α = .892)- demonstrated high internal consistency. In addition, McDonald's omega was calculated for the overall scale, indicating excellent reliability (ω = .972).

Step 8 – Confirmatory Factor Analysis. To validate the factor structure revealed by the EFA, confirmatory factor analysis was conducted on an independent sample (N = 179). Although maximum likelihood estimation (MLE) is frequently employed as the estimation method in confirmatory factor analysis (CFA) within the literature, Newsom (2017) argues that when Likert-type scales include seven or more response categories, the measurements can reasonably be treated as continuous, whereas in cases with fewer response categories, it is more

appropriate to treat the measurements as ordinal. In light of this consideration, and given that the scale developed in the present study employs a four-point rating format and that partial violations of the assumption of multivariate normality were observed, the WLSMV estimation method was selected. The model fit was assessed through a variety of indices, including the chi-square statistic relative to degrees of freedom (χ^2/df = 2.68), Comparative Fit Index (CFI = .916), Tucker-Lewis Index (TLI = .902), Root Mean Square Error of Approximation (RMSEA = .097), and Standardized Root Mean Square Residual (SRMR = .059). These results suggested an acceptable level of model-data fit, particularly for a model derived from self-report data in the field of education. All standardized factor loadings were statistically significant ($p < .001$), ranging between .58 and .81, supporting the structural validity of the proposed three-factor model.

Demographic form

In addition to the Test Score Inflation Scale, a demographic form was used to collect information regarding participants' gender, professional seniority, subject taught, postgraduate education status, administrative role in school, and whether they had received training in measurement and evaluation.

Data Analysis

In the first aim of the study, whether there was a significant difference in test score inflation tendencies among teachers with different characteristics was examined. MANOVA was used for the comparison of groups at the comparison stage. Before proceeding with the analyses, the assumptions of measurement level, normal distribution, and homogeneity of variances were considered (Field, 2013; Howell, 2007). Accordingly, the measurement of test score inflation is at the interval level. The assumption of normal distribution was tested both graphically and statistically. Z-scores were calculated and found to range between -1.95 and +1.71, which is within acceptable limits (±3). In addition, the Shapiro–Wilk test was applied to the total scores, and the result was non-significant ($p > .05$), indicating that the distribution did not significantly deviate from normality. In the test of homogeneity of variances, the variables for which variances were assumed to be homogeneous were having postgraduate education, holding administrative positions at the school, and receiving training on measurement and evaluation. When interpreting the tables, the similarity of variances in subgroups was considered.

In the second aim of the study, which examined individual characteristics, measurement and evaluation competencies, personality traits, and the physical and sociological structures of the schools where teachers worked, predicted test score inflation tendencies were examined. For this purpose, multiple linear regression analysis was conducted. Before proceeding with the analyses, missing values and outliers were examined, and no missing or outlier values were encountered. Normal distribution was examined both graphically and statistically, and no violation of normal distribution was observed. After the models were built, the examination revealed tolerance values ranging from .71 to .98, and VIF values ranging from 1.02 to 1.41. Therefore, it was concluded that there was no multicollinearity problem.

FINDINGS

The first aim of the study was to examine whether there were significant differences among groups of teachers based on their gender, seniority, type of school they work in, status of having postgraduate education, administrative positions at the school, teaching areas, and whether they received training on measurement and evaluation, regarding their test score inflation tendencies as measured by the Test Score Inflation Scale.

To examine the differences between groups, the MANOVA technique was employed. In this study, comparisons were made based on the total scores obtained from the dimensions of the developed scale, which had a three-dimensional structure. Accordingly, the test score inflation tendencies of teachers were examined using MANOVA for differences between groups based

on the scores obtained from the Avoiding Pressure, Student Support, and Fairness dimensions of the Test Score Inflation Scale, with respect to their gender, type of school they work in, status of having postgraduate education, managerial positions at the school, and whether they received training on measurement and evaluation. The findings are presented in Table 2.

Table 2. *MANOVA results for two-category independent variables.*

Dimensions (Variables)	Gender	n	$\bar{X}$	sd	F	p	η^2
Avoidance of Pressure	Female	265	22.14	4.47	112.66	≤ .05	.18
	Male	242	17.65	5.05			
Student Support	Female	265	11.40	2.30	126.27	≤ .05	.20
	Male	242	9.14	2.21			
Fairness	Female	265	20.76	3.84	132.90	≤ .05	.20
	Male	242	16.50	4.48			
Dimensions (Variables)	School Type	n	$\bar{X}$	sd	F	p	η^2
Avoidance of Pressure	Public	248	18.36	5.30	51.80	≤ .05	.09
	Private	259	21.56	4.71			
Student Support	Public	248	9.68	2.38	34.05	≤ .05	.06
	Private	259	10.95	2.50			
Fairness	Public	248	17.43	4.83	40.33	≤ .05	.07
	Private	259	19.97	4.15			
Dimensions (Variables)	Graduate Education Status	n	$\bar{X}$	sd	F	p	η^2
Avoidance of Pressure	Yes	312	20.06	5.21	0.12	.72	.00
	No	195	19.89	5.33			
Student Support	Yes	312	10.38	2.43	0.44	.50	.00
	No	195	10.23	2.66			
Fairness	Yes	312	18.90	4.52	1.20	.27	.00
	No	195	18.44	4.89			
Dimensions (Variables)	Managerial Position	n	$\bar{X}$	sd	F	p	η^2
Avoidance of Pressure	Yes	191	20.06	5.51	0.05	.82	.00
	No	316	19.96	5.10			
Student Support	Yes	191	10.52	2.56	1.79	.18	.00
	No	316	10.21	2.50			
Fairness	Yes	191	19.09	4.87	1.87	.17	.00
	No	316	18.50	4.53			
Dimensions (Variables)	Receiving Education on Measurement and Evaluation	n	$\bar{X}$	sd	F	p	η^2
Avoidance of Pressure	Yes	445	19.95	5.25	0.24	.62	.00
	No	62	20.31	5.32			
Student Support	Yes	445	10.36	2.51	0.66	.41	.00
	No	62	10.08	2.58			
Fairness	Yes	445	18.70	4.65	0.07	.79	.00
	No	62	18.87	4.82			

When Table 2 was examined, a significant difference was found in terms of gender and school type among the sub-dimensions of the Test Score Inflation Scale, while no significant difference was found among the variables of having postgraduate education, receiving training on measurement and evaluation, and managerial position. In the dimensions of Avoidance of Pressure, Student Support, and Fairness, it was observed that the female study group had significantly higher scores in the gender variable, and the private school study group had significantly higher scores in the school type variable. The findings obtained by Kolevzon (1981) indicated that the scores given to students did not accurately reflect academic achievement.

Table 3. *MANOVA results for senority and field variables.*

Dimensions (Variables)	Seniority	n	$\bar{X}$	sd	F	p	η^2
Avoidance of Pressure	1-5	218	19.92	5.16	0.50	.73	.00
	6-10	66	19.95	5.52			
	11-15	28	19.93	5.31			
	16-20	92	19.55	5.26			
	21+	103	20.59	5.31			
Student Support	1-5	218	10.28	2.52	1.35	.24	.01
	6-10	66	10.17	2.54			
	11-15	28	11.18	2.05			
	16-20	92	10.57	2.55			
	21+	103	10.07	2.58			
Fairness	1-5	218	18.62	4.65	0.13	.97	.00
	6-10	66	19.05	4.61			
	11-15	28	18.75	3.98			
	16-20	92	18.85	4.88			
	21+	103	18.62	4.78			
Dimensions (Variables)	Field	n	$\bar{X}$	sd	F	p	η^2
Avoidance of Pressure	Social Studies	62	19.55	5.50	0.25	.85	.00
	Mathematics	175	20.07	4.85			
	Science	58	19.74	5.63			
	Turkish	212	20.13	5.42			
Student Support	Social Studies	62	9.79	2.69	2.40	.06	.01
	Mathematics	175	10.61	2.26			
	Science	58	9.84	2.52			
	Turkish	212	10.38	2.65			
Fairness	Social Studies	62	18.06	4.66	0.61	.60	.00
	Mathematics	175	18.73	4.54			
	Science	58	18.55	4.55			
	Turkish	212	18.96	4.80			

This research has shown that female teachers tend to inflate test scores. According to Kolevzon's findings, since female teachers have lower chances of being recruited and promoted, it has been determined that the effect of teacher success is clearly seen in the evaluation of student success. This situation supports the idea that female teachers contribute more to test score inflation. In other words, the results obtained indicate that a significant portion of the inflation can be explained depending on the gender of the instructors.

MANOVA tests were conducted when the number of dependent variables was more than one category. MANOVA was conducted to determine whether there was a significant difference in terms of teachers' seniority and fields in terms of the scores obtained from the dimensions of the Test Score Inflation Scale, and the findings are presented in Table 3. When examining Table 3, no significant difference was observed in the test score inflation tendencies of teachers based on their seniority and the field they teach. This indicates that seniority and expertise in teaching do not affect test scores in evaluating student performance, showing similar evaluation tendencies among teachers.

Table 4. *Regression analysis results for the dependent variable avoidance of pressure.*

Model	Variable	*B*	SH_B	*β*	*t*	*p*
1	Constant	17.65	0.31		57.70	≤ .001
	Gender	4.49	0.42	.43	10.61	≤ .001
2	Constant	19.00	0.40		47.05	≤ .001
	Gende	3.91	0.43	.37	9.09	≤ .001
	School Type	-2.14	0.43	-.20	-4.98	≤ .001
3	Constant	15.32	1.11		13.82	≤ .001
	Gender	3.74	0.43	.36	8.75	≤ .001
	School Type	-1.76	0.44	-.17	-4.03	≤ .001
	School's Physical Characteristics	0.20	0.06	.15	3.56	≤ .001
4	Constant	13.75	1.29		10.67	≤ .001
	Gender	3.74	0.43	.36	8.78	≤ .001
	School Type	-1.76	0.44	-.17	-4.11	≤ .001
	School's Physical Characteristics	0.21	0.06	.15	3.76	≤ .001
	Frequency of Exams	0.62	0.26	.09	2.37	.018
5	Constant	15.56	1.55		10.07	≤ .001
	Gender	3.74	0.42	.36	8.82	≤ .001
	School Type	-1.73	0.44	-.17	-3.98	≤ .001
	School's Physical Characteristics	0.22	0.06	.16	3.93	≤ .001
	Frequency of Exams	0.56	0.26	.08	2.16	.031
	Interaction with Other Teachers in Measurement and Evaluation Topics	-0.80	0.04	-.08	-2.11	.035

Regression analysis, used to explore relationships between variables, attempted to determine the causal effect of one variable on another. This analysis provides information about the relationship between variables and the strength of this relationship. Multiple regression analysis is used when multiple independent variables are used to predict the values of the dependent variable. In this research, multiple independent variables were used in predicting the scores of the Test Score Inflation Tendency. Therefore, multiple linear regression analysis was performed. The predictive power of teachers' scores on the Avoidance of Pressure subscale of the Test Score Inflation Scale by their individual characteristics, measurement and evaluation competencies, personality traits, and the physical and sociological structures of the schools where they work was examined. The results of the stepwise multiple linear regression are presented in Table 4.

When Table 4 is examined, it can be observed that gender (t = 8.82; p < .05), school type (t = -3.98; p < .05), physical characteristics of the school (sum of features such as the need for repairs to the school building, lack of heating system in winter) (t = 3.93; p < .05), teachers' frequency of conducting exams (t = 2.16; p < .05), and interaction with other teachers in measurement and evaluation topics (t = -2.11; p < .05) significantly predicted teachers'

avoidance behaviors. However, seniority, having a postgraduate education, managerial position, field of expertise, training in measurement and evaluation, frequency of using measurement and evaluation practices, measurement and evaluation competencies, sociological structure of the school, personality traits of teachers, and measurement and evaluation competencies of teachers were not found to be significant predictors of their avoidance behaviors.

The predictive capacity of individual characteristics, measurement and evaluation competencies, personality traits, and the physical and sociological structures of the schools where teachers work on the scores obtained from the Student Support dimension, which is the second dimension of the Test Score Inflation Scale, has been examined. Stepwise multiple linear regression results for the scores of the Test Score Inflation Scale are presented in Table 5.

Table 5. *Results of multiple regression analysis with student support as the dependent variable.*

Model	Variable	*B*	SH_B	*β*	*t*	*p*
1	Constant	9.15	0.15		62.92	≤ .001
	Gender	2.26	0.20	.45	11.24	≤ .001
2	Constant	5.96	0.41		14.53	≤ .001
	Gender	1.95	0.19	.39	10.12	≤ .001
	School's Sociological Structure	0.18	0.02	.31	8.22	≤ .001
3	Constant	4.80	0.49		9.78	≤ .001
	Gender	1.88	0.19	.37	9.86	≤ .001
	School's Sociological Structure	0.13	0.02	.23	5.35	≤ .001
	School's Physical Characteristics	0.12	0.03	.18	4.16	≤ .001

When Table 5 is examined, it is observed that teachers' scores on supporting students were significantly predicted by their gender ($t = 9.86$; $p < .05$), the sociological structure of the school ($t = 5.35$; $p < .05$), and the physical characteristics of the school ($t = 4.16$; $p < .05$). However, variables such as seniority, type of school, status of having postgraduate education, managerial status, field of expertise, receiving training on measurement and evaluation, frequency of conducting exams, frequency of using measurement and evaluation practices, competence in measurement and evaluation, interaction with other teachers on measurement and evaluation issues, personality traits of teachers, and competence of teachers in measurement and evaluation did not significantly predict teachers' scores on supporting students.

The scores obtained by teachers from the third dimension of the Test Score Inflation Scale, which is the fairness dimension, were examined for their predictability based on their individual characteristics, their competence in measurement and evaluation, their personality traits, and the physical and sociological structure of the school. The stepwise multiple linear regression results for the fairness dimension scores are presented in Table 6.

According to Table 6, , teachers' fairness scores were significantly predicted by their gender ($t = 9.59$; $p < .05$), the physical features of their schools (such as the total of features like the need for school building repairs, lack of heating system in winter) ($t = 3.79$; $p < .05$), school type ($t = 2.60$; $p < .05$), the sociological structure of the school ($t = 2.52$; $p < .05$), and the teachers' field ($t = 1.99$; $p < .05$). Teachers' tenure, graduate education status, managerial status, participation in training on measurement and evaluation, frequency of examination administration, frequency of use of measurement and evaluation practices, measurement and evaluation competencies, interaction with other teachers on measurement and evaluation issues, personality traits of teachers at the school, and measurement and evaluation competencies of teachers at the school were not found to be significant predictors.

Table 6. *Results of multiple regression analysis with fairness as the dependent variable.*

Model	Variable	*B*	SH_B	*β*	*t*	*p*
1	Constant	16.50	0.27		61.71	≤ .001
	Gender	4.26	0.37	.46	11.53	≤ .001
2	Constant	11.29	0.86		13.16	≤ .001
	Gender	3.86	0.36	.41	10.68	≤ .001
	Physical Features of the School	0.30	0.05	.25	6.36	≤ .001
3	Constant	12.43	0.96		12.94	≤ .001
	Gender	3.63	0.37	.39	9.81	≤ .001
	Physical Features of the School	0.27	0.05	.22	5.59	≤ .001
	School Type	-0.98	0.38	-.11	-2.58	.010
4	Constant	11.45	1.04		11.06	≤ .001
	Gender	3.53	0.37	.38	9.52	≤ .001
	Physical Features of the School	0.21	0.05	.17	3.86	≤ .001
	School Type	-0.94	0.38	-.10	-2.49	.013
	Sociological Structure of the School	0.11	0.05	.11	2.44	.015
5	Constant	10.58	1.12		9.43	≤ .001
	Gender	3.55	0.37	.38	9.59	≤ .001
	Physical Features of the School	0.21	0.05	.17	3.79	≤ .001
	School Type	-0.98	0.38	-.11	-2.60	.010
	Sociological Structure of the School	0.12	0.05	.11	2.52	.012
	Field	0.32	0.16	.08	1.99	.047

DISCUSSION and CONCLUSION

The first question of this study is whether teacher characteristics lead to differences in test score inflation tendencies. When the scores obtained from the Test Score Inflation Scale by female and male teachers were examined, significant differences were found among the dimensions of Avoidance of Pressure, Student Support, and Fairness. The findings obtained from these analyses indicate that female teachers have a greater tendency to inflate test scores compared to male teachers. The gender effect on the Fairness dimension was moderate (partial $\eta^2 = .08$), consistent with recent findings that female teachers demonstrate stronger relational accountability in assessment (Griffith & Sovero, 2021). The effect of school type on Avoidance of Pressure was smaller (partial $\eta^2 = .03$) but aligns with evidence that private school contexts amplify parental expectations and grading pressures (Milovanska-Farrington, 2022). Jewell & McPherson (2012) reached the conclusion in their university study that male faculty members were less inclined to inflate grades compared to female faculty members. Furthermore, in the same study, female instructors expressed a significantly higher rate of test score inflation compared to male instructors. The findings of this study are also parallel to the findings of Jewell and McPherson (2012) study.

This tendency may reflect not only individual choices but also broader institutional expectations and gendered professional roles in teaching. Female teachers often experience greater emotional labor, relational accountability, and parental scrutiny, which can foster compensatory grading behaviors as a form of relational maintenance or role conformity (Bennett, 1982; Griffith & Sovero, 2021). Additionally, according to the results of the MANOVA conducted based on the school type variable, there was a significant difference between private and public schools. Couch *et al*. (1993) shared that in private schools, parents have more influence on grades, hence scores are more manipulated by teachers. There is evidence of grade inflation in private schools in many studies (Grove & Wasserman, 2004; Kezim *et al*., 2005). Rojstacher and Healy (2012) stated that in private schools, there is an increase in grades without

improvement in student performance. Barksdale and Thomas (2000) reported that teachers feel pressure from students, school administrators, and parents regarding issues such as teaching test-taking techniques, focusing on topics that will appear on high-stakes tests rather than course content, and preparing exam questions to ensure students pass their classes. These findings underline the argument that in highly competitive or consumer-driven school environments, such as private schools, grade inflation may be instrumentalized as a tool to satisfy stakeholders and maintain institutional image (Rothstein, 2007; Walsh, 2010).

In this study, no significant difference was found in test score inflation tendencies among teachers based on their professional tenure. However, Sonner and Sharland (1993) shared the finding that less tenured teachers give higher grades than more tenured teachers. Although tenure did not emerge as a significant predictor in this sample, prior research suggests that novice teachers may be more susceptible to external pressures or may inflate scores to foster rapport with students, highlighting the need for mentorship in ethical assessment practices.

According to the independent samples MANOVA results based on the variable of whether teachers hold managerial positions at school, there is no significant difference between teachers who hold administrative roles and those who do not in terms of the dimensions of the Test Score Inflation Scale. Smith and Fleisher (2011) found a significant relationship between expected scores and achievement scores. They noted that some school administrators adopt the principle of "student as a customer," using student success as a measure of educator evaluation. Dhuey and Smith (2018), in their research on school principals and teacher autonomy, stated that principals have a significant influence, especially on the scores of tests related to mathematics and reading comprehension.

According to the variable of "receiving training on measurement and evaluation," no significant difference was found among the dimensions of the Test Score Inflation Scale for teachers. The majority of teachers in the sample of this study have received training on measurement and evaluation. This may be because teachers receive training on measurement and evaluation during their undergraduate and graduate studies, as well as seminars and workshops they attend after starting their careers. Cizek *et al*. (1995) found in their study that teachers who did not receive training on measurement and evaluation were lacking in classroom assessment practices but were able to develop assessment tools as they gained professional experience.

Regarding the variable of subject area, no statistically significant difference was observed among the dimensions of the scale. Au (2011) noted that curriculum content becomes standardized in programs that are shaped by high-stakes exams, further limiting instruction for students with different learning styles. Berliner (2011) highlighted that the use of math and reading questions associated with high-stakes tests restricts all subjects that used to connect students with society (such as going on field trips or participating in projects that investigate local pollution). Additionally, he pointed out that subjects focused on by high-stakes tests are given more hours, leading students to perceive success as scores obtained from these tests.

The second question of the study is to determine the variables that predict teachers' test score inflation tendencies. For this purpose, multiple regression analysis was conducted with the scale scores as the dependent variable. According to the accepted model, variables such as gender (female-male), physical characteristics of the school, type of school (public-private), the sociological structure of the school, and interaction with teachers regarding measurement and evaluation predict the dependent variable.

In this study, it has been observed that the scores obtained from the dimensions of the Test Score Inflation Scale, namely Avoidance of Pressure, Student Support, and Fairness, are significantly higher for female teachers compared to male teachers. In light of this finding, it can be stated that gender is the most important variable influencing test score inflation tendencies. Griffith and Sovero (2021) observed that female teachers are more concerned about school management and job security, and therefore inflate scores with the understanding that

"if the student is happy with a high grade, the school administrator will also be satisfied with the teacher." Kolevzon (1981) found that the majority of teachers who engage in score inflation in certain areas is often female. Jewell and McPherson (2012) believe that women give extra points to avoid the pressure of not being able to get promoted, find a position, or be perceived as unsuccessful teachers. Howard and Henney (1998) stated that female educators give extra points to students for their attendance and participation in class. Groves (2003) found that female teachers give additional assignments and tasks to increase students' motivation and well-being. Bennett (1982) found that students expect female teachers to provide both different course content and teaching processes and to show a bias towards students during the evaluation process. Additionally, female teachers may increase their test score inflation tendencies by providing extra credit assignments, teaching test techniques, and preparing for high-stakes tests to support students. Warwick and Jatoi (1994) stated that female teachers teach more through practice assignments and problem-solving, thus increasing their students' success. In addition, female teachers, feeling unable to prevent injustice due to the difficulties imposed by the education system, may be inclined to inflate their students' success by preparing them for high-stakes exams.

Rothstein (2007) stated in his research on the effects of school competition that among the reasons for choice, school administrators promise to meet parental demands. Therefore, when parental requests regarding the school are not met, the school fulfils these promises by inflating test scores. In this study, it has been observed that the physical characteristics of the school have a significant effect on predicting test score inflation tendencies. The physical characteristics of the school account for a 7% effect in explaining test score inflation tendencies. Accordingly, the good physical condition of the school (such as the absence of the need for repairs to the school building, the presence of heating systems, uncrowded classrooms, and teachers' workspaces) leads to an increase in test score inflation tendencies. A study conducted in the United States indicated that teachers in low-income or disadvantaged schools show less tendency for test score inflation (Gershenson, 2018).

According to the accepted model of regression analysis, the type of school variable is another variable that can explain test score inflation tendencies. While the type of school contributes 2% to explaining the scale scores of test score inflation, its effect on the dimensions of Avoidance of Pressure and Fairness has been found. Chan *et al*. (2007) stated that competitive interactions between schools and the perspective of schools on competition affect grade inflation. Additionally, it has been suggested that the grading policies of schools during competition affect the well-being of students and that schools engage in test score inflation because of students' preference for schools based on their grades. Walsh (2010) stated in his study that increased competition encourages school administrators to lower grading standards. It has also been noted that teachers are more inclined to engage in test score inflation behaviors for students from high-income families to appease parents who insist on test score inflation (Gershenson, 2018). In a study conducted in Türkiye (Kaya & Nartgün, 2016), parents' expectations from private schools were examined, and generally, parents expressed that high academic achievement is an important factor. These expectations can influence private school administrators and teachers. The findings of this study indicate that the tendency for test score inflation in private schools is significantly higher than in public schools. Overall, the regression analysis highlights that score inflation is not merely a product of individual characteristics but also an outcome of systemic and contextual dynamics within schools. The multidimensional structure of the Test Score Inflation Scale -Avoidance of Pressure, Student Support, and Fairness- suggests that inflation behaviors are often motivated by tensions between pedagogical goals and institutional accountability. For instance, inflating scores to support students or to avoid institutional reprimands may be interpreted as adaptive strategies rather than deviant behavior. This perspective aligns with McMillan (2001), who

emphasized that grading practices often reflect ethical ambiguity and institutional contradictions rather than intentional deception.

The sociological structure of the school can be defined in various ways. In this study, factors such as valuing education in the school, high student achievement, absence of parental and administrative pressure, teachers not feeling pressured, focus on tests and exams, and the belief that grades affect the student's future were used to determine the sociological structure of the school. It has been found in the research that the sociological structure of the school predicts the teacher's test score inflation tendency. Accordingly, the sociological structure of the school contributes 1% to explaining the test score inflation tendency. This is a relatively small contribution and should be interpreted with caution. In addition, as scores obtained from the sociological structure of the school that affect the dimensions of Support Student and Fairness increase, or in other words, as the sociological structure of the school improves, teachers believe that the tendency for test score inflation increases. In this study, the good sociological structure of the school increases the teacher's test score inflation tendency. This situation may also be related to the type of school variable. In private schools in Türkiye, especially considering the exams for secondary and higher education placement, it is observed that students have higher academic achievement (MoNE, 2018). These results further suggest that sociological privilege and institutional pressure may inadvertently legitimize score inflation under the guise of maintaining academic excellence. Rather than being a simple artifact of weak teacher judgment, inflation here appears to be embedded in the very structure of school expectations.

Smith and Fleisher (2011) suggest that teachers sometimes set unreachable academic performance goals, which may create a sense of wrongdoing among educators for assigning low grades. In this case, interacting with other teachers to discuss assessment tools and evaluation processes may reduce the tendency for grade inflation. In this study, interaction with other teachers regarding measurement and evaluation is another variable that predicts scores obtained from the Avoidance of Pressure dimension of the grade inflation scale. The effect of this variable indicates an expected outcome. Interaction among teachers in the same school, such as reviewing course objectives, discussing new measurement and evaluation practices, writing exam questions, examining the alignment of questions with objectives, grading exam papers, enhancing student achievement, solving questions in test books, and solving questions from transition exams to high school, reduces grade inflation tendencies. This situation can be considered as a positive effect of the widespread dissemination of knowledge in the field of measurement and evaluation. Therefore, it is thought that encouraging teachers to engage more in interaction regarding measurement and evaluation topics would be beneficial for education. Germaine *et al*. (2011) argue that instructions and assignment deadlines in tasks given to prevent requests for grade increases from students or objections to the grades received should be clear, and the expectations of the teaching process should be clearly expressed. They suggest that by doing so, students will perceive their teachers as guides rather than adversaries who want to deduct points from them. Teachers with high measurement and evaluation skills provide clearer tasks and instructions to students. When clear and specific criteria are shared with students, they tend to trust the teacher's competence more and have lower expectations of high grades. This reinforces the idea that inflation tendencies can be curbed through professional dialogue and shared responsibility. When teachers co-construct assessments, align expectations, and exchange feedback on grading standards, they not only reduce individual bias but also foster a more ethical and transparent assessment culture (Germaine *et al*., 2011).

LIMITATIONS

This study has several limitations that should be acknowledged. First, the data were collected using self-report measures, which may be subject to social desirability bias. Despite the anonymity provided, teachers may have responded in a manner they perceived as socially acceptable rather than fully reflecting their actual practices or beliefs. Second, the study was

limited to middle school teachers in Türkiye, which restricts the generalizability of the findings to other educational levels, regions, or cultural contexts. Third, although the scale development followed DeVellis's (2012) recommended procedures, including expert reviews and factor analyses, the psychometric properties of the scale could be further validated through test-retest reliability or cross-validation with independent samples. Fourth, while this study considered several background variables (e.g., gender, school type, professional experience), other potentially influential variables -such as teachers' personal grading philosophy, school accountability pressures, or perceptions of fairness norms- were not examined. Finally, the cross-sectional design of the study limits causal interpretations; longitudinal studies could provide deeper insights into how test score inflation tendencies evolve over time or across different policy implementations.

SUGGESTION

Based on the findings of this study, the following recommendations are provided:

(1) Teachers engage in inflating scores through various means. This practice may involve creating exams that are easy for students to answer, preparing exams based on questions covered in class, or composing exams consisting of content beyond the scope of the curriculum, among other methods. Therefore, it is suggested that teacher competence should not be evaluated based solely on student achievement, as this notion may contribute to the inflation of test scores, where student grades are seen as indicators of teacher competency. Given that gender significantly predicted test score inflation (β = .38), professional development programs should include gender-sensitive approaches addressing relational pressures in assessment.

(2) Teachers' beliefs that mock exams are more effective in preparing students for multiple-choice exams lead them to make inappropriate preparations for tests. It is important to educate teachers, school administrators, and parents about what constitutes inappropriate test preparations and their consequences. Otherwise, when significant learning objectives are excluded from the teaching process, an educational system that does not contribute to the development of advanced skills in students will be adopted, and schools will produce more unqualified individuals in society.

(3) Awareness among education stakeholders about grade inflation is important, as literature review has shown insufficient research on this concept. In this regard, it is recommended that this concept be included in the curriculum of undergraduate and graduate programs in education faculties.

(4) Although grade inflation is carried out by teachers, there are internal and external factors that contribute to this situation. Sometimes teachers may inflate grades to reduce students' anxiety, while other times they may do so to avoid pressure from school management or parents. To address the problem, all factors influencing the issue should be addressed collectively, and decisions should be made based on the opinions of all stakeholders. However, one of the significant challenges at this point is the lack of longitudinal research in the field. More evidence can be provided to the decision-making process with data obtained from longitudinal studies. The finding that physical school conditions predicted inflation (β = .25) suggests that resource allocation policies may indirectly shape teachers' grading behaviors.

(5) National education ministries should make decisions that emphasize a process-oriented education approach that prioritizes learning rather than a result-oriented education approach that emphasizes grades.

Acknowledgments

This study is part of the first author's master's thesis, supervised by the second author.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Ankara University, 22/12/2021-387.

Contribution of Authors

Deniz Gökçe Güngör: Conceptualization, Literature Review, Data Collection, Data Analysis, Writing-original draft, and, Writing-review & editing. **Ömer Kutlu:** Conceptualization, Methodology, Supervision.

Orcid

Deniz Gökçe Güngör https://orcid.org/0000-0002-3783-9243
Ömer Kutlu https://orcid.org/0000-0003-4364-5629

REFERENCES

Alderson, J.C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: a study of washback. *Language Testing*, *13*(3), 280-297. https://doi.org/10.1177/026553229601300304

Au, W. (2011). Teaching under the new Taylorism: high-stakes testing and the standardization of the 21st century curriculum. *Journal of Curriculum Studies*, *43*(1), 25-45. https://doi.org/10.1080/00220272.2010.521261

Barksdale, M.A., & Thomas, K.F. (2000). What's at stake in High-Stakes testing. *Journal of Teacher Education*, *51*(5), 384-397. https://doi.org/10.1177/0022487100051005006

Bennett, S.K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, *74*(2), 170-179. https://doi.org/10.1037/0022-0663.74.2.170

Berliner, D.C. (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

Bangert-Drowns, R.L., Kulik, C.C., Kulik, J.A., & Morgan, M. (1991). The instructional effect of feedback in Test-Like events. *Review of Educational Research*, *61*(2), 213-238. https://doi.org/10.3102/00346543061002213

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education Principles Policy and Practice*, *5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Blatchford, P., Bassett, P., Goldstein, H., & Martin, C. (2003). Are class size differences related to pupils' educational progress and classroom processes? Findings from the institute of education class size study of children aged 5-7 years. *British Educational Research Journal*, *29*(5), 709-730. https://doi.org/10.1080/0141192032000133668

Caruth, D.L., & Caruth, G.D. (2013). Grade inflation: An issue for higher education? *Turkish Online Journal of Distance Education*, *14*(1), 102-110. https://dergipark.org.tr/tr/pub/tojde/issue/16895/176017

Chan, W., Li, H., & Suen, W. (2007). A signaling theory of grade inflation. *International Economic Review*, *48*(3), 1065-1090. https://doi.org/10.1111/j.1468-2354.2007.00454.x

Cizek, G.J., Robert, R., & Fitzgerald, S. (1995, April 18-22). *Further investigation of teachers' assessment practices* [Paper presentation]. The Annual Meeting of the American Educational Research Association, San Francisco, CA, United States.

Couch, J.F., Shughart, W.F., & Williams, A.L. (1993). Private school enrollment and public school performance. *Public Choice*, *76*(4), 301-312. https://doi.org/10.1007/bf01053301

Creswell, J.W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education.

Darling-Hammond, L., & Wise, A.E. (1985). Beyond standardization: state standards and school improvement. *Elementary School Journal*, *85*(3), 315-336. https://doi.org/10.1086/461408

Deci, E.L., Koestner, R., & Ryan, R.M. (1999). A meta analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627-668. https://doi.org/10.1037/0033-2909.125.6.627

Devellis, R. (2012). *Scale development theory and applications*. Sage Publications.

Dhuey, E., & Smith, J. (2017). How school principals influence student learning. *Empirical Economics*, *54*(2), 851-882. https://doi.org/10.1007/s00181-017-1259-9

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. SAGE.

Fraenkel, J.R., & Wallen, N.E. (2006). *How to design and Evaluate research in Education*. McGraw-Hill Companies.

Frisch, V.V. (2018). *Do pop quizzes have a positive effect on exam grades?* [Master's thesis, Otterbein University]. Digital Commons @ Otterbein. https://digitalcommons.otterbein.edu/stu_master/16/

Gehringer, E.F. (2001). Assignment and quality control of peer reviewers. In *Proceedings of the 2001 American society for engineering education annual conference & exposition* (pp. 6.224.1-6.224.8). https://doi.org/10.18260/1-2%2D%2D8941

Germaine, R., Pacis, D., Bustillos, T., Keough, P., Wheeler, D., & Weegar, M.A. (2011). Lessons learned through a review of grade appeal data. *Journal of Research in Innovative Teaching*, *4*(1), 39-46. https://www.nu.edu/wp-content/uploads/2018/11/journal-of-research-in-innovative-teaching-volume-4.pdf

Glover, S.H., Bumpus, M.A., Sharp, G.F., & Munchus, G.A. (2002). Gender differences in ethical decision making. *Women in Management Review*, *17*(5), 217-227. https://doi.org/10.1108/09649420210433175

Griffith, A.L., & Sovero, V. (2021). Under pressure: How faculty gender and contract uncertainty impact students' grades. *Economics of Education Review*, *83*, 102-126. https://doi.org/10.1016/j.econedurev.2021.102126

Grove, W.A., & Wasserman, T. (2004). The life-cycle pattern of collegiate GPA: Longitudinal cohort analysis and grade inflation. *The Journal of Economic Education*, *35*(2), 162-174. https://doi.org/10.3200/jece.35.2.162-174

Groves, J. (2003). Perceptions of extra credit by hospitality and tourism faculty. *Journal of Hospitality & Tourism Education*, *15*(3), 38-43. https://doi.org/10.1080/10963758.2003.10697026

Haladyna, T.M., Nolen, S.B., & Haas, N.S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, *20*(5), 2-7. https://doi.org/10.2307/1176395

Howard, J.R., & Henney, A.L. (1998). Student participation and instructor gendering the Mixed Age College classroom. *The Journal of Higher Education*, *69*(4), 384-405. https://doi.org/10.1080/00221546.1998.11775141

Howell, D.C. (2007). *Statistical methods for Psychology*. Thomson.

Jewell, R.T., & McPherson, M. (2012). Instructor- Specific grade inflation: Incentives, gender, and ethnicity. *Social Science Quarterly*, *93*(1), 95-109. https://doi.org/10.1111/j.1540-6237.2011.00827.x

Jewell, R.T., McPherson, M., & Tieslau, M. (2013). Whose fault is it? Assigning blame for grade inflit ion in higher education. *Applied Economics*, *45*(9), 1185-1200. https://doi.org/10.1080/00036846.2011.621884

Kaya, A.B., & Nartgün S.S. (2016). Özel okul velilerinin beklentileri doğrultusunda okul imajı oluşturma [Creating a school image in line with the expectations of private school parents]. *Journal of Research in Education and Teaching*, *5*(2), 153-167. http://www.jret.org/FileUpload/ks281142/File/17.senay_nartgun.pdf

Kezim, B., Pariseau, S.E., & Quinn, F. (2005). Is grade inflation related to faculty status? *Journal of Education for Business*, *80*(6), 358-364. https://doi.org/10.3200/joeb.80.6.358-364

Kolevzon, M.S. (1981). Grade inflation in higher education: A comparative study. *Research in Higher Education*, *15*(3), 195-212. https://doi.org/10.1007/bf00976416

Koretz, D. (2005). Alignment, high stakes, and the inflation of test scores. *Yearbook of The National Society for The Study of Education*, *104*(2), 99-118. https://doi.org/10.1111/j.1744-7984.2005.00027.x

Kulas, J.T., Stachowski, A.A., & Haynes, B.A. (2008). Middle response functioning in Likert-responses to personality items. *Journal of Business and Psychology*, *22*(3), 251-259. https://doi.org/10.1007/s10869-008-9064-2

Kulik, J.A., Kulik, C.C., & Bangert, R.L. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal*, *21*(2), 435-447. https://doi.org/10.3102/00028312021002435

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, *92*(10-11), 2083-2105. https://doi.org/10.1016/j.jpubeco.2008.02.009

Lipnevich, A.A., Guskey, T.R., Murano, D.M., & Smith, J.K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education: Principles, Policy & Practice*, *27*(5), 1-21. https://doi.org/10.1080/0969594X.2020.1799190

Marshall, K. (2003, October 1). *Test Prep-The Junk Food of Education*. Education Week. https://www.edweek.org/teaching-learning/opinion-test-prep-the-junk-food-of-education/2003/10

McDaniel, M.A., Anderson, J.L., Derbish, M.H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *The European Journal of Cognitive Psychology*, *19*(4-5), 494-513. https://doi.org/10.1080/09541440701326154

McMillan, J.H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement Issues and Practice*, *20*(1), 20-32. https://doi.org/10.1111/j.1745-3992.2001.tb00055.x

Milovanska-Farrington, S. (2022). Do parents expect too much or is it all about grades? The discrepancy between parents' aspirations and child's academic performance, and parental satisfaction with the school. *Cogent Economics & Finance*, *10*(1). https://doi.org/10.1080/23322039.2022.2079178

Ministry of National Education [MoNE]. (2018). *2018 Liselere Geçiş Sistemi (LGS)* [2018 High School Transition System]. https://www.meb.gov.tr/meb_iys_dosyalar/2018_12/17094056_2018_lgs_rapor.pdf

Moore, W.P. (1993, April 13-15). *Preparation of students for testing: Teacher differentiation of appropriate and inappropriate practices* [Paper presentation]. The annual meeting of the national council on measurement in education, Atlanta, GA, United States. https://files.eric.ed.gov/fulltext/ED357033.pdf

Morrell, W.E. (1945). Quizzes to encourage preparation for lectures. *Journal of Chemical Education*, *22*(8), 402-405. https://doi.org/10.1021/ed022p402

Newsom, J.T. (2017). Structural models for binary repeated measures: Linking modern longitudinal structural equation models to conventional categorical data analysis for matched pairs. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 626-635. https://doi.org/10.1080/10705511.2016.1276837

Nichols, S.L., & Berliner, D.C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Harvard Education Press.

Gershenson, S. (2018). *Grade inflation in high schools (2005–2016)*. Thomas B. Fordham Institute.

Pfeifer, J.H., Brown, C.S., & Juvonen, J. (2007). Prejudice reduction in schools teaching tolerance in schools: Lessons learned since Brown v. Board of Education about the development and reduction of children's prejudice. *Social Policy Report*, *21*(2), 1-24. https://doi.org/10.1002/j.2379-3988.2007.tb00051.x

Popham, W.J. (1991). Appropriateness of teachers' test preparation practices. *Educational Measurement: Issues and Practice*, *10*(4), 12-15. https://doi.org/10.1111/j.1745-3992.1991.tb00211.x

Popham, W.J. (2001, March 1). *Teaching to the test?* ASCD. https://www.ascd.org/el/articles/teaching-to-the-test

Posner, D. (2004). What's wrong with teaching to the test? *Phi Delta Kappan*, *85*(10), 749-751. https://doi.org/10.1177/003172170408501009

Pintrich, P.R. (2003). A Motivational Science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*(4), 667-686. https://doi.org/10.1037/0022-0663.95.4.667

Pynes, C.A. (2014). Seven arguments against extra credit. *Teaching Philosophy*, *37*(2), 191-214. https://doi.org/10.5840/teachphil20144414

Reardon, S.F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane & G. Duncan (Eds.), *Whither opportunity? Rising inequality and the uncertain life chances of low income children* (pp. 91-116). Russell Sage Foundation Press.

Rojstaczer, S., & Healy, C. (2012). Where a is ordinary: The evolution of American College and university grading, 1940-2009. *Teachers College Record*, *114*(7), 1-23. https://doi.org/10.1177/016146811211400707

Rothstein, J. (2007). Does competition among public schools benefit students and taxpayers? comment. *The American Economic Review*, *97*(5), 2026-2037. https://doi.org/10.1257/aer.97.5.2026

Saha, L.J., & Dworkin, A.G. (2009). *International handbook of research on teachers and teaching*. Springer.

Shepard, T. (2010, October 21-23). *The use of extra credit to improve course design* [Paper Presentation]. 2010 North Midwest Section, Mankato, Minnesota, United States https://doi.org/10.18260/1-2-1114-36472

Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta analytic review of research. *Review of Educational Research*, *75*(3), 417-453. https://doi.org/10.3102/00346543075003417

Smith, D.E., & Fleisher, S. (2011). The implications of grade inflation: Faculty integrity versus the pressure to succeed. *Journal of Research in Innovative Teaching*, *4*(1), 32-38. http://www.nu.edu/assets/resources/pageResources/journal-of-research-in-innovative-teaching-volume-4.pdf

Sonner, B.S., & Sharland, A. (1993). Grading differences between graduate teaching assistants and faculty members in the introductory marketing class. *Journal of Marketing Education*, *15*(2), 44-49. https://doi.org/10.1177/027347539301500207

Stajkovic, A.D., Luthans, F. (1997. Business ethics across cultures: A social cognitive model. *Journal of World Business*, *32*(1), 17-34. https://doi.org/10.1016/s1090-9516(97)90023-7

Walsh, P. (2010). Does competition among schools encourage grade inflation? *Journal of School Choice*, *4*(2), 149-173. https://doi.org/10.1080/15582159.2010.483918

Warwick, D P., & Jatoi, H. (1994). Teacher gender and student achievement in Pakistan. *Comparative Education Review*, *38*(3), 377-399. https://doi.org/10.1086/447257

Woodruff, D.J., & Ziomek, R.L. (2004). *High School Grade Inflation from 1991 to 2003.* ACT.

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education*, *11*(3-4), 186-193. https://doi.org/10.1016/j.iheduc.2008.06.005

Ziomek, R.L., & Svec, J.C. (1997). High school grades and achievement: Evidence of grade inflation. *NASSP Bulletin*, *81*(587), 105-113. https://doi.org/10.1177/019263659708158716