

RESEARCH

Open Access



# A study of EFL teachers' classroom grading practices in secondary schools and private institutes: a mixed methods approach

Majid Nowruzi

Correspondence: [m-nowruzi@phd.araku.ac.ir](mailto:m-nowruzi@phd.araku.ac.ir); [Lasttorch@gmail.com](mailto>Lasttorch@gmail.com)  
Department of English Language,  
Faculty of Foreign Languages, Arak  
University, Arak, Iran

## Abstract

This explanatory sequential mixed methods study aimed at exploring the grading decision-making of Iranian English language teachers in terms of the factors used when assigning grades and the rationales behind using those factors. In the preliminary quantitative phase, a questionnaire was issued to 300 secondary school and private institute EFL teachers. Quantitative data analyses showed that teachers attached the most weight to nonachievement factors such as effort, improvement, ability, and participation when determining grades. Next, follow-up interviews were conducted with 30 teachers from the initial sample. The analyses of interview data revealed that teachers assigned hodgepodge grades on five major grounds of learning encouragement, motivation enhancement, lack of specific grading criteria, pressure from stakeholders, and flexibility in grading. Data integration indicated that teacher grading decision-making was influenced by both internal and external factors, with adverse consequences for grading validity. Eliciting explanations for the use of specific grading criteria from the same teachers who utilized those criteria in their grading in a single study added to the novelty of this research. Implications for grade interpretation and use, accountability in classroom assessment, and teachers' professional development are discussed.

**Keywords:** Explanatory sequential design, Grading decision-making, Hodgepodge grading, Achievement factors, Nonachievement factors

## Introduction

Grades are unquestionably the primary indicators of student performance within schools (Guskey & Link, 2019). They represent the most popular currency exchanged within educational systems worldwide (Pattison et al., 2013). They summarize student learning and have influenced various high-stakes educational decisions about students such as college or university admissions (Brookhart et al., 2016; DeLuca et al., 2017; Guskey, 2015). Despite the growing use of grades for and their pervasive influence on educational decision-making (Brookhart et al., 2016; Pattison et al., 2013), research has revealed that teachers rely on various achievement and nonachievement data when making grading decisions (Guskey, 2011; Nowruzi & Amerian, 2020; Randall &

Engelhard, 2009, 2010; Yesbeck, 2011). Doubts and concerns about ineffective grading practices and conflated grades have increasingly been expressed by teachers, parents, educational administrators, and researchers (Baird, 2013; Black & William, 1998; Brookhart, 2004, 2013; Guskey & Bailey, 2001; Smail, 2013). It has been argued that using various sources of evidence contributes to the multidimensionality of grades and complicates the meanings they convey (Cross & Frary, 1999).

The vast majority of grading research reveals that, contrary to the measurement community's recommendations to grade on achievement only (Airasian, 2000; Stiggins, 2001), teachers base grades on ancillary traits of achievement such as effort, participation, ability, improvement, behavior, personality traits, and work habits (Brookhart et al., 2016; Cizek, 1996; Duncan & Noonan, 2007; Guskey, 2011; Guskey & Link, 2019; McMillan, 2001; McMillan et al., 2002; Nowruzi & Amerian, 2020; Randall & Engelhard, 2009, 2010; Stiggins et al., 1989; Sun & Cheng, 2013; Svennberg et al., 2014; Yesbeck, 2011). Previous research also documented that academic knowledge contributed only marginally to the grade assigned (Bowers, 2011; Brennan et al., 2001; Linn, 2000; Willingham et al., 2002; Woodruff & Ziomek, 2004). Likewise, Cizek (1996) states that even though "grades continue to be relied upon to communicate important information about [academic] performance and progress ... they probably don't" (p. 104).

Although Iranian EFL teachers' grades are ultimately transformed into report card grades that impact students' university admissions and career opportunities, the nature of these grades has long remained unknown. The problem is that it is unclear to what extent grades that are presumed by all stakeholders to represent academic performance are, in reality, based on academic factors. More importantly, Iranian EFL teachers' reasons for assigning hodgepodge grades have not been explored, an issue that perpetuates the troubles with classroom assessment and problematizes the validity of teacher-assigned grades.

Studying teachers' grading practices has particularly significant implications in the Iranian context considering the socially accepted role of grades and the lasting influences they have on students' educational lives and parents' perceptions of academic achievement. Extensive grade-based instructional decision-making by Iranian instructors on a daily basis, on the one hand, and concerns about what a grade represents, on the other hand, highlight the significance of researching grading in the under-studied Iranian setting. Although investigating EFL teachers' grading practices is still in its infancy (Brindley, 2007; Rea-Dickins, 2004), little is known about Iranian EFL teachers' grading decision-making. Thus, conducting this study can be an effort to bridge the gap in the literature about the scarcity of cross-cultural grading research expressed by Brookhart et al. (2016). Likewise, the results of keyword searches in scholarly search engines such as *Eric*, *Google Scholar*, and *tandfonline* indicated that very limited studies addressed Iranian teachers' classroom assessment practices. Of these, studies that centered on Iranian English language teachers' grading practices utilizing a mixed methods design were almost nonexistent.

The purpose of this explanatory sequential mixed methods study (Creswell & Plano Clark, 2018) was to gain understanding of the factors that Iranian EFL secondary school and private institute teachers considered when assigning grades. Specifically, the quantitative phase of this study aimed at unpacking the grading criteria that teachers utilized in determining grades, while the qualitative phase sought to explore, in more depth, the

rationales behind teachers' amalgamated grading. The initial purpose of selecting a two-phase design was to provide explanations for and in-depth understanding of the mechanisms embedded in teachers' grade-giving practices.

### Grading research literature

To date, various studies have been done on teachers' grading practices by exploring the factors considered in grading and teachers' rationales for assigning conflated grades to students. A number of these studies investigated the use of academic and nonacademic factors in determining grades and their impacts on grade interpretation and use (Brookhart, 1993, 1994; Frary et al., 1993; McMillan, 2001; Svennberg et al., 2014). Overall, grading research can be broken down into two major categories: (a) studies conducted in ESL contexts where achievement criteria were dominant when determining grades, despite the relative importance of nonachievement factors (e.g., Guskey & Link, 2019; McMillan, 2001, 2003; McMillan et al., 2002; McMillan & Nash, 2000), and (b) subject-specific research in EFL settings in which nonachievement factors were found to be the principal components of grades that English language teachers assigned (e.g., Cheng & Sun, 2015; Nowruzi & Amerian, 2020; Sun & Cheng, 2013). What follows is a review of a number of these studies.

The first tentative model for teachers' classroom assessment and grading practices was proposed by McMillan and Nash (2000) who examined the grading practices of 24 elementary and secondary mathematics and English teachers. Their model comprised six themes, three of which including (a) teacher beliefs and values, (b) realities of classrooms, and (c) external factors pertained to the rationale behind teachers' grading decision-making. They reported constant tension between internal factors such as teachers' philosophies of teaching and learning and external criteria such as parents, standardized testing, and classroom constraints. McMillan and Nash (2000) found that teachers constantly struggled to strike a balance between internal and external grading influencers. However, probable differential grading practices of teachers across two basically different subject matters of math and English in elementary and secondary schools were not taken into account, an issue which might have partially skewed their findings. This issue was resolved in the current study by focusing solely on English language teachers' grading practices.

In another study, McMillan (2001) surveyed 2293 secondary school teachers of mathematics, English, science, and social studies about the factors they used to determine grades. Four major grading components emerged including *academic achievement*, *academic enablers* (i.e., factors that contribute to achievement), *external criteria*, and *extra credit for borderline cases*. Although academic performance weighted the most heavily in constructing grades, academic enablers (i.e., effort, ability, improvement, participation) were found to be important contributors to grades as well, as verified by other research (Frary et al., 1993; McMillan et al., 2002; Stiggins & Conklin, 1992). Likewise, Guskey and Link (2019) studied the grading factors used by 943 school teachers and found that they utilized a multitude of academic and nonacademic factors such as effort, class participation, students' work habits, and neatness to assign grades. Not only did these two studies combine the grading results across different subject matters, which should have been done with extreme caution, but they also viewed grading only quantitatively and did not explore teachers' rationales behind using the grading

components. Although such findings are interesting per se, they do not provide practitioners with any clues as to why grading objectivity is at stake and how it can be remedied.

Kunnath (2016) conducted an explanatory sequential mixed methods research and found that teachers strived to boost their grading objectiveness by using pedagogical practices most compatible with their educational philosophies. Concerning the teachers' rationales for using nonacademic factors in grading, he found that teachers used nonacademic criteria to (a) justify their pedagogical practices, (b) encourage student success, and (c) contribute to fairness in grading. The first two themes tended to reflect teachers' beliefs and values, whereas the third theme can be interpreted in light of teachers' attempts to accommodate various external pressures from stakeholders such as parents.

To address the need for subject-specific grading research, some studies examined English language teachers' grading decision-making in ESL/EFL contexts such as Canada, Hong Kong, and China (Cheng & Sun, 2015; Cheng & Wang, 2007; Sun & Cheng, 2013). Sun and Cheng (2013) explored grade meaning in the grading practices of 350 Chinese secondary school English language teachers. Grades were found to reflect teachers' perceptions of (a) effort, homework quality, and fulfillment of duty; and (b) the extent of learning as judged by academic enablers, improvement, learning processes, and achievement. Additionally, teachers considered either what was fair or what contributed to learning as rationales for determining grades. The results indicated that teachers ascribed the most importance to nonachievement factors for assigning grades. Academic achievement happened to be only "part of the construct, but not the whole of it" (Brookhart, 1993, p. 139). What seems problematic is that in their study, Sun and Cheng (2013) provided thoughtful reasoning for separating effort from other academic enablers and considering enablers and achievement factors of equal importance in measuring learning.

In a later study, similar results were reported by Cheng and Sun (2015), showing that although teachers incorporated both academic and nonacademic factors in grading, they placed the strongest emphasis on the latter. Nowruzi and Amerian (2020) obtained comparable findings after studying the grading practices of five Iranian EFL institute teachers qualitatively. Of the 92 grading constructs elicited in their study using the repertory grid technique of Kelly's (1991) personal construct theory (PCT), more than two-thirds were nonacademic, pointing to English language teachers' extensive use of nonachievement factors in determining grades. However, the small number of participants limited the external validity of the study. Such findings conflict with those of previous research where achievement was the principal determinant of grades (e.g., Cheng & Wang, 2007; McMillan, 2001; McMillan et al., 2002). This discrepancy in the findings of grading research conducted in ESL versus EFL settings was one of the reasons that motivated the present study.

Cheng and Sun (2015) also reported three grading components including (a) *norm/objective-referenced factors* such as mastery of learning objectives, class participation, and grading compared to other teachers' grades, (b) *effort factor* which comprised of effort, disruptive behavior, and homework; and (c) *performance factors* including academic and nonacademic performance, cognitive abilities, and performance compared with peers. It remains unclear, however, how participation, effort, and nonacademic

performance belonged to three separate grading components, whereas, in effect, they all tend to enable achievement. Also, the idea of placing disruptive behavior in the same component with homework and effort seems hard to grasp. Besides, the rationale behind Chinese EFL teachers' hodgepodge grading was not elaborated on. Such shortcomings highlight the need for undertaking new grading research that can present the readership with more firmly established grading components.

This study was designed and conducted to bridge some of the existing gaps in the grading literature. The majority of grading research in ESL contexts were done quantitatively (Brookhart et al., 2016), increasing the risk of "seeing just the forest but not the trees" (Saito & Inoi, 2017, p. 217). In addition, many of these studies' findings were combined across various subject matters, which might have problematized their implications, knowing that teachers commonly utilize different grading schemes for different courses. The observed discrepancy among different research results concerning the reliance on achievement or nonachievement factors, classification of grading factors on thoughtful reasoning, and the scarcity of mixed methods studies that enable researchers to investigate the grading factors and teachers' rationales and interpret the combined results in a single study were some of the gaps that this study attempted to narrow. To address these problems, the following research questions were formulated:

- 1) What factors do Iranian secondary EFL teachers consider when determining grades?
- 2) What factors do Iranian private EFL institute teachers consider when determining grades?
- 3) What are Iranian EFL teachers' reasons for assigning hodgepodge grades to student work?
- 4) How can the qualitative findings help provide a deeper understanding of teachers' grading practices?

## **Method**

### **Context of the study**

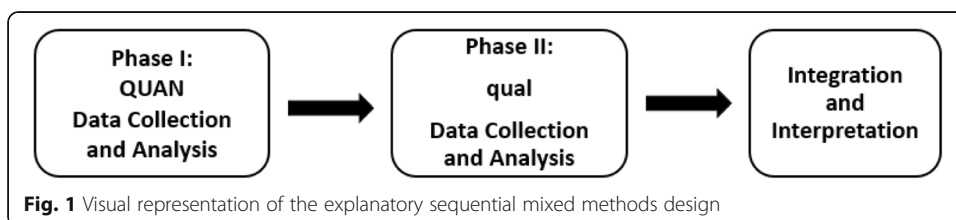
English language instruction in Iran officially starts at secondary education. Iranian students study English for six consecutive years from grade 7 to 12 prior to taking part in the university entrance exam known as *Konkooor*. Teaching English in schools is mainly limited to teaching skills that are assessed in *Konkooor* such as reading comprehension, grammar, and vocabulary. Thus, students' abilities in using English for communicative purposes remain underdeveloped by the end of mainstream schooling. Therefore, many students pursue foreign language learning in private EFL institutes and schools simultaneously. This turns institutes into very important venues for instructional research because the success or failure of countless number of learners and their motivation for learning English are heavily influenced by teacher-assigned grades. For these reasons, this study focused on the grading practices of English language instructors in private institutes as well as those in secondary schools to broaden the scope of investigating teachers' grade-giving practices.

**Design and rationale of the study**

An explanatory sequential mixed methods design (Creswell & Plano Clark, 2018; Tashakkori & Teddlie, 1998) consisting of an initial cross-sectional survey design (McMillan, 2000) followed by a basic interpretative qualitative design (Ary et al., 2014) was used for data collection and analysis, as shown in Fig. 1. Whereas the quantitative phase was aimed at identifying the factors used by Iranian EFL teachers when determining grades, the follow-up qualitative phase elaborated on the initial numerical results by exploring participants’ views (Ivankova et al., 2006). Ultimately, the findings of the two phases were integrated and interpreted (Creswell et al., 2003) to provide insights into the way teachers made grading decisions. It was believed that neither quantitative nor qualitative data alone could adequately unpack the nuanced meanings that teachers attached to grades in the Iranian context. Another reason for mixing numeric and text data in this study was to foster complementarity (Greene et al., 1989; Johnson & Turner, 2003; Tashakkori & Teddlie, 1998). The priority (Creswell et al., 2003), however, was given to the quantitative phase due to the scarcity of Iranian grading studies. It was assumed that examining teachers’ grading practices quantitatively and then seeking explanations for such practices from the same participants within a single study would warrant more accurate results than combining separate research findings.

**Participants**

Three hundred Iranian EFL teachers were recruited for the quantitative phase of this study through convenience sampling. Schools and institutes in urban areas with the largest number of EFL teachers teaching in them had priority. Sixty-two percent of the sample taught English in secondary schools, while 38% were institute teachers. Also, the majority of the teachers (61%) were female, with 39% male teachers. The participants aged 20-49 with a mean of 35 (*SD* = 8.5). In addition, 52% had between 5 and 20 years of teaching experience, with 4% novice teachers in their first year of teaching. The majority of the participants (58%) were academically certified in majors such as translation, literature, or linguistics and 57% were TEFL graduates (*n* = 171). The number of participants with no academic qualifications was negligible. A purposeful subsample of 30 teachers, 15 secondary school, and 15 private institute teachers, was selected for the follow-up interviews. The selection was guided by the criterion sampling technique (Ary et al., 2014) from among those whose responses to survey items were representative of the reported means for the nonacademic factors of the survey. All participants consented to take part in interviews.



**Fig. 1** Visual representation of the explanatory sequential mixed methods design



### Instrument

A 34-item Likert type questionnaire titled *Classroom Assessment and Grading Practices Survey* adapted for use from McMillan (2001) was used in the quantitative phase of this study (Additional file 1). The questionnaire includes four sections. The first section explains the research purposes and gathers participants' demographic data. The next three sections address (a) factors used in grading, (b) assessment types used for making grading decisions, and (c) cognitive ability levels of students measured by teachers' classroom assessments. However, only findings from the first subscale (grading factors) were reported in this study. The grading subscale consists of 19 items on a 6-point scale ranging from 1 as *Not at all* to 6 as *Completely*. Teachers were requested to select a number from 1 to 6 for each grading item based on the frequency with which they considered that item when giving grades. Subsequently, the means of each of the items were computed, indicating the degree of teachers' reliance on each factor when grading.

The validity of the questionnaire was secured by asking a panel of 10 teachers, five from each setting, to examine the questionnaire items for content validity and item wording prior to data collection. The panel recommended that the Persian translation of the survey items accompany the original English version to ensure accurate understanding of the survey items. The questionnaire was also piloted with 20 teachers and minor modifications were made in the questionnaire. Cronbach's alpha reliability coefficient for the grading subscale was .86.

### Quantitative data collection and analysis

Of the 400 questionnaires distributed both electronically and manually, 330 questionnaires were returned, indicative of a response rate of 82%. After discarding the incomplete questionnaires, 300 fully completed questionnaires, 187 from secondary schools and 113 from private institutes, were kept for data analysis. The electronic questionnaires, created using Google Forms, reached teachers via their groups on *WhatsApp* and *Telegram* once permissions were obtained from group admins. The response rates for the electronic and manual data collection methods were 79% and 61%, respectively, indicating teachers' preference for taking online surveys. Teachers were informed beforehand that their participation was voluntary. They were also ensured that their responses would remain confidential. It took each respondent nearly 20 min to complete the questionnaire. The data collection took place between October and December 2019.

Once the data collection ended, percentages, means, and standard deviations were computed for each of the grading items to detect possible grading trends. Subsequently, two principal component analyses (PCA) with Varimax rotation were performed separately for each dataset to create overarching grading components that were more manageable and enhanced data interpretability. Prior to conducting the PCAs, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were calculated to assess the data factorability. The KMO was .86 and Bartlett's sphericity test was significant at  $p < .0001$ , pointing to the suitability of data for factor analysis. Additionally, the case-item ratio was 10 to 1, indicative of the suitability of grading items for factor extraction. Components with eigenvalues of 1 and higher were

retained and labeled based on the items with the highest loadings and in line with the literature. SPSS 24.0 was used for data analysis in the study and the critical alpha value was set to  $\alpha = 0.05$ .

### **Qualitative data collection and analysis**

The interview protocol (Additional file 2) consisting of five open-ended questions was developed after the quantitative data analysis to have teachers elaborate on and explain the preliminary numeric data including the elicited grading components (Creswell et al., 2003). The questions were then pilot-tested with four teachers from the initial sample to identify any likely flaws in the protocol prior to the actual data collection. The piloted data were excluded from the final analysis. Next, the interview protocol was discussed with the teachers and slight modifications were made to it. It was agreed that all interviews be conducted in the candidates' native language, i.e., Persian, to minimize the chances of providing wrong information. Later, probes were added to the protocol to help elicit in-depth information about interviewees' grading practices. All interviews were conducted, audio recorded, translated into English and then transcribed verbatim by the researcher. Qualitative data collection took place in the winter of 2020.

Interview data were analyzed using QSR NVivo v. 10, a qualitative data analysis software used for coding and theme development. Once the translated transcripts were approved by the interviewees, the researcher read and reread the transcripts and noted down any preliminary concepts. Then, the data were coded by segmenting and labeling the transcripts using NVivo. Next, the initial in vivo codes were refined by grouping similar codes into overarching themes and subthemes. To ensure the trustworthiness of the coding process, the same coding was repeated by the researcher in a 2-week interval to calculate the intra-coder agreement index. In addition, a colleague was requested to code some transcripts to report the inter-coder agreement index. The intra- and inter-coder agreement indexes were .84 and .73, respectively. Disputed codes were subsequently resolved by discussion. In the end, the generated themes were returned to the interviewees for member checking (Guba & Lincoln, 1989) to ensure authenticity, credibility, trustworthiness, and robustness of the outcomes. Also, teachers' quotes accompanied all elicited themes.

## **Results**

### **Quantitative results**

In this section, first, the quantitative results including the descriptive and factor analysis outputs obtained from the questionnaires will be reported for each setting separately. Then, the qualitative findings including the themes and subthemes that emerged from content analyses of interview transcripts alongside exemplar quotes will be presented.

### ***Secondary EFL teachers' grading factors***

Table 1 presents the percentages of teachers' responses to each of the 19 grading items across the 6-point scale along with the relevant means and standard deviations. The means of the grading items ranged from 4.17 for *effort* to 2.52 for *grade distributions of*



**Table 1** Descriptive statistics of grading items used by secondary EFL teachers

Grading item	Percentage (%)						Descriptive		
	Not at all (1)	Very little (2)	Some (3)	Quite a bit (4)	Extensively (5)	Completely (6)	N	M	SD
1. Student effort-how much the student tried to learn	0	10	24	21	28	17	187	4.17	1.25
2. Class participation	2	11	23	13	37	14	187	4.14	1.34
3. Improvement of performance	1	7	32	15	31	14	185	4.10	1.25
4. Ability levels of the students	2	13	29	19	21	16	183	3.93	1.37
5. Quality of graded homework	3	8	28	26	24	11	185	3.92	1.24
6. Specific learning objectives mastered	1	18	23	22	25	11	185	3.84	1.32
7. Effort, improvement, behavior and other non-test factors for borderline cases	5	11	36	23	19	6	185	3.60	1.23
8. Academic performance as opposed to other factors	4	21	29	17	25	4	182	3.50	1.29
9. Completion of ungraded homework	5	13	33	29	16	4	184	3.50	1.20
10. Work habits and neatness	0.0	20	38	20	17	5	187	3.49	1.15
11. Performance compared to other students	6	20	32	20	19	3	183	3.36	1.25
12. Extra credit for academic performance	11	16	27	27	16	3	184	3.31	1.31
13. Performance compared to a set scale	14	19	31	12	20	4	183	3.19	1.41
14. Disruptive student behavior	10	26	29	13	18	4	183	3.16	1.37
15. School or district policy for grading	15	20	33	20	9	3	183	2.98	1.29
16. Performance compared to students from previous years	24	25	20	9	16	6	185	2.85	1.57
17. Inclusion of Zeros for incomplete assignment	20	23	34	10	13	0.0	187	2.75	1.28
18. Extra credit for nonacademic performance	26	25	27	14	7	1	184	2.55	1.27
19. Grade distributions of other teachers	28	26	25	11	7	3	187	2.52	1.36

*other teachers*. The top five items on the list with the largest means were *effort* ( $M = 4.17$ ), *participation* ( $M = 4.14$ ), *improvement* ( $M = 4.10$ ), *ability* ( $M = 3.93$ ), and *graded homework* ( $M = 3.92$ ), all of which were considered nonacademic. While nearly two-thirds (66%) of the 187 secondary teachers frequently considered student effort when determining grades, none excluded effort from consideration when grading (*Not at all* = 0). Large means and SDs revealed that even the least popular items with the lowest means played parts in teacher grading. Also, inconsistencies were observed when comparing the means and response percentages of various factors such as *academic performance* and *mastery of learning objectives*. For instance, while nearly one fifth of the teachers did not consider *mastery of specific learning objectives* for assigning grades, slightly over one third of them used it extensively to determine grades.

**PCA findings for secondary EFL teachers**

Table 2 summarizes the outputs of PCAs that generated four components with eigenvalues above one. However, only three components were retained and the fourth was discarded because only a single item loaded on it. The first component, *academic enablers* (McMillan, 2001), consisted of the largest number of items (eight items), the majority of which were nonacademic and contributed substantially to grades. It also explained the largest grading variance (41%). The second and third components included seven and three items each and explained 12% and 7% of the grading variance, respectively. The second component was labeled *external benchmarks and homework* (McMillan, 2001), since its underlying items, as shown in Table 2, drew on comparisons between teachers’ judgments and factors external to the classroom such as grade distributions of other teachers or students’ performances in previous years. The third factor termed *classroom-management grading* consisted of three items as *disruptive student behavior*, *zeroes for incomplete assignments*, and *extra credit for nonacademic performance*. Its underlying items related to the use of grades by teachers for reward or punishment or more broadly for punitive and behavioral purposes in class.

**Table 2** PCA outputs for secondary EFL teachers’ dataset

Grading item	Factor loading			
	1	2	3	4
<b>Factor 1: Academic enablers</b>				
Student effort-how much the student tried to learn	<b>.85</b>	.09	.22	.09
Improvement of performance	<b>.82</b>	.07	.20	.16
Class participation	<b>.80</b>	.30	.06	.18
Ability levels of the students	<b>.78</b>	.19	.20	-.19
Effort, improvement, behavior and other non-test factors for borderline cases	<b>.71</b>	-.03	-.01	.33
Academic performance as opposed to other factors	<b>.60</b>	.54	.04	-.17
Specific learning objectives mastered	<b>.59</b>	.41	-.25	.25
Work habits and neatness	<b>.48</b>	.27	.34	.30
<b>Factor 2: External benchmarks and homework</b>				
Performance compared to students from previous years	.08	<b>.77</b>	.30	.19
Performance compared to other students	.28	<b>.73</b>	.15	-.22
Performance compared to a set scale	.41	<b>.67</b>	-.13	.24
Grade distributions of other teachers	-.09	<b>.61</b>	.29	.15
School or district policy for grading	.132	<b>.58</b>	.48	.17
Completion of ungraded homework	.35	<b>.55</b>	.30	.26
Quality of graded homework	.52	<b>.53</b>	.10	.32
<b>Factor 3: Classroom-management grading</b>				
Disruptive student behavior	.15	.17	<b>.78</b>	-.10
Inclusion of zeros for incomplete assignment	.16	.13	<b>.74</b>	.07
Extra credit for nonacademic performance	-.01	.29	<b>.61</b>	.47
Extra credit for academic performance	.60	.19	.07	<b>.74</b>
<b>Eigenvalue</b>	7.82	2.34	1.38	1.08
<b>Percent of variance accounted for</b>	41.07	12.31	7.24	5.67
<b>Alpha reliability coefficient</b>	.90	.83	.71	na

Note. N = 187. Factor loadings above .40 are in bold. Rotation converged in 8 iterations

**Private EFL institute teachers' grading factors**

Table 3 presents the descriptive statistics and response percentages for the 19 grading items used by private institute teachers. The items that impacted teachers' grading the most (means > 4) were *improvement, effort, participation, graded and ungraded homework, ability, non-test factors for borderline cases, and mastery of learning objectives*. Except for the last item, *mastery of learning objectives*, the rest are considered nonacademic. Additionally, the means ranged from 2.14 to 4.50 (SD > 1), indicating that all the grading factors contributed, at varying degrees, to the grades assigned. Although a proportionately low mean of  $M = 2.38$  was reported for *student disruptive behavior*, the majority of teachers (80%) considered it to 'some' extent, 'quite a bit,' or 'extensively' in determining grades. None of the teachers believed that student conduct should be excluded from consideration when giving grades (*Not at all* = 0). Large

**Table 3** Descriptive statistics of grading items used by private EFL institute teachers

Grading item	Percentage						Descriptive		
	Not at all (1)	Very little (2)	Some (3)	Quite a bit (4)	Extensively (5)	Completely (6)	N	M	SD
1. Improvement of performance	0.0	5	15	20	44	16	113	4.50	1.10
2. Student effort-how much the student tried to learn	2	3	15	24	37	19	113	4.47	1.17
3. Class participation	2	11	13	23	30	21	113	4.32	1.35
4. Quality of graded homework	0.0	9	15	32	24	20	111	4.30	1.21
5. Ability levels of the student	0.0	8	20	23	32	17	113	4.29	1.20
6. Completion of ungraded homework	1	7	25	23	32	12	111	4.14	1.19
7. Effort, improvement, behavior, and other non-test factors for borderline cases	5	7	21	25	22	20	113	4.10	1.41
8. Specific learning objectives mastered	2	10	15	31	32	10	113	4.10	1.20
9. Extra credit for academic performance	10	14	18	16	27	15	109	3.80	1.57
10. Performance compared to other students	7	10	16	43	17	7	111	3.77	1.24
11. Academic performance as opposed to other factors	0.0	15	33	24	21	7	109	3.73	1.17
12. Work habits and neatness	3	16	20	35	23	3	111	3.68	1.16
13. Performance compared to a set scale	5	12	34	24	20	5	111	3.55	1.22
14. Inclusion of zeros for incomplete assignment	22	21	25	16	10	6	113	2.88	1.49
15. Performance compared to students from previous years	19	24	23	21	8	5	111	2.88	1.39
16. School or district policy for grading	23	24	23	14	16	0.0	107	2.77	1.37
17. Grade distributions of other teachers	29	23	21	19	3	5	104	2.59	1.41
18. Disruptive student behavior	0.0	20	40	30	4	6	111	2.38	1.05
19. Extra credit for nonacademic performance	52	12	16	12	7	1	110	2.14	1.40

standard deviations reported for each of the grading items were indicative of extensive grading variation among teachers.

**PCA findings for private EFL institute teachers**

The outcomes of factor analyses with Varimax rotation for private institutes’ dataset are summarized in Table 4. Four components with eigenvalues of at least 1 were extracted. The component with the largest number of items (8 items) was labeled *academic enablers* due to the dominance of nonacademic items loading on it. It accounted for the largest variance (36%) in teachers’ grading, nearly three times larger than the variance reported for component two. The next factor that explained 12% of the grading variance was labeled *external benchmarks* because most of its items (3 out of 4) focused on comparing student performance with external criteria such as set scales or student performance in previous years. Component three was termed *classroom-management grading* because the majority of items loading on it such as *extra credit for nonacademic performance, disruptive student behavior, and inclusion of zeros for*

**Table 4** PCA outputs for private EFL institute teachers’ dataset

Grading item	Factor loading			
	1	2	3	4
<b>Factor 1: Academic enablers</b>				
Student effort—how much the student tried to learn	<b>.83</b>	.01	.16	.20
Improvement of performance	<b>.82</b>	.05	.13	−.01
Ability levels of the students	<b>.79</b>	.27	.01	.01
Specific learning objectives mastered	<b>.75</b>	.21	−.03	.14
Class participation	<b>.70</b>	.36	−.11	.15
Quality of graded homework	<b>.67</b>	.31	.10	.30
Effort, improvement, behavior, and other non-test factors for borderline cases	<b>.65</b>	.06	−.05	.37
Completion of ungraded homework	<b>.48</b>	.15	.28	.16
<b>Factor 2: External benchmarks</b>				
Performance compared to a set scale	.31	<b>.79</b>	−.06	.02
Performance compared to students from previous years	.03	<b>.72</b>	.49	−.05
Performance compared to other students	.38	<b>.62</b>	.29	.13
Work habits and neatness	.33	<b>.56</b>	.02	.49
<b>Factor 3: Classroom-management grading</b>				
Extra credit for nonacademic performance	−.06	.03	<b>.67</b>	.07
Disruptive student behavior	−.01	.10	<b>.65</b>	−.04
Grade distributions of other teachers	.06	.46	<b>.61</b>	.32
School or district policy for grading	.18	.24	<b>.58</b>	−.39
Inclusion of zeros for incomplete assignment	.36	−.29	<b>.56</b>	.09
<b>Factor 4: Academic performance</b>				
Extra credit for academic performance	.30	−.01	.07	<b>.80</b>
Academic performance as opposed to other factors	.38	.42	.01	<b>.56</b>
<b>Eigenvalue</b>	6.82	2.36	1.49	1.14
<b>Percent of variance accounted for</b>	35.92	12.42	7.86	5.98
<b>Alpha reliability coefficient</b>	.90	.80	.68	.61

Note. N = 113. Factor loadings above .40 are in bold. Rotation converged in 8 iterations

*incomplete assignments* aimed at the specification of sanctions for student conduct in class. This component accounted for 8% of the variance in grading. The last component, *academic performance*, comprised only two items, *extra credit for academic performance* and *academic performance as opposed to other factors*, and explained the least variance in grading.

**Qualitative findings**

***Rationales behind hodgepodge grading***

Table 5 presents the themes and subthemes generated from the analysis of interview data along with interviewees’ quotes and occurrence percentages. The themes included (1) encouraging learning, (2) enhancing motivation, (3) lack of specific grading criteria, (4) pressure from stakeholders, and (5) flexible grading. The most frequently referenced theme (29.5%) was *encouraging learning* that was broken down into two subthemes of (a) *inseparability of achievement and enablers* and (b) *grades as payment for student*

**Table 5** Reasons why Iranian EFL teachers used nonacademic factors in grading (N = 30)

Theme and subtheme	Example quote	Frequency (%)
<b>Encouraging learning</b>	“Those students who participate more and try harder also learn better and more.”	<b>29.5</b>
Inseparability of achievement and enablers	“I guess it is wrong to think of enablers and achievement as separate entities because they feed on each other.”	
Grades as payment for student work	“In my opinion, a school is like a factory. Therefore, students should get paid for good work and punished for bad work. We [teachers] pay them grades.”	
<b>Enhancing motivation</b>	“Look, when the student knows that his/her efforts, abilities, or even class attendance are seen and counted by the teacher, definitely he/she will have more motivation to learn.”	<b>23.0</b>
Providing students with feedback	“In my idea, opening a discussion with a student about their grades and what they do in class that leads to those grades is the best way to let them know what their strengths and weaknesses are. Otherwise, they might not care that much what you say.”	
<b>Lack of specific grading criteria</b>	“Until now no one has given me any specific standards to base my grades on, maybe very generally.”	<b>16.4</b>
<b>Pressure from stakeholders</b>	“Many people, if not all, believe that their children should get better grades when they try more and are active. They drive you crazy if your grade doesn’t reflect this.”	<b>16.0</b>
Parents	“I am afraid of parents who come and talk to me about their son or daughter who failed even though he/she tried hard. They give me a lot of stress. They expect their children to be passed.”	
Students	“Students who regularly attend class or do their homework neatly expect to pass the course . . . no matter if they didn’t learn well.”	
School/institute administrators	“On several occasions the school principal has come to me saying: ‘If possible, let this student pass because he has good manners or is very neat.’”	
<b>Flexible grading</b>	“A teacher should not be strict in giving grades on achievement only. We live in a complex world. What are our grades supposed to change?”	<b>15.1</b>
Everything counts in grading	“I think many factors make a grade, not just one and the teacher has the responsibility to take as many factors into account to be fair.”	
Weakness compensation grading	“Considering ability, effort, or good behavior in grades can benefit those who perform poorly, but shouldn’t fail.”	

*work*. The second theme, *motivation enhancement*, focused on how the inclusion of nonachievement criteria in grading increased student motivation. It consisted of a sub-theme that was concerned with the role of feedback in motivating students. Together, the first two themes, learning encouragement and motivation enhancement, constituted the most important reasons as to why EFL teachers integrated nonacademic factors into their grading. The third theme, *lack of specific grading criteria*, was elicited from teachers' complaints about the absence of any grading guidelines to which they refer for grading. In teachers' opinions, the presence of such criteria could enhance grading by providing teachers, particularly novice teachers, with a frame of reference. *Pressure from stakeholders* was the fourth theme that constituted three subthemes that centered on pressure from (a) parents, (b) students, and (c) school/institute administrators. Finally, the *flexible grading* theme, which was mentioned the least by interviewees (15.1%) yielded two minor themes as (a) everything counts in grading and (b) weakness compensation grading.

#### ***Encouraging learning***

The majority of teachers believed that using nonacademic factors in grading, particularly enablers, enhanced learning. One teacher approved of this by saying: "Learning manifests itself through effort . . . . Where there is some effort, there should be some learning, too." Extensive use of nonachievement grading factors for learning was similarly endorsed by other teachers. "Those [students] who participate more and try harder also learn better and more," was an experienced teacher's response to why he valued effort in grading. Also, teachers thought that since improvement was the by-product of learning, failing to consider improvement in grading would discourage learning. One teacher rhetorically asked, "How can the teacher see improvement [in student work] and remain indifferent [to it]?" Even questioning the role of improvement or effort as grading criteria faced criticism by some teachers. For them, learning was the superordinate goal that justified teachers' reliance on various grading factors to determine grades. The analysis of additional comments produced the following subthemes.

**Inseparability of achievement and enablers** Several teachers believed that academic and nonacademic factors coalesced as a grading system and it was hard to separate them. For example, a teacher commented that "Effort, ability, improvement, and learning feed on each other and are interwoven." Another teacher pointed to the fusion of all grading factors this way: "I always thought effort meant improvement and improvement quite often meant learning . . . . like a chain . . . . Grading should capture all." The chain analogy demonstrates the inseparability of grading factors and justifies their use for advocating learning. Grading, for many teachers, was just a means to promote learning. Accordingly, one teacher remarked, "There's no effort without result and grades should reflect it [effort]." In a similar tone, another teacher declared, "Grades that do not take effort, improvement, participation into account have a limited meaning."

**Grades as payment for student work** Grades were seen by many teachers as payments in exchange for student effort. Teachers likened their grading to a type of transaction



between the work done and the grade earned. This was evidenced when a teacher explained, “In my opinion, a school looks like a factory. They [students] should get paid for good work and punished for bad work. We [teachers] pay them grades.” Other teachers approved of the grade-as-payment notion when emphasizing that they ‘pulled for students’ (McMillan, 2001) by raising their low grades in return for efforts expended, particularly in borderline cases. A teacher noted that she visualized her students and all their individual contributions to class when promoting failing grades, saying, “Students should reap what they sew during the term.” Similar comments constituted a significant portion of interview contents.

### ***Enhancing motivation***

The second important theme was using nonacademic grading factors as motivators of student learning. Teachers clearly indicated that integrating factors such as effort, improvement, and participation into grades raised student motivation to learn. One teacher commented: “If you mind your students’ efforts, they will be more motivated to attend the class.” Another teacher said: “The student who makes an effort that is reflected in her grade will be better motivated to attend class.” Even when the interviewer reminded a teacher that such amalgamation conflated grade meaning, he rhetorically responded: “How else can we appreciate students’ efforts *meaningfully* [emphasis added by the researcher] and keep their morale high if not by grades?” However, whether or not mixing academic and nonacademic factors into grades enhances motivation remains open to skepticism.

**Providing students with feedback** Some interviewees stated that considering nonacademic factors in grading broadened their opportunities to give students feedback they needed to stay motivated. A teacher commented, “Talking about their [students’] effort or how much they have improved makes my pupils want to do better and better.” Another teacher said that one of the most efficient ways for her to interact with learners about their performance was to hold conferences with them about what more they needed to do to improve and how this could influence their grades. Also, many teachers viewed grade-based interactions with students as chances to communicate their expectations of what mattered the most in their classroom assessments to students.

### ***Lack of specific grading criteria***

Another rationale for assigning amalgamated grades to students was *lack of specific grading criteria* that accounted for 16.4% of all elicited codes (see Table 5). Many teachers acknowledged that they had received no specific training in grading during teacher education programs. One teacher reported: “So far, no specific standards were given to me, or to any other teacher, to base our grades on.” In fact, some teachers looked perplexed when asked about official grading factors. One interviewee indirectly referred to teachers’ reliance on their gut feelings for assigning grades by stating: “Grades are based on what works to the best interest of students.” He added: “When you become a teacher, this is *you* [emphasis added] who should learn how to grade. It’s a trial and error game.” Few respondents referred to some forms of grading schemes proposed by heads of schools or institutes, but they failed to elaborate on them.

### ***Pressure from stakeholders***

As shown in Table 5, *pressure from stakeholders* was another reason given by interviewees to explain or justify their amalgamated grading. Students and parents exerted pressure on teachers to accommodate their grades. For instance, one teacher stated that: “Many parents, if not all, think that their children deserve higher grades when they appear to be trying harder. Some of them drive you [teachers] crazy if your grades do not reflect this [student effort].” Similarly, another teacher agreed that students who actively participated in class discussions or did their homework neatly expected to earn higher grades. Considering this, one of the teachers said: “They [students who made an effort] expect to get good grades, no matter if they did or didn’t learn enough.” One institute teacher approved of parental pressure by stating: “I’m afraid of parents who come and talk me into promoting their child’s failing grade when they think he/she should not have failed.” Furthermore, some teachers complained about school or institute administrators for pressurizing them to accommodate grades. An experienced teacher admitted that on several occasions the school principal had asked him to consider raising some students’ grades without legitimate reasons.

### ***Flexible grading***

The final theme was concerned with the use of nonacademic factors to ensure grading flexibility. Many teachers explained that they considered a wide variety of factors in their grading to maximize the chances for students to succeed. Accordingly, one teacher stated that she believed teachers should be “strict in teaching, but lenient in grading.” She clarified her argument by adding, “We live in a complex world and this complexity will be reflected in the factors influencing grades, too.” Teachers also believed that in order for grades to be equitable indicators of student performance, they should capture all that a student demonstrated in class. One teacher remarked: “I think many factors make a grade, not just one, to be as fair as possible.” Another teacher asked: “If grades should be based on achievement only, then how should student effort be appreciated?” Furthermore, some teachers considered nonacademic factors in grading as a strategy to compensate for weaknesses in students’ performances. Nonachievement factors gave teachers reasons to raise the grades of students who did not deserve receiving failing grades. One teacher commented, “Considering ability, effort, or good behavior in grades can benefit those who perform poorly, but shouldn’t fail.”

## **Discussion**

The purpose of this explanatory sequential mixed methods research (Creswell & Plano Clark, 2018) was to examine the grading practices of Iranian English language teachers in secondary schools and private EFL institutes. Specifically, the quantitative phase of this study aimed at identifying the factors teachers used to determine grades. The follow-up qualitative phase then elaborated on teachers’ rationales for assigning ‘hodge-podge grades’ (Brookhart, 1991) to students. The findings from both phases were subsequently integrated with the aim of providing more insight into EFL teachers’ grading decision-making.

### Hodgepodge grading reiterated

In response to research questions 1 and 2, the results of both descriptive and factor analyses showed that, contrary to measurement experts' recommendations, teachers attached the most weight to nonachievement factors when determining grades in both settings. This finding was not surprising and was reported in numerous earlier research (Brookhart et al., 2016; Duncan & Noonan, 2007; Guskey, 2011; Guskey & Link, 2019; Nowruzi & Amerian, 2020; Randall & Engelhard, 2009, 2010; Sun & Cheng, 2013; Yesbeck, 2011). What was surprising, however, was that achievement factors such as mastery of learning objectives and academic performance were quite marginalized in Iranian EFL teachers' grading practices, similar to what was reported in the Chinese EFL instruction context (e.g., Cheng & Sun, 2015; Sun & Cheng, 2013). This finding contrasted with what McMillan (2001) and McMillan et al. (2002) had reported where academic achievement was the main grading factor even when enablers' influences on grades were significant.

Academic enablers were found to be the primary grading component in this study. Contrary to what Guskey and Link (2019) reported, student effort had the heaviest weighting in determining grades here. This was consistent with Brookhart et al. (2016) who referred to effort as the "key element in grading" in their review (p. 22). From the social constructivist perspective, teachers may consider effort and participation as their primary grading criteria on the grounds that they believe engagement in learning is a true indicator of achievement or it contributes to learning. If this is true, Iranian teachers' grading practices tend to be more pedagogically oriented than measurement-oriented. Teachers appear to be primarily concerned with the consequences of grades for instruction and learning rather than grade meaning and use (Brookhart, 1993; Sun & Cheng, 2013). This would problematize teachers' high-stakes instructional decisions (DeLuca et al., 2017).

The second grading component used by Iranian EFL teachers, *external benchmarks*, centered on comparing students' current performances with their previous performance or with those of their peers. This component may be comparable with what Cheng and Sun (2015) termed *norm/objective-referenced factor*. However, class participation does not belong here, contrary to what Cheng and Sun reported. This might be because participation counts as some effort by the student to learn or to manifest learning and, therefore, is considered to be an enabling factor (McMillan, 2001). With regard to performance comparisons, what matters is the mechanism behind drawing such comparisons. The interview data suggested that it was highly improbable that student performances are compared systematically and objectively. It seems more likely that such comparisons are made subjectively, with reference to images formed in teachers' minds about students' past performances. Such mental representations are frequently influenced by teachers' beliefs and values about what counts as academic performance and thus, tend to be highly individualized (Randall & Engelhard, 2010).

Still another key grading component in the Iranian context was referred to as *classroom-management grading*. It appears that teachers employed grades for behavioral purposes with the ultimate goal of managing their classes more efficiently (Bonner & Chen, 2009; Brookhart, 1993, 1994; Nowruzi & Amerian, 2020). This can occur either directly by the inclusion of students' disruptive behavior in grading or indirectly by assigning zeros rather than partial credit for incomplete homework. In reality, teachers

tend to use grades to canalize student behavior with the intention of creating environments that are conducive to learning rather than to measure achievement. One possible explanation is that in teachers' views or even in the views of parents and students, effective classroom management pertains to teachers' competence in managing their classrooms. It is likely that assigning zeros instead of partial credit for incomplete homework is consistent with the punitive uses of grading (Dyrness & Dyrness, 2008; Reeves, 2004), whereas assigning extra credit for nonacademic performance is the other end of the spectrum, i.e., the use of grades for encouraging positive behavior.

Overall, the subjectivity of nonacademic factors on the one hand, and extensive idiosyncrasies in considering such factors for determining grades on the other hand, suggest that Iranian secondary and private institute teachers assign hodgepodge grades (Brookhart, 1991) of effort, improvement, ability, participation, and achievement. This poses a significant threat to the validity of the various interpretations and uses that stakeholders make of grades because grades no longer seem to communicate what students, parents, and even teachers expect them to communicate, i.e., achievement. At this point, it would be helpful to look at reasons for such hodgepodge grading from the teachers' own perspectives.

#### **Rationale behind Iranian EFL teachers' hodgepodge grading**

In response to the third research question as to why Iranian EFL teachers assign hodgepodge grades to student work, the qualitative analyses of the interviews revealed that teachers prioritized nonacademic factors in grading for five main reasons including learning encouragement, motivation enhancement, lack of specific grading criteria, pressure from stakeholders, and maintenance of grading flexibility.

Considering nonachievement factors in grading to encourage learning that was referred to in other studies as one of the rationales behind conflated grading (e.g., Kunnath, 2016; McMillan, 2001, 2003; Sun & Cheng, 2013) probably stems from teachers' belief in that there is a trade-off between the degree of engagement in learning activities and terminal learning outcomes. Such reasoning seems to be consistent with the social constructivist theory of learning. It appears that many teachers give priority to learning and use classroom assessment as a means of advocating further learning rather than measuring the extent of learning (McMillan & Nash, 2000; Sun & Cheng, 2013). As Kunnath (2016) mentioned, classroom assessment and grading is subsumed under teachers' overarching teaching and learning philosophy. Also, it seems that such a philosophy originates from teachers' beliefs and values that McMillan (2003) and McMillan and Nash (2000) referred to in their classroom assessment models. In other words, teachers' beliefs and values that are distilled from sociocultural and educational values of the society in which they live tend to play important roles in shaping grades as a by-product of classroom assessments.

The second reason for inflating grades, i.e., enhancing motivation, which was verified by previous research (Black & William, 1998; Brookhart, 1994; Crooks, 1988; McMillan, 2003; McMillan & Nash, 2000; Oosterhof, 2001) can be discussed in a similar vein. This finding is consistent with Kelly's (2008) warning that awarding failing grades results in poor motivation and low engagement in learning. Based on teachers' beliefs, it appears that as participation in class activities enhances learning, it can similarly raise

motivation. Therefore, encouraging learning and raising motivation are classified as two internal factors (McMillan, 2001, 2003; Simon et al., 2010) that are dependent on teachers' beliefs. However, what gains prominence, from the classical measurement theory perspective, is that formative assessments and grades arising from them will not problematize measurement as long as they are not used for summative purposes (Airasian, 2000). In other words, teachers should beware of acting as coaches and judges simultaneously (Bishop, 1992).

The third and fourth reasons for amalgamated grading, lack of specific grading criteria and pressure from stakeholders, can be seen as external factors (McMillan, 2003) that influence grades. Teachers do not act only on the basis of internal factors to make grade-based decisions; external factors such as parental pressure and the absence of distinct grading criteria are classroom realities that cause teachers not to put all their assessment eggs in the basket of achievement (Cheng & Wang, 2007; Davison, 2004). That is why teachers decide to consider an array of factors rather than a single factor in determining grades, a process which contributes to assigning multidimensional grades by combining different academic and nonacademic factors (Brookhart, 1993; Cheng & Sun, 2015; Nowruzi & Amerian, 2020).

The role of the fifth reason, flexibility in grading, gains special importance here. Teachers' flexibility in integrating various factors into grades can be interpreted as a leeway for them to strike a balance between internal and external forces, as reported by McMillan and Nash (2000). A number of studies referred to this as an effort by the teacher to assign fair grades to students (Kunnath, 2016; Sun & Cheng, 2013). Kunnath (2016) stated that integrating nonachievement factors into grades enhances fairness in grading. However, from the measurement experts' views, when grades reflect characteristics other than achievement, interpretations and uses arising from them are not valid and, most probably, such grades are not fair, as well. Thus, it appears that laxity in grading is an effort by the teacher to align the forces that shape grades rather than attempts to enhance grading fairness.

The fourth research question was concerned with how the qualitative findings provide a better insight into the quantitative results in this mixed methods study. The first point to mention is that qualitative findings explain why nonachievement factors have always been and will probably be an indispensable part of grades, even when teachers have been trained to base their grades on achievement or similar grading guidelines have been available (Cross & Frary, 1999; Duncan & Noonan, 2007; Guskey, 2009). Such findings show that one of the most influential internal factors that strongly influences grades is teachers' long-held beliefs and values that do not change overnight. The fact that grades are multidimensional (Bowers, 2009) lends itself to the impacts of strong internal and external factors that determine the nature of grades in the long run.

## Conclusion

Although this study offers new insights into EFL teachers' grading practices, some limitations exist. The first limitation is that this study addressed the grading practices of only EFL teachers. Broadening the scope of the study to include teachers teaching different subject matters and elementary school teachers can be more enlightening. The second limitation concerns participant selection for the qualitative phase. The results could have been even more reliable if the sample was selected using randomization,

rather than convenience sampling. Still another limitation concerns combining the grading findings of secondary EFL teachers in both junior and senior high schools. Iranian senior EFL teachers' grading practices are likely to be more heavily influenced by external factors such as the university entrance examination. Combining the results for all secondary teachers might have confounded the research outcomes.

### Implications and future directions

The implications of this mixed methods study are threefold. The first implication is that because grades were found to be inaccurate indicators of students' academic performance (Baird, 2013; Riley & Ungerleider, 2019; Smail, 2013), great caution should be exercised when using them for making summative instructional decisions. Future research should focus on finding ways to encourage teachers to critically evaluate their core educational beliefs and values and the impacts of such beliefs on grades they assign. Such introspection can help teachers become more measurement-oriented when utilizing classroom assessments. Secondly, the findings of this study provided concrete evidence that teachers used grades formatively to improve motivation and learning. Future research needs to tap on the distinction between formative and summative assessment types to foster transparency in grading and accountability in assessment. This can help minimize the risk of using the right assessment for the wrong purposes. Also, a re-conceptualization of traditional measurement theories to create classroom-friendly assessment packages can be on the agenda for any upcoming research (Brookhart, 2003; Moss, 2003). The third implication concerns the absence of grading standards. Providing teachers with non-prescriptive grading guidelines can help grades become more accurate indicators of achievement, resulting in more objectivity and fairness in grading.

### Abbreviations

CA: Classroom assessment; EFL: English as a foreign language; ESL: English as a second language; PCA: Principal component analysis; PCT: Personal construct theory

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40468-021-00145-2>.

**Additional file 1.** Teachers' classroom assessment and grading practices survey

**Additional file 2.** Grading interview protocol

### Acknowledgements

Special thanks to the editor and reviewers.

### Author's contributions

The author read and approved the final manuscript.

### Funding

No funding was received from any specific funding agencies.

### Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

### Declarations

#### Competing interests

The author declares that he has no competing interests.



Received: 4 August 2021 Accepted: 30 September 2021  
Published online: 19 October 2021

## References

- Airasian, P. (2000). *Assessment in the classroom: A concise approach*, (2nd ed., ). Boston: McGraw-Hill.
- Ary, D., Jacobs, L. C., Sorensen, C. K., & Walker, D. A. (2014). *Introduction to research in education*, (9th ed., ). Belmont: Wadsworth, Cengage Learning.
- Baird, J.-A. (2013). Judging students' performances. *Assessment in Education: Principles, Policy & Practice*, 20(3), 247–249. <https://doi.org/10.1080/0969594X.2013.812396>.
- Bishop, J. H. (1992). Why U.S. students need incentives to learn. *Educational Leadership*, 49(6), 15–18.
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>.
- Bonner, S. M., & Chen, P. P. (2009). Teacher candidates' perceptions about grading and constructivist teaching. *Educational Assessment*, 14(2), 57–77. <https://doi.org/10.1080/10627190903039411>.
- Bowers, A. J. (2002). Reconsidering grades as data for decision making: More than just academic knowledge. *Journal of Educational Administration*, 47(5), 609–629. <https://doi.org/10.1108/09578230910981080>.
- Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation*, 17(3), 141–159. <https://doi.org/10.1080/13803611.2011.597112>.
- Brennan, R. T., Kim, J. S., Wenz-Gross, M., & Siperstein, G. N. (2001). The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts Comprehensive Assessment System (MCAS). *Harvard Educational Review*, 71(2), 173–216. Retrieved from <http://hepg.org/her/abstract/104>. <https://doi.org/10.17763/haer.71.2.v51n6503372t4578>.
- Brindley, G. (2007). Editorial. *Language Assessment Quarterly*, 4(1), 1–5. <https://doi.org/10.1080/15434300701348268>.
- Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35–36. <https://doi.org/10.1111/j.1745-3992.1991.tb00182.x>.
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30(2), 123–142. <https://doi.org/10.1111/j.1745-3984.1993.tb01070.x>.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279–301. [https://doi.org/10.1207/s15324818ame0704\\_2](https://doi.org/10.1207/s15324818ame0704_2).
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12. <https://doi.org/10.1111/j.1745-3992.2003.tb00139.x>.
- Brookhart, S. M. (2004). *Grading*. Upper Saddle River: Pearson Education.
- Brookhart, S. M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69–90. <https://doi.org/10.1080/0969594X.2012.703170>.
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., ... Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848. <https://doi.org/10.3102/0034654316672069>.
- Cheng, L., & Sun, Y. (2015). Teachers' grading decision making: Multiple influencing factors and methods. *Language Assessment Quarterly*, 12(2), 213–233. <https://doi.org/10.1080/15434303.2015.1010726>.
- Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, 4(1), 85–107. <https://doi.org/10.1080/15434300701348409>.
- Cizek, G. J. (1996). Grades: The final frontier in assessment reform. *NASSP Bulletin*, 80(584), 103–110. <https://doi.org/10.1177/019263659608058416>.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research*, (3rd ed., ). Los Angeles, CA: SAGE Publications, Inc.
- Creswell, J. W., Plano Clark, V. L., Gutmann, M., & Hanson, W. (2003). Advanced mixed methods research designs. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook on mixed methods in the behavioral and social sciences*, (pp. 209–240). Thousand Oaks: Sage Publications.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481. <https://doi.org/10.3102/00346543058004438>.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, 12(1), 53–72. [https://doi.org/10.1207/s15324818ame1201\\_4](https://doi.org/10.1207/s15324818ame1201_4).
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305–334. <https://doi.org/10.1191/0265532204lt286oa>.
- DeLuca, C., Braund, H., Valiquette, A., & Cheng, L. (2017). Grading policies and practices in Canada: A landscape study. *Canadian Journal of Educational Administration and Policy*, 184, 4–22.
- Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research*, 53(1), 1–21.
- Dymness, R., & Dymness, A. (2008). Making the grade in middle school. *Kappa Delta Pi Record*, 44(3), 114–118. <https://doi.org/10.1080/00228958.2008.10516507>.
- Frary, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice*, 12(3), 23–30. <https://doi.org/10.1111/j.1745-3992.1993.tb00539.x>.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255–274. <https://doi.org/10.3102/01623737011003255>.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park: Sage.
- Guskey, T. R. (2009). *Bound by tradition: Teachers' views of crucial grading and reporting issues*. Paper presented at the Annual Meeting of the American Educational Research Association San Francisco, CA.
- Guskey, T. R. (2011). Stability and change in high school grades. *NASSP Bulletin*, 95(2), 85–98. <https://doi.org/10.1177/0192636511409924>.
- Guskey, T. R. (2015). *On your mark*. Bloomington: Solution Tree Press.

- Guskey, T. R., & Bailey, J. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks: Corwin.
- Guskey, T. R., & Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice*, 26(3), 303–320. <https://doi.org/10.1080/0969594X.2018.1555515>.
- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1), 3–20. <https://doi.org/10.1177/1525822X05282260>.
- Johnson, B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook on mixed methods in the behavioral and social sciences*, (pp. 297–320). Thousand Oaks: Sage Publications.
- Kelly, G. A. (1991). *The psychology of personal construct*. London: Routledge.
- Kelly, S. (2008). What types of students' effort are rewarded with high marks? *Sociology of Education*, 81(1), 32–52. <https://doi.org/10.1177/003804070808100102>.
- Kunnath, J. P. (2016). A critical pedagogy perspective of the impact of school poverty level on the teacher grading decision-making process. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 10007423)
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4–16. <https://doi.org/10.2307/1177052>.
- McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, 7(8). <https://doi.org/10.7275/5kc4-jy05> Available at: <https://scholarworks.umass.edu/pare/vol7/iss1/8>.
- 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>.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34–43. <https://doi.org/10.1111/j.1745-3992.2003.tb00142.x>.
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95(4), 203–213. <https://doi.org/10.1080/00220670209596593>.
- McMillan, J. H., & Nash, S. (2000). *Teacher classroom assessment and grading practices decision making*. In Paper presented at the National Council on Measurement in Education. LA: New Orleans.
- Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13–25. <https://doi.org/10.1111/j.1745-3992.2003.tb00140.x>.
- Nowruzi, M., & Amerian, M. (2020). Exploring the factors Iranian EFL institute teachers consider in grading using personal construct theory. *Journal of Teaching Language Skills*, 38(4), 123–164. <https://doi.org/10.22099/jtls.2020.36293.2780>.
- Oosterhof, A. (2001). *Classroom application of educational measurement*. Upper Saddle River: Prentice Hall.
- Pattison, E., Grodsky, E., & Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher*, 42(5), 259–265. <https://doi.org/10.3102/0013189x13481382>.
- Randall, J., & Engelhard, G. (2009). Differences between teachers' grading practices in elementary and middle schools. *The Journal of Educational Research*, 102(3), 175–186. <https://doi.org/10.3200/joer.102.3.175-186>.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26(7), 1372–1380. <https://doi.org/10.1016/j.tate.2010.03.008>.
- Rea-Dickins, P. (2004). Editorial: Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249–258. <https://doi.org/10.1191/026533204lt283ed>.
- Reeves, D. (2004). The case against the zero. *Phi Delta Kappan*, 86(4), 324–325. <https://doi.org/10.1177/003172170408600418>.
- Riley, T., & Ungerleider, C. (2019). Imputed meaning: An exploration of how teachers interpret grades. *Action in Teacher Education*, 41(3), 212–228. <https://doi.org/10.1080/01626620.2019.1574246>.
- Saito, H., & Inoi, S. I. (2017). Junior and senior high school EFL teachers' use of formative assessment: A mixed-methods study. *Language Assessment Quarterly*, 14(3), 213–233. <https://doi.org/10.1080/15434303.2017.1351975>.
- Simon, M., Tierney, R. D., Forgetter-Giroux, R., Charland, J., Noonan, B., & Duncan, R. (2010). A secondary school teacher's description of the process of determining report card grades. *McGill Journal of Education*, 45(3), 535–554. <https://doi.org/10.7202/1003576ar>.
- Small, E. (2013). Moderating New Zealand's national standards: Teacher learning and assessment outcomes. *Assessment in Education: Principles, Policy & Practice*, 20(3), 250–265. <https://doi.org/10.1080/0969594X.2012.696241>.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5–15. <https://doi.org/10.1111/j.1745-3992.2001.tb00065.x>.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teacher's hands: Investigating the practices of classroom assessment*. Albany: State University of New York.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5–14. <https://doi.org/10.1111/j.1745-3992.1989.tb00315.x>.
- Sun, Y., & Cheng, L. (2013). Teachers' grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326–343. <https://doi.org/10.1080/0969594x.2013.768207>.
- Svennberg, L., Meckbach, J., & Redelius, K. (2014). Exploring PE teachers' 'gut feelings'. *European Physical Education Review*, 20(2), 199–214. <https://doi.org/10.1177/1356336x13517437>.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks: Sage.
- Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39(1), 1–37. <https://doi.org/10.1111/j.1745-3984.2002.tb01133.x>.
- Woodruff, D. J., & Ziomek, R. L. (2004). *High school grade inflation from 1991 to 2003 (Research report series 2004–04)*. Iowa City: ACT.
- Yesbeck, D. M. (2011). *Grading practices: Teachers' considerations of academic and non-academic factors*. (Doctoral dissertation). Retrieved from ProQuest (913076079)

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.