Factors associated with Chinese undergraduates' perceptions of the ethical issues in classroom assessment practices—a mixed methods study

Ruiqin Gao[1] · **Xiumei Liu[2]** · **Xumei Fan[1]**

Abstract

Although studies have been conducted on educators' perceptions of assessment practices, few studies explored students' perceptions of the ethical issues in classroom assessment. A mixed design research method was used to examine factors associated with students' perceptions of the ethicality of classroom assessment practices. A sample of 1996 undergraduate students enrolled in 177 colleges and universities in China participated in the quantitative phase of the study to complete a survey measuring students' perceptions regarding the ethicality of classroom assessment. In the qualitative phase, 579 participants responded to the open-ended questions concerning their justification of the ethicality of the individual assessment situations. Quantitative analyses indicated that students' gender, grade level, major, and program were associated with their perceptions of the ethicality of multiple assessment practices (i.e., multiple assessment, surprise items, considering effort and attendance in grading, and giving feedback). Qualitative analyses showed that conflicting needs of different stakeholders in classroom assessment (i.e., student needs vs. assessment needs, teacher needs vs. assessment needs, student needs vs. student needs) were associated with their perceptions. Findings of the current study offer insights for teachers regarding how to make classroom assessment practices ethical based on diverse needs of stakeholders involved in assessment.

Keywords Classroom assessment · Undergraduates · Ethical issues · Mixed methods

✉ Ruiqin Gao
rgao@email.sc.edu

[1] Department of Educational Studies, University of South Carolina, Wardlaw 140, 820 South Main Street, Columbia, SC 29208, USA

[2] School of Law and Humanities, China University of Mining and Technology, Beijing, People's Republic of China

1 Introduction

Ethics is defined as the social responsibility of a profession as well as the personal responsibility of those practicing the profession (Oakland, 2004). Assessment is defined as a systematic process of using diverse methods to make decisions about what students know and can do (Brookhart & Nitko, 2018). Ethics in educational assessment is defined as the rules of behaviors that educators follow in assessment practices (Thorndike et al., 1991). A teacher's professional responsibility in assessment is to make appropriate decisions about students' learning by using high-quality information (Brookhart & Nitko, 2018). As assessment has a great impact on students' learning, motivation, curriculum development, and teaching process (Harlen, 2007; Lyon, 2013; White, 2009), educators should follow rules of behaviors and norms in their assessment practices to make an ethical decision of student learning (Johnson et al., 2008). Ethical assessment practices improve student learning and prevent potential harm to students (Green & Johnson, 2010). Therefore, morality and ethics should be considered in classroom assessment practices (Pope, 2006).

Ethical issues related to classroom practices have been examined in different countries including China (Fan et al., 2020), South Africa (Beets, 2012), England (O'Leary, 2008), Turkey (Özbek, 2013), Canada (Tierney, 2014), and the USA (Johnson et al., 2017; Pope, 2006; Gao et al., 2019, Gao et al., 2021). For example, student cheating is a common practice in higher education assessment of Australia and Turkey (Bretag et al., 2019; Yazici et al., 2011). The practice of narrowing curriculum to improve students' test score is common in countries such as the USA, England, and China (Berliner, 2011; Liu & Wu, 2006; OECD, 2013). Excluding a certain group of students from the test to increase the aggregate average score of schools is common in such countries as the USA, Canada, and Netherlands (Ferrer-Esteban, 2013). Maximizing student scores on all forms of evaluation is common in China (OECD, 2013). These previous studies indicated that the ethical issues in assessment are universal.

Given this situation, the continued exploration on this issue in specific culture is necessary. In particular, this topic is worth examining within the Chinese higher education context. Approximately 30.32 million undergraduate students were enrolled in degree programs at public colleges and universities in China in, 2019 (Textor, 2020a, 2020b). With such a huge number of undergraduates experiencing assessment in Chinese higher education, the investigation into their perceptions of the ethical issues in classroom assessment is significant for practitioners, policy makers, and educational researchers to have an in-depth understanding of the ethical issues and address the issues in appropriate ways. Besides, very little research has explored students' perceptions of the ethical issues in classroom assessment and factors affecting their viewpoints. To bridge the gap in literature, we investigated into Chinese undergraduates' perceptions of the ethical issues in classroom assessment and factors impacting their perceptions.

2 Theoretical framework

The present study draws on theoretical and empirical foundations in the areas of both ethics and assessment. Classroom assessment is an ongoing process in which teachers gather information and make decisions about student learning (Tierney, 2014). It is mainly composed of formative assessment and summative assessment. Formative assessment is defined as a process of monitoring student learning, providing feedback, and informing students of their learning during the course of a unit of instruction (Brew et al., 2009; Weurlander et al., 2012). In contrast, summative assessment is referred to as an activity of using assessment to summarize student achievement at the end of a unit of instruction (Boud & Falchikov, 2006; Hernández, 2012). Teachers are expected to practice assessment in alignment with standards and guidelines recommended by experts of educational assessment to improve student learning.

Ethics in assessment refers to the professional responsibility for gathering and using assessment information appropriately by following standards and ethical principle guiding assessment practices (Brookhart & Nitko, 2018). As teachers' assessment decisions are affected by diverse factors including teachers' gender, proficiency in the subject (Biberman-Shalev et al., 2011), teaching experiences (Duncan & Noonan, 2007), and students' gender and socioeconomic background (Lekholm, 2011; Peterson & Kennedy, 2006; Skelton et al., 2006), there has been a gap between the recommended and actual assessment practices (Alsarimi, 2000). Under this situation, ethical issues inevitably arise from assessment (Pope, 2006). Green et al. (2007) identified seven major categories of assessment practices involving ethical issues. They included bias/fairness (e.g., knowing students' identity while grading), communication about grading (e.g., using surprise items in assessment), multiple assessment methods (e.g., using single or multiple assessment formats), grading practices (e.g., considering student attendance/effort in grading), confidentiality (e.g., peer marking), standardized test preparation (e.g., training students in test-taking skills), and test administration (e.g., reminding students of something relevant to assessment).

Standards and guidelines have been developed for guiding teachers' assessment practices. Fairness is a major issue in classroom assessment. It is interpreted as giving students the opportunities to learn and/or demonstrate learning (Tierney, 2014). Unbiased and fair assessment is recommended by Joint Committee on Standards for Educational Evaluation (JCSEE, 2015). The principle of *do no harm* is recommended by Taylor and Nolen (2005). *Do no harm* suggests that teachers should avoid acting in a way that may cause harm to students or other individuals in schools. Standards such as meeting student needs, treating students with respect, incorporating the principle of fairness are also suggested (Airasian, 2005). Under these guidelines, practices such as addressing only students' strengths or weaknesses in feedback or considering students' effort or family background in grading are considered unethical.

With regard to the communication about grading, transparency is emphasized in the standards for classroom assessment (JCSEE, 2003). Teachers should

communicate clearly with students about the assessment (Airasian, 2000; Guskey & Jung, 2009; McMillan, 2011). Involving students in the assessment process was also recommended (Chappuis & Stiggins, 2002; Falchikove, 2013; Gao et al., 2019; Green & Johnson, 2010). Stiggins and Chappuis (2005) suggested sharing expectations and criteria with students. Under these guidelines, the assessment practice of sharing grading rubrics with students is considered ethical while using surprise items in a test is considered an unethical practice.

Assessment decisions should provide a complete picture of student achievement (Tierney, 2014). As most students do not perform consistently across assessments, a variety of assessment formats should be used for sufficient information (Gronlund, 2003). Therefore, multiple assessment methods are recommended as they help teachers make good decisions about students' achievement (e.g., Brookhart & Nitko, 2018; Rasooli et al., 2019; Smith, 2003; Tieney, 2016). Besides, students need different opportunities to demonstrate their learning in different ways (Heritage, 2013). According to these guidelines, using multiple assessment formats is an ethical practice while adopting a single assessment format is considered unethical.

With regard to grading practices, teachers should score students accurately and fairly (Brookhart & Nitko, 2018). *Avoid score pollution* is suggested for student assessment (Haladyna et al., 1991; Popham, 1991). *Avoid score pollution* requires that assessment results should accurately reflect students' actual mastery of content. Students' scores are polluted by factors unrelated to achievement (Green & Johnson, 2010). Non-achievement factors including attitude, effort, attendance, and student ability should not contribute to grading directly (Brookhart, 2004; Oosterhof, 2009). Therefore, it is an unethical practice to consider students' attendance in grading or give minimum scores to students regardless of their achievement level.

Confidentiality in assessment requires that only persons who have authorized rights can have access to student scores (JCSEE, 2003). Test developers should develop and implement procedures for ensuring the confidentiality of scores (Joint Committee on Testing Practices, 2004). Teachers should protect the confidentiality of information that identifies individuals and the rights to privacy of individuals involved in the assessment process (NCME, 1995). Under these guidelines, practices such as peer marking and disclosing students' scores are unethical.

Test administration is one of the important assessment-related activities. Those who administer assessment should ensure that assessment is administered fairly and accurately (Brookhart & Nitko, 2018). Procedures for test administration should be standardized and constant across test users and test takers (Kline, 2013). Popham (2017) discussed that indicating in any way that students' answers may be wrong during test administration is inappropriate. Therefore, it is unethical practice for teachers to remind students of anything relevant to assessment during test administration.

The investigation into students' view of the ethicality of assessment practices is significant for improving the quality of classroom assessment practices as well as student learning. Strong agreement among students regarding the ethicality of assessment practices would suggest that student needs in assessment are well met. Evidence of weak agreement would indicate that further dialogue between different assessment stakeholders is necessary. As students are the direct participants in

classroom assessment practices, teachers need to consider their perceptions in order to better facilitate student learning and communicate better with students the benefits and drawbacks of certain assessment practices.

3 Literature review

3.1 Ethical issues in assessment

The ethical issues occurred in assessment has motivated researcher to conduct study on relevant topic. Previous studies focused on different assessment stakeholders' perceptions of the ethical issues in classroom assessment. The stakeholders included pre-service teachers (Bergman, 2013; Bergman, 2018; Bergman, 2020; Fan et al., 2019; Liu et al., 2016; McGlory, 2013), in-service teachers (Green et al., 2007; McGlory, 2013), and the educational leaders (Johnson et al., 2008) and students (Fan et al., 2020). In particular, studies compared the perception of pre-service teachers in China and the USA regarding the ethical issues in assessment (e.g., Fan et al., 2019, Liu et al., 2016). Studies also examined Chinese university professors' view of the ethical issues in classroom assessment (e.g., Fan et al., 2017; Fan et al., 2020). Cirlan (2017) examined Finnish pre-service and in-service teachers' perceptions about the ethicality of the classroom assessment practices. The study by Fan et al. (2020) explored Chinese college students' perception of the ethical issues in classroom assessment. All these study results indicated that there was no consensus regarding ethicality of the assessment practices among different stakeholders involved in assessment.

Besides the ethicality of the assessment practices, studies also examined ethical dilemmas in classroom assessment. Teachers' ethical dilemmas in classroom assessment mainly arose from the conflicts between teachers' perceptions of institutional demands and student needs (Pope et al., 2009). Johnson et al. (2017) developed an ethical decision-making model demonstrating how to deal with ethical dilemmas. This ethical decision-making model was validated with authentic classroom assessment scenarios (Gao et al., 2019) and expanded as well (Gao et al., 2021). These models examined the common ethical dilemmas in assessment, guidelines used for resolving the ethical dilemmas, and the impact of teacher's decision-making. The study by Beets (2012) in South Africa suggested infusing principles of ubuntu to strengthen ethics in educational assessment. These studies contributed a lot in offering guidance to educators in considering diverse elements involved in ethical dilemmas in classroom assessment.

Among all these studies on the ethical issues in classroom assessment, only the study by Fan et al. (2020) examined students' views of ethical issues in classroom assessment. While assessment was used as a way to determine student learning (Earl, 2012; Sambell et al., 2012), it is also used for student learning (Dann, 2014; Earl, 2012). Students' perceptions of assessment methods are important as they affect the ways students learn (Lizzio & Wilson, 2013), their engagement with the course materials (Gibbs, 2006), and the involvement in learning process (Biggs, 2003). Therefore, it is of great necessity to use different samples and

data resources to further explore how students perceive the ethicality of classroom practices and what are the factors affecting their perception. The knowledge of student perceptions and factor affecting their perceptions is likely to improve learning and assessment in classroom.

3.2 Students' perceptions of assessment in higher education

Different aspects of student learning including what they attended to, how much work they did, how they spent their time, and how they prepare for an assessment were determined by how they perceive assessment (Brown & Hirschfeld, 2008; Gielen et al., 2003). Students' perceptions of assessments have both positive and negative impact on learning (Brown & Hirschfeld, 2008; Loyens et al., 2008). Negative feelings toward assessment such as stress, anxiety, and fear reduce students' academic achievement (Craddock & Mathias, 2009).

Studies have investigated students' perceptions of different aspects of assessment in higher education within different cultural contexts. For instance, Pan (2020) examined Taiwan university students' perceptions of summative and formative classroom assessment in English courses. The study indicated that students have more positive perceptions of the summative assessment and cooperative group assessment, and a combination of summative and formative assessment tasks was beneficial for student learning. Göktürk Saglam and Yalçin Duman (2020) explored undergraduates' perceptions of the source-based writing assessment with a Turkish English for Academic Purposes (EAP). A positive association was identified between students' perceptions and students' writing proficiency. Brown and Wang (2013) found that undergraduates in Hong Kong showed strong awareness of the evaluative and controlling role that assessment plays in their lives. Aldrich et al. (2018) found that students attending face-to-face course believed that oral presentations were more effective for student learning than those in a hybrid class format.

These previous studies explored students' perceptions of diverse aspects of the assessment. To our knowledge, the study by Fan et al. (2020) is the only one which explored students' perceptions of the ethical issues in classroom assessment. However, this study had three noteworthy limitations. First, the study employed only quantitative data analysis, so a qualitative method of research is necessary for further investigation. The use of a mixed-methods approach in the current study should provide a more reliable and valid picture of these ethical issues. Second, the study included both undergraduates and graduates, and these groups have different learning foci and different assessment formats. Some assessment situations in undergraduate classrooms may not apply to graduates, so the current study mainly focuses on undergraduates. Third, the study focused on students' views of the ethical issues in classroom assessment practices without exploring the reasons why students have different perceptions. The current study mainly explores factors associated with students' perceptions regarding the ethicality of the classroom assessment practices using both quantitative and qualitative data.

3.3 Factors associated with student's perceptions of assessment

As students' perceptions of assessment are associated with their learning process and finally with their academic success (Struyven et al., 2006), studies were conducted to explore whether various student characteristics affect their perceptions of different aspects of assessment. With regard to gender, female students tended to prefer constructed-response assessment format than male students do (Aldrich et al., 2018). Male and female students have different perceptions regarding the scaffolding strategies in formative mathematics assessment (Wafubwa & Ochieng, 2021). Male and female students showed differences in how they perceived group work. Specifically, female students are more likely to think that efforts are not fairly apportioned in groups (Ludlum et al., 2021). Gao (2012) found that female students have more positive perceptions regarding assessment authenticity and transparency. Wurf and Povey (2020) discussed that boys tended to perceive that assessments were more transparent and congruent with their learning.

Students of different age groups differ in their perceptions of writing assessments and multiple-choice questions formats (Aldrich et al., 2018). Upper-grade level students tended to have a more positive perception of assessment with regard to the congruence with planned learning and transparency in assessment (Dhindsa et al., 2007). Students' perceptions of assessment differ as a function of subject area (Iannone & Simpson, 2013). Business majors were more likely to work longer hours and be assigned group projects in their classes relative to non-business majors (Ludlum et al., 2021). Besides, students' gender, grade level, and subject areas have interaction effects on congruence with planned learning and transparency of assessment (Alkharusi & Al-Hosni, 2015).

The aforementioned studies showed that students' characteristics such as gender, grade level, and subject area have an impact on their perceptions of assessment. As ethical issue is an important aspect of classroom assessment, these personal factors may also impact their perceptions of ethical issues in classroom assessment. If students from different groups favor different assessment practices that produce the same learning outcome, knowing students' preference may help teachers and administrator make appropriate decisions regarding which assessment practices to choose. Therefore, the present study intended to extend previous research by questioning how students' perception of ethical issues in classroom assessment varies as a function of gender, grade level, major, program, and additional potential factors.

3.4 Assessment context in China

In Chinese education system, the evaluative function of assessment was greatly emphasized as schools focus much on students' success in high-stakes assessments used to select student for entry to better schools or for valuable opportunities for further improvement (He et al., 2011; Liu and Qi, 2005; Niu, 2007). Assessment has long been used as a means of improving social and personal life (Ye et al., 2007) and motivating effort (Kennedy et al., 2008). Summative assessments have been

used as the primary assessment format (Cheng et al., 2015). However, this traditional assessment was criticized as it failed to represent the overall students' abilities. With curriculum reform in China, the idea of assessment for all-round development of students was highly advocated. The new assessment system goes beyond the high-stakes exams to advocate the holistic learning progress of students. With this transformation, teachers might experience a conflict between the traditional and new ethical principles guiding their assessment practices. Knowledge of students' perceptions may provide insightful information for teachers' professional development and ethical assessment practices. Besides, as students are the direct participants of the assessment, knowing their perceptions may promote the conversation between teachers and students regarding how to use assessment to facilitate student learning.

3.5 Purposes and research questions

Based on the previous research studies and the Chinese university context, the purpose of this study is to investigate factors associated with Chinese college students' perceptions regarding the ethicality of assessment practices. The quantitative phase of the current study focused on how students' demographic characteristics (gender, grade level, major, and program) are associated with their perceptions of the ethicality of assessment situations. The qualitative phase used open-ended questions concerning students' justification of their decisions to explore additional factors. This study sought to answer the following questions:

- How do Chinese undergraduates perceive the ethical issues in classroom assessment?
- What are the factors associated with Chinese undergraduates' perceptions of the ethical issues in classroom assessment?

4 Methods

This study used a convergent parallel mixed-method design to provide a unified picture of factors associated with students' perception of the ethical issues in assessment practices. A convergent parallel design entails that a researcher concurrently collects both quantitative and qualitative data, analyzes the data separately, and compares or relates the quantitative statistical results and qualitative findings (Creswell & Creswell, 2017). The quantitative data were used to learn about how students perceived the ethicality of the assessment scenarios, and how their demographic characteristics were associated with their perceptions. The qualitative method was used to find out additional factors associated with their perceptions.

4.1 Participants

Participants in this study consisted of 1996 undergraduate students enrolled in 177 colleges and universities from 23 provinces and 4 municipalities in China. Among

these participants, 1359 (68%) were females, and 637 (32%) were males. Approximately 75% ($n = 1493$) were freshmen and sophomores, and 25% ($n = 503$) were juniors and seniors. About 67% ($n = 331$) of the students majored in humanities and social sciences, and 33% ($n = 1656$) majored in STEM. About 88% ($n = 1769$) of participants were in a non-teacher preparation program, and 11% ($n = 226$) were in a teacher preparation program. All these participants participated in the quantitative study.

A total of 579 out of 1996 participants responded to the follow-up open-ended questions. Among these participants, 36.2% ($n = 149$) were males and 63.8% ($n = 263$) were females; 72.1% ($n = 297$) of participants were freshmen and sophomores, and 27.9% ($n = 115$) were juniors and seniors, while 62.4% ($n = 257$) and 37.6% ($n = 155$) were majored in humanities and science respectively. Most of the participants (92.2%) were from non-teacher preparation program.

4.2 Data collection

Participants were selected based on two criteria. First, each participant must be an undergraduate student participating directly in classroom assessment practices. Second, we selected participants from different grade level, major, and programs by applying a maximum variation sampling strategy. Researchers emailed faculty members, student service staff, and students in Chinese universities, explaining the purpose of the study and stating that their participation would be voluntary. Survey-Monkey link was then sent to potential participants. Data collection was conducted in the spring semester of 2018. The final response rate was 62%.

The quantitative data were collected from a survey questionnaire with 15 scenarios about ethical issues in classroom assessment (see Appendix Table 3). The survey was developed based on scenarios developed by Green et al. (2007), Johnson et al. (2008), Liu, et al., (2016), and Fan et al., (2017). These scenarios were developed using resources from the assessment literature, researchers' experiences, and anecdotes gathered from students (Green et al., 2007). For example, one scenario states that "To enhance self-esteem, a professor addresses only students' strengths when giving feedback to her students' assignments since she believes that positive feedback is good for students' growth. The scenarios were revised to fit the classroom assessment context in Chinese higher education settings. The experts' view of the ethicality of the scenarios was used as a reference point to discuss students' perceptions. Experts' views were obtained from the review of the textbooks, teaching standards, and published journal articles (e.g., Brookhart & Nitko, 2018; Gronlund, 2003; Popham, 2017; Taylor & Nolen, 2005). The original scale was a 4-point Likert scale. The data were analyzed by dichotomizing them into ethical and unethical. A higher score from the survey indicated a student was more likely to agree with experts on the ethicality of assessment practices. As experts viewed three scenarios as ethical and the other twelve ones as unethical, we reverse-coded the twelve items so a higher score indicated higher agreement with experts. Demographic information, including gender, grade level, major, and program, was obtained as well.

Qualitative data were collected from open-ended questions regarding students' decision of the ethicality of classroom assessment scenarios and the justification of their decision (e.g., "how do you justify your decision on the ethicality of this scenario").

4.3 Data analysis

For quantitative study, logistic regression analysis was performed to explore the association of students' gender, grade level, major, and program with their perceptions. Logistic regression is a method used for analyzing a dataset in which dependent variables entail two possible answers, and which explores the relationship between one or more explanatory variables and binary outcome variables (Hosner & Lemeshow, 1989). The logistic regression model is a robust model as it does not require independent variables to be normally distributed or to have equal variances in each group. In addition, it does not assume a linear relationship between independent variables and dependent variables (King, 2003). In the current study, four blocks of independent variables (gender, grade level, major, and program) were successively entered into the model. Students' decisions of the ethicality of each scenario were regressed on the group variables. The logistic regressions were repeated for individual assessment scenarios to explore the relationship between students' perceptions of the ethicality of each assessment scenario and their demographic characteristics. All statistical analyses were carried out using Statistical Packages for Social Science (Version 26). A two-tailed p value of < 0.05 was considered statistically significant. The odds ratio was used to estimate the change in the odds of membership in the dependent variable for every one-unit increase of the independent variable. If the odds ratios are significant at the 0.05 level, the 95% confidence interval does not contain 1.

For the qualitative data analysis, we followed procedures in qualitative data analysis suggested by Creswell and Creswell (2017). First, two research team members organized all the responses in Excel for data analysis and read responses to get a general view of students' perceptions of each scenario. Second, we created a codebook after reading all the responses to the open-ended questions. Third, we coded the responses independently by following the same framework outlined by the codebook. We created 15 separate Excel sheets to record students' responses to each scenario. Within each scenario sheet, we created two columns recording students' evaluations of the ethicality (ethical and unethical). Within each ethicality category, we color-coded the responses to extract the main themes (e.g., assessment needs or student needs). Fourth, we combined all the responses to each scenario theme by theme (e.g., assessment needs, teacher needs, and student needs). Fifth, we read across the ethical categories and themes to find out why students perceived each scenario in a different way (e.g., conflict between student needs and assessment needs). The same coding procedure was repeated for each of the scenarios.

After two rounds of independent coding, a third team member cross-checked the result. All team members discussed any discrepancies and subsequently came up with the final coding results.

5 Results

5.1 Quantitative analysis results

One of the objectives of the study is to model students' perceptions of the ethicality of the classroom assessment scenarios. To evaluate factors that contribute to the probability of the occurrence of Ethical or Unethical level, the rate of estimation is calculated. Due to the specific characteristics of individual scenarios, we hypothesized that students from different demographic backgrounds would have different decisions of the ethicality of each scenario. Logistic regressions were repeated for individual assessment scenarios. The study revealed gender, grade level, major, and program significantly affect students' perceptions of the ethicality of 10 of the 15 assessment situations (Table 1).

Gender was significantly associated with students' perceptions of the ethicality of 7 assessment situations. Female students were more likely to agree with experts in terms of ethicality of Scenario 8 ("Considering effort in grading") (OR = 1.87, 95% CI = 1.51–2.32), Scenario 9 ("Showing score sheets to students") (OR = 1.38, 90% CI = 1.12–1.71), Scenario 10 ("Bumping students' grade") (OR = 1.54, 95% CI = 1.26–1.89), Scenario 11 ("Sharing rubrics with students") (OR = 1.57, 95% CI = 1.13–2.20), and Scenario 14 ("Using single test format") (OR = 1.35, 95% CI = 1.10–1.68). Male students were more likely to agree with experts regarding the ethicality of Scenario 4 ("Giving a student a higher grade based on mastery level rather than assignment submission status") (OR = 0.77, 95% CI = 0.60–0.98), and Scenario 15 ("Counting attendance as part of final grades") (OR = 0.70, 95% CI = 0.52, 0.95).

Students' major had a significant relationship with students' perceptions of the ethicality of 5 assessment situations. Students majoring in STEM were more likely to agree with experts than those majoring in humanities and social sciences regarding the ethicality of Scenario 4 ("Giving a student a higher grade even based on mastery level rather than the assignment submission status") (OR = 1.35, 95% CI = 1.01–1.73), Scenario 5 ("Using surprise item") (OR = 1.57, 95% CI = 1.25–1.98), Scenario 8 ("Considering effort in grading") (OR = 1.54, 95% CI = 1.24–1.92), Scenario 9 ("Showing score sheets to students") (OR = 1.34, 95% CI = 1.08–1.67), and Scenario 13 ("Reminding students to check answer in test administration") (OR = 1.39, 95% CI = 1.11–1.74).

Grade level was significantly associated with students' perceptions of the ethicality of 3 assessment situations. Higher grade level students (juniors and seniors) were more likely to agree with experts than lower grade level students (freshmen and sophomores) regarding Scenario 1 ("Giving only positive feedback") (OR = 1.28, 95% CI = 1.05–1.57), Scenario 8 ("Considering effort in grading") (OR = 1.29, 95% CI = 1.04–1.61), and Scenario 9 ("Showing score sheets to students") (OR = 1.46, 95% CI = 1.17–1.82).

Being in a teacher preparation program significantly affected students' perceptions of one assessment situation. Students who were not in a teacher preparation program were more likely to agree with experts than those who were in the

Table 1 Binary logistic regression model (significant variables)

Assessment scenarios (expert view)	Gender (ref = male) OR (95%CI)	Major (ref = humanities) OR (95%CI)	Grade (ref = lower grade) OR (95%CI)	Program (ref = non-teacher program) OR (95%CI)
Scenario 1 (Unethical)	1.05 (0.86,1.30)	1.00 (0.81,1.23)	*1.28 (1.05,1.57)	1.16 (0.87,1.55)
Scenario 4. (Ethical)	*0.77 (0.60,0.98)	*1.35 (1.005,1.73)	1.06 (0.83,1.35)	*0.70 (0.51, 0.95)
Scenario 5. (Unethical)	0.83 (0.70,1.05)	***1.57 (1.25,1.98)	0.79 (0.62,1.02)	0.90 (0.63,1.28)
Scenario 8. (Unethical)	***1.87 (1.51,2.32)	***1.54 (1.24,1.92)	*1.29 (1.04,1.61)	1.07 (0.80,1.45)
Scenario 9. (Unethical)	**1.38 (1.12,1.71)	**1.34 (1.08,1.67)	**1.46 (1.17,1.82)	0.978 (0.73,1.32)
Scenario10. (Unethical)	***1.54 (1.26,1.89)	0.97 (0.79,1.20)	1.13 (0.62,1.39)	1.05 (0.79,1.4)
Scenario 11. (Ethical)	**1.57 (1.13,2.20)	1.23 (0.86,1.76)	1.16 (0.81,1.67)	0.98 (0.60,1.58)
Scenario 13. (Unethical)	1.13 (0.91,1.41)	**1.39 (1.11,1.74)	1.17 (0.94,1.46)	0.88 (0.66,1.19)
Scenario 14. (Unethical)	**1.35 (1.10,1.68)	1.09 (0.88,1.36)	1.05 (0.84,1.30)	1.37 (1.0,1.88)
Scenario 15. (Unethical)	*0.70 (0.52,0.95)	0.93 (0.67,1.28)	1.23 (0.90,1.67)	1.31 (0.87,1.99)

OR odds ratio, *95% CI* 95% confidence interval

$*p < 0.05$; $**p < 0.01$; $***p < 0.00$

program regarding the ethicality of Scenario 4 ("Giving a student a higher grade based on mastery level rather than the assignment submission status") (OR = 0.7, 95% CI = 0.51–0.95).

5.2 Qualitative analysis results

The quantitative data results showed students across groups differed on the ethicality of 10 assessment scenarios. In this qualitative phase, we focused on these 10 assessment scenarios to explore additional factors affecting their perceptions and their justification of their decision on the ethicality of assessment practices (Table 2).

Students viewed the ethicality of assessment scenarios from the perspectives of different needs of stakeholders involved in assessment. When the needs of stakeholders intersected, the disagreement among students regarding ethicality occurred accordingly. The conflict mainly came from the conflicts between student needs and assessment needs, teacher needs and student needs, different student needs, and different assessment needs.

Regarding Scenario 1 ("To enhance self-esteem, a professor addresses only students' strengths when giving feedback to her students' assignments since she believes that positive feedback is good for students' growth"), Students had split opinions. A total of 255 (62%) of participants agreed with experts that the practice was unethical. They justified their viewpoint by saying positive-only feedback gave students a false picture of their learning and may stimulate students to become arrogant. Students need negative feedback to learn to deal with frustrations and to reflect on their progress, etc. For example, one student stated, "*It (positive-only feedback) will make students have a false perception of themselves, leading students to feel satisfied and confident without being able to realize their own drawbacks.*" The other 153 students (38%) thought the practice was ethical as positive feedback helped enhance students' self-confidence, self-esteem, and learning motivations. The disagreement arose from the conflict between student needs themselves. While students need encouragement from teachers, they also need to know their weaknesses for their overall development.

For Scenario 4 ("As a professor finalizes grades, she notices the grade of a student is in between B + and an A − . She gave the student an A − because tests and papers showed the student had mastered the course objectives even though he had not completed some of his homework assignments"), 316 (77%) students agreed with experts the practice was ethical. The disagreement arose from the conflict between assessment needs and student needs, and assessment needs themselves. From the perspective of assessment, assessing students' mastery level of knowledge should be the final purpose. Multiple assessment formats provide a more comprehensive picture of students' strengths and weakness. For example, one student responded, "*teachers should practice multiple ways of assessment. Single assessment provides a limited source of information, which is not good for either students or teachers.*" However, other students thought teachers should follow the rubric strictly to maintain objectivity and fairness to all students. Different assessment needs conflicted in this case. When viewing the assessment from the perspectives of student needs,

Table 2 Student justification of the ethicality of the assessment practices

Assessment scenarios(expert view)	Ethical	Unethical	Conflicts
Scenario 1 (Unethical)	1. Positive feedback enhances students' self-confidence, self-esteem, and learning motivation	1. Positive-only feedback will give students a false picture of their learning and may make them become arrogant	Student needs *vs*. Student needs
		2. Negative feedback is good for improving students' anti- frustration ability and reflection ability	
Scenario 4 (Ethical)	1. Mastering the learning goal should be the primary purpose of assessment	1. Completion of assignments reflects students' attitude toward learning	Assessment needs *vs*. Student needs
	2. Using multiple assessment formats provides a more comprehensive picture of a student's strengths and weaknesses	2. Teachers should follow rubric strictly to maintain objectivity and fairness	
Scenario 5 (Unethical)	1. Surprise items motivate students' active learning and improve their adaptability	1. Surprise items may hurt students' self-confidence and increase their test anxiety	Student needs *vs*. Student needs Student needs *vs*. Assessment needs
	2. Including surprise items measures what students have really learned from daily classroom learning activities rather than what guidelines cover	2. Surprise items might be unfair for those students who did not focus on the learning points covered by surprise items	
Scenario 8 (Unethical)	1. Considering students' effort motivates them to increase effort	1. Teachers' perception of effort might be arbitrary, subjective, and unfair	Student needs *vs*. assessment needs
	2. Effort is a good supplement to the final exam to assess the overall performance of students	2. Assessment should primarily reflect students' mastery of learning goals	
Scenario 9 (Unethical)	1. Knowing other students' scores helps students know more about their performance by comparing themselves with other students	1. Teacher should take the responsibility to protect students' privacy	Student needs *vs*. student needs
	2. Knowing other students' scores motivates students to work harder	2. It may harm students who do not perform well	

Table 2 (continued)

Assessment scenarios(expert view)	Ethical	Unethical	Conflicts
Scenario10 (Unethical)	1. The practice showed teachers' concern and encouragement for students	1. The practice is not fair to other students 2. Assessment did not accurately reflect students' performance	Teacher needs vs. assessment needs
Scenario 11 (Ethical)	1. Involving students in rubric development makes the rubric more student-oriented, objective, fair, and easy to be followed 2. Involving students in rubric development promotes student–teacher relationship	1. Involving students in rubric development increases the subjectivity of rubric 2. Involving students in rubric development fosters rubric-oriented learning	Student needs vs. assessment needs
Scenario 13 (Unethical)	1. Reminding students of the missed items is fair to everyone if the teacher reminds all students	1. Teachers should follow the test administration rules 2. It is not fair to students in other test rooms 3. Reminding weakens the function of test	Assessment needs vs. teacher needs
Scenario 14 (Unethical)	1. Multiple choice assessment format mainly measures the low cognitive skills of students 2. The probability of guessing the multiple-choice items right is higher than that of other assessment formats, thus reducing the validity of the assessment	1. It is objective, easy to administer, and easy to grade	Assessment needs vs. assessment needs
Scenario 15 (Unethical)	1. Attendance is irrelevant to students' mastery of knowledge 2. Twenty percent gives too much credit and may appear to obligate students to attend class, which may hinder students' active learning	1. Attendance stimulates students to learn, motivation, regulate themselves, and be punctual 2. Attendance reflects students' attitude toward learning 3. Attending class shows students' respect for classmates, teachers, and knowledge	Assessment needs vs. teacher needs Assessment needs vs. student needs

they thought the completion of the assignment reflected a student's attitude toward learning, which was important for overall development. The conflict between student needs and assessment needs makes students have split opinions.

With regard to Scenario 5 ("For the class-level final exam, a professor uses a few surprise items about additional topics that were covered in class but were not listed in the study guide"), only 21% of participants agreed with experts that the assessment situation was unethical. This group of students justified their choice that surprise items may hurt students' self-confidence and increase students' test anxiety. In addition, it may be unfair to other students who have not focused on the knowledge assessed by the item. Students who treated the practice as being ethical thought surprise items can enhance students' learning motivation and adaptability. Different student needs intersected at this point. Moreover, using surprise items is a good way to evaluate students' mastery of knowledge uncovered in a test guideline; the final purpose of assessment is to see whether students have achieved the learning goal rather than something only covered in the guideline. This disagreement originated from the conflict between student needs and assessment needs.

While discussing Scenario 8 ("In grading a final exam, a professor always reads the student's name and considers effort in assigning grades"), 70% of the participants agreed with experts the practice was unethical. Teachers' perceptions of effort might be arbitrary, subjective, and therefore unfair. Moreover, assessment should primarily reflect students' mastery of learning goals; a grade should not directly relate to students' effort. On the other hand, students who viewed the practice as ethical thought non-academic factors were important criteria to assess students' performance. The final grade was not adequate to reflect the overall performance of students. Considering students' effort in grading encouraged more effort from students. The conflict between student needs and assessment needs led to the disagreement among students.

In terms of Scenario 9 ("At the beginning of the class, when a student requests to see her grade of a final exam, her professor shows the student the whole score sheet that includes all students' final scores"), 71% of the participants agreed with the experts that the practice was unethical; teachers should protect students' privacy. The practice also may harm students who are performing poorly; 29% of the students thought the practice was ethical as knowing other students' scores may help them compare their performance with that of other students. Conflicts between diverse student needs were involved in this case.

For Scenario 10 ("A professor who knows a student had a bad week because of problems at home bumps the student's participation grade up a few points to compensate for his bad score on a quiz"), 56% of the students agreed with experts that this practice was unethical. The major reason was that this practice would be unfair to other students. For example, one student responded, "*Dealing with bad experiences in life is an important life skill for students. Teachers can provide support and guidance to students in other ways to help students develop in a healthy way both physically and mentally.*" Other students thought the practice was ethical as teachers showed concern, encouragement, and care to students, which was important for students' psychological development. Teacher needs and assessment needs conflicted in this assessment situation.

With regard to Scenario 11 ("At the beginning of the semester, a professor shares with students the rubrics for each task. The professor leads students in a discussion about the rubrics, makes changes to the rubrics according to students' feedback, and gives students the final versions to guide their completion of the course tasks"), 90% of the participants viewed the practice to be ethical, which corresponded with the view of experts. These students justified their options in different ways. Involving students in the development made the rubric more student-oriented, objective, fair, and easy to be followed. For example, one student stated, "*It emphasized the active role students play in the assessment process. The communication with students on the feasibility of rubrics helps maintain the fairness and objectivity of rubrics and promotes the critical thinking ability of students.*" Second, student involvement promoted the communication between students and teachers, thus enhancing the student–teacher relationship. A small percentage of students considered the practice to be unethical. They believed students' perceptions of the rubric might be subjective due to the lack of professional knowledge. Besides, informing students of the rubric would encourage rubric-oriented learning. Assessment needs and student needs intersected in this assessment scenario.

Seventy-seven percent of the participants agreed with experts the practice in Scenario 13 ("While administering a class-level mid-term test, a professor notices that most students missed the same question. The professor reminds all students to check their answers to that question one more time") was unethical. Teachers should strictly follow the test administration rules, and the practice would be unfair to other students who were not reminded of what they missed. Third, reminding of the missed item might produce an inaccurate reflection of students' achievement. For example, one participant responded, "A *test is to evaluate students' mastery level of knowledge. Telling a student of the missed item may weaken the function of test. In addition, it is not fair to those students who already answered the item correctly, and even mislead this group of students.*" Twenty-three percent of the students viewed the practice as ethical if teachers showed concern by reminding all students in the room. Thus, the teacher needs and assessment needs conflicted in this assessment practice.

Regarding Scenario 14 ("An instructor uses only multiple-choice questions in the end-of-course exam. She justifies this practice by stating multiple choice questions can be graded objectively and efficiently"), 79% of the participants agreed with experts the practice was unethical. A multiple-choice assessment format mainly measured the low cognitive skills of students, and higher education should emphasize higher cognitive skills. Multiple assessment formats were necessary for assessing the overall achievement of students. Second, the probability of guessing the multiple-choice items right was higher than for other assessment methods, thus reducing the validity of this assessment. For example, one student responded "*Teachers should encourage higher cognitive skills of students, such as analyzing, creating and evaluating, by adopting other performance-based assessment formats. The multiple-choice assessment is easy to grade but prohibits students' learning motivation and restricts students' creative thinking. Besides, it is inadequate to assess students' overall performance.*" A small portion of students perceived the practice as ethical because multiple-choice assessment does have some advantages. It is easy to

administer, easy to grade; grading results tend to be objective. In this case, different assessment needs conflicted.

Scenario 15 ("A college professor counts students' attendance as 20% of their final grades") was treated as ethical by most students. Attendance enhances students' learning motivation, self-regulation ability, and the positive behavior habit of being punctual. Second, as attendance reflects students' attitude toward learning, it should be an important assessment criterion. In addition, attendance displays respect for classmates, teachers, and knowledge. Only 12% of the students agreed with experts the practice was unethical. This minority thought attendance was irrelevant to students' mastery of knowledge. Besides, a weight of 20% was excessive and may obligate students to be present, which may hinder their active learning. The intersection between assessment needs and students needs caused disagreement among students.

6 Discussion and conclusions

6.1 Factors associated with students' perceptions of the ethical issues

The current study investigated the underlying factors associated with Chinese college students' perceptions of ethicality of classroom assessment practices. Results from quantitative analysis indicated students' gender, grade level, major, and program were significantly associated with their perceptions of the ethicality of multiple assessment situations (10 of the 15 assessment scenarios in the survey). In general, female students, higher grade level students (juniors and seniors), those majoring in STEM, and students who were not in a teacher preparation program have significantly higher agreement with experts with regard to the ethicality of most assessment situations (7 of the 10 assessment scenarios). These results lead to a conclusion that students' perceptions of the ethical issues in classroom assessment are directly related to their differences in gender, grade level, major, and program type. In particular, gender is the key element affecting students' perception of the ethicality of assessment practices. The results supported the findings of previous studies concerning the differences in students' perceptions of assessment as a function of gender (Dhindsa et al., 2007; Gao, 2012), subject area (Iannone & Simpson, 2013), and grade level (Dhindsa et al., 2007).

The present study also supported the value of incorporating students' perceptions of the ethical issues to improve the validity of classroom assessment as assessment should take into account the systematic differences among groups (Green & Johnson, 2010). The present findings implied that university teachers, administrators, and policymakers should pay attention to the nature of class gender, grade level, major, and program type regarding the design of classroom assessment tasks and professional training programs for teachers. Besides, professional development programs should focus on increasing teachers' awareness of the ethicality of their classroom assessment practices.

Qualitative data analysis evaluated in-depth exploration of students' justification of their perceptions (why they think certain scenarios are ethical or unethical).

Students justified their decisions on the ethicality of the scenarios from the perspectives of assessment needs, student needs, and teacher needs. Student needs indicates that teachers should consider students' personal needs such as effort, family background, academic background, and physical conditions in assessment. Teacher needs involves teachers' personal needs or desires as they play the role of being teachers. Assessment needs requires that assessment should accurately reflect students' mastery level of knowledge and skill (Gao et al., 2019; Green & Johnson, 2010; Pope et al., 2009). When these needs intersected, students disagreed regarding the ethicality of the assessment practices. The findings from both qualitative and quantitative study showed demographic characteristics of students as well as the conflicting needs of stakeholders in assessment were associated with students' perceptions of the ethicality of the assessment practices.

Addressing only students' strengths in feedback as described in Scenario 1 was viewed as unethical by experts as students need to strengthen areas in which they are weak (JCSEE, 2015). The current study indicated that juniors and seniors viewed this assessment practice as unethical, while freshmen and sophomores considered it as ethical. Students viewed the practice from different needs of students. Positive feedback enhanced students' self-confidence and learning motivation. Negative feedback also helps students to be able to reflect and to deal with frustration. Different student needs intersected in this assessment practice, causing disagreement among students. The implications of this are particularly interesting to faculty members who teach the higher grades of universities. As recommended by Brookhart and Nitko (2018) that effective feedback should inform students not only what they are doing well but also what they need to improve, teachers, especially those who teach higher grade level students, should provide more objective feedback that students can use to improve their learning and focus more on students' personal needs in assessment.

Male students, those majoring in STEM, and those who were not in a teacher preparation program tended to think giving students a higher grade based on their mastery level even if students did not submit some of the assignments as described in Scenario 4 was ethical. The conflict between assessment needs and student needs made students come up with different decisions. From the perspective of assessment needs, students' scores should reflect their levels of achievement (O'Connor, 2017). Homework and work done for practice should not be counted as part of the grade as they only display students' developing skills and knowledge rather than expertise (Taylor & Nolen, 2005). From the perspective of student needs, assignment submission is important to address as part of developing students' sense of responsibility and positive attitude toward learning as well. The findings implied that teachers teaching a female-dominated class, those majored in non-STEM, and those in teacher preparation programs should focus more on students' personal needs. In practice, teachers need to balance the assessment needs and student needs as recommended by Green and Johnson (2010) that teachers learn why students have not submitted assignments and assist or supervise make-up work.

Teachers should inform students of the grading plan and the content in the tests (Brookhart & Nitko, 2018). In this sense, including surprise items in test as described in Scenario 5 is not recommended by experts, and students majoring

in STEM agreed with experts. Including surprise items in a test may increase test anxiety of students and harm their self-esteem. However, students of other majors thought including surprise items would stimulate their learning motivation. If a student only focuses on the contents covered by study/test guide, he/she will ignore those not covered in the guide. Students viewed these assessment practices from different student needs. Teachers, especially those teaching STEM, may need to raise their awareness by giving students enough information before assessment and take into account diverse needs of students.

The effort is the most common non-academic factor teachers consider when assigning grades (Cox, 2011; Marzano, 2000). However, as assessment should primarily reflect students' mastery level of knowledge and skills (Green and Johnson, 2010; Guskey & Bailey, 2001), considering effort in grading will artificially inflate students' grades (Stiggins et al., 2004). Moreover, teachers' perceptions of effort might be arbitrary and unfair (Green & Johnson, 2010). Most female students, higher grade level students, and STEM students agreed with experts that considering students' effort in grading as described in Scenario 8 was unethical. However, some students held the opinion that considering students' effort stimulates students to increase effort. The intersection between student needs and assessment needs led to disagreement among students. Teachers teaching male-dominated classes, lower grade level students, and non-STEM students may consider using alternative assessment formats such as formative assessment for increasing student effort than assigning a grade conveying information about students' behavior.

Brookhart and Nitko (2018) stated posting all students' assessment results would do more harm than good, so evaluation should be kept confidential. Female students, students from higher grade levels, and those majoring in STEM stated that showing all students' scores to other students as described in Scenario 9 was unethical as their privacy should be protected. A few respondents stated that knowing other students' scores might help a student know his/her rank in class, and thereby stimulate his/her learning motivation. The disagreement among student arose from the conflict between diverse student needs. Brookhart and Nitko (2018) recommended student score records should not be transferred to any third party without authorization from either students or guardians. Male students, lower grade level students, and those majored in non-STEM should increase their awareness of confidentiality in classroom assessment.

Grades should represent students' mastery level of the learning goals (Green & Johnson, 2010), but grade alteration is a common assessment practice (Tierney, 2015). The current study showed that female students were more likely to consider the practice of altering students' grades due to family problems as described in Scenario 10 unethical. From the perspective of teacher needs, they want to show their concern and support for students. In contrast, from the perspective of assessment, grades should accurately reflect students' mastery of knowledge. In this situation, the intersection between assessment needs and student needs caused disagreement. Teachers, especially those teaching male-dominated classes, should show more concern for students' personal needs by giving them another chance to retake a test to show their true mastery.

The purpose of involving students in the assessment process is to help them use this information to monitor their learning (Chappuis & Stiggins, 2002). Most female students perceived the practice as described in Scenario 11 to be ethical as involving students in designing rubrics helps foster the active role of students in the learning process and perceive grading is fair. Teachers teaching classes with predominantly male students might need to convince these classes that having them help in developing rubrics is beneficial to students' learning. For example, teachers can discuss with students the areas of content on which exams will focus, the relative weight of different areas, and the number of questions or percentage of points. This may help reduce test anxiety and give students general ideas about the relative importance of different topics (Green & Johnson, 2010). On the other hand, a small percentage of students stated this practice as unethical, stating student involvement may increase subjectivity in the rubric as students lack professional training in this skill. To avoid possible bias, teachers should play the role of leading and guiding the whole process of rubric development.

While administering assessments, the teachers' responsibility is to ensure the administration process is fair to every student and will produce interpretable results (Brookhart & Nitko, 2018). Giving hints to students in any way that certain answers might be wrong or inappropriate (Popham, 1991). Most STEM students agreed with experts this practice as described in Scenario 13 was unethical. They stated teachers should follow testing rules strictly. STEM students also discussed that reminding examinees of the missed items in one classroom would be unfair to those in other classrooms. A few students viewed the practice as ethical if teachers only reminded the student of the missing items without telling the correct answers. Students perceived the ethicality of this assessment situation from the perspective of assessment needs and teacher needs, which conflicted. With regard to assessment needs, any factors reducing the validity of the assessment should be avoided. From the perspective of teacher needs, proctors might believe they need to remind students of these items so they will not miss points. The findings indicated teachers should consider their major responsibility of giving students sufficient information about the assessment procedures rather than the assessment content.

Teachers need sufficient information from multiple assessments to accurately evaluate each student's achievement (Smith, 2003); McMillan (2000) suggested that assessment should be versatile. Use of different assessment methods leads students to different learning approaches (Struyven et al., 2005). Students also vary in their preference for different assessment formats (Xu et al., 2016). Using multiple assessment formats provides a comprehensive picture of students' mastery level (Brookhart & Nitko, 2018; Waugh & Gronlund, 2013). Most female students felt using only multiple-choice questions in the final exam as described in Scenario 14 was unethical. The findings corresponded with the previous study results that female students viewed constructed-response assessment formats as being more effective (Aldrich et al., 2018). Students believed the practice was unethical because multiple-choice items mainly assess the lower cognitive skills of students while the focus of higher education should be on higher cognitive skills such as creating and evaluating. Students who favored multiple-choice assessment stated multiple-choice

assessment was easy to grade and administer. The contradiction between assessment needs led to the disagreement among students. Teachers, especially those who taught classes with mostly female students, should use multiple assessment formats to increase the validity of assessments and help students improve their weaker skills and take into consideration different assessment needs.

Scoring should not be affected by factors unrelated to mastery of the learning goals such as student effort, growth, behavior, and attendance (Oosterhof, 2009). The current study showed most female students agree with experts that counting students' attendance as 20% of their final grade as described in Scenario 15 was unethical. These students stated attendance was irrelevant to students' learning goals and assigning a 20% value to the final grade imposes pressure on students' active learning. In contrast to females, most male students viewed it as ethical. Attendance, as one of the important rules to follow, represents students' attitude toward learning and respect for knowledge. It promotes students' learning motivation and self-regulation ability. Student disagreement arose from the conflict between assessment needs and student needs. Teachers, especially those who teach male-dominated classes, can try creating a separate report on students' non-academic achievement so scores can primarily reflect mastery of learning goals as recommended by Winger (2005). In this way, both assessment needs and student needs were considered.

6.2 Implications for assessment practice

The current study contributes to finding out the factors associated with undergraduates' perception of the ethical issues in classroom assessment. The results of this study suggest several implications. First, as assessment is practiced in a specific context and situation, students' agreement regarding the ethicality of each individual scenario was hard to achieve. In the process of assessment practices, teachers should consider students' demographic characteristics, including gender, major, grade level, and program so assessment can meet individual student learning needs. Second, as stakeholders in assessment play different roles in the process, they have different needs including assessment needs, teacher needs, and student needs. These needs might conflict in specific assessment contexts. Teachers need to balance the diverse needs of different stakeholders in assessment in order to assess ethically. Third, stakeholders should be informed of the information related to ethical assessment standards, policies, and guidelines so that they can make appropriate decisions about student learning using assessment results. Fourth, discussion on ethical issues in assessment should be included in professional development programs to increase teachers' awareness of ethical assessment.

Overall, this study explored differences in students' perception of ethical issues as a function of a student's characteristics and the intersection of different stakeholders' needs involved in classroom assessment. The findings offered insight to both pre-service and in-service teachers. Teacher education programs should address the ethical issues in pre-service teachers' coursework, especially,

how to meet students' needs based on their diverse characteristics as well as the diverse needs of different stakeholders. As teachers lack formal assessment training (Stiggins, 1999; Tienken and Wilson, 2000), more professional development programs on assessment for in-service teachers are needed. In the training session, teachers can discuss the scenarios mentioned in this study to see how they perceive the ethicality of those assessment practices. Suggestions can be offered to educators who teach students of a specific gender, grade level, or major group. Guidelines and standards related to classroom assessment should be mentioned and discussed to increase stakeholders' awareness of increasing ethicality in assessment practices. Teachers can incorporate the discussion of ethical issues in instruction as well to increase students' awareness of ethical and fair assessment. These in-service training will help enhance the theoretical and practical knowledge of assessment stakeholders on ethical issues in classroom assessment.

6.3 Research limitations

The current study has several limitations. First, the number of scenarios in the survey measuring Chinese college students' perceptions of ethical issues in classroom assessment was limited. As assessment is specific to a context, additional scenarios should be added to the scale so the scenarios can represent various assessment practices in Chinese higher education. Future research can focus on improving and then validating the scale. Second, the qualitative data were collected from open-ended questions. Lack of direct communication with participants may have produced subjectivity and inaccuracy in coding. Future research can interview students to obtain in-depth information about their perceptions through face-to-face communication. Third, the current study evaluated only the underlying factors associated with students' perceptions of the ethicality of assessment practice. Applicable and contextualized ethical guidelines should be developed to provide assessment stakeholder with guidelines for solving the ethical issues in assessment. Fourth, even if we did find some common factors associated with students' perceptions of the ethical issues in classroom assessment, assessment practices may vary across majors or program types. Future research may limit the target population to students from the same major or program to find out specific perceptions of students from particular major or program.

Despite the described limitations, this study aims to fill a gap in the scarcity of research exploring differences in students' perceptions of ethical issues in classroom assessment as a function of gender, grade level, major, program type, and conflicts between diverse stakeholder needs. Further research is necessary to expand these results beyond the sample used in this study and provide additional insights about students' expectations and preferences for assessment practices. In general, ethical assessment practices enable teachers to use assessment information appropriately to improve student learning.

Table 3 Survey scenarios

1. To enhance self-esteem, a professor addresses only students' strengths when giving feedback to her students' assignments since she believes that positive feedback is good for students' growth

2. A professor does not grade all class-level quizzes. Instead, he lets students grade each other's paper and then share the results in groups

3. A professor uses observational checklists, anecdotal notes, and interviews in assessing students

4. As a professor finalizes grades, she notices the grade of a student is in between B + and an A − . She gave the student an A − because tests and papers showed the student had mastered the course objectives even though he had not completed some of his homework assignments

5. For the class-level final exam, a professor uses a few surprise items about additional topics that were covered in class but were not listed in the study guide

6. While administering a final exam, a professor notices that a student has skipped a problem and is recording all of her answers out of sequence on the answer sheet. The professor shows the student where to record the answer she is working on and instructs the student to put the answer to each question with the same number on the answer sheet

7. When assigning a team project to work on collaboratively, a professor does not provide rubrics on how it will be graded, stating instead that he will assign a score based on students' overall performance on the project

8. In grading a final exam, a professor always reads the student's name and considers effort in assigning grades

9. At the beginning of the class, when a student requests to see her grade of a final exam, her professor shows the student the whole score sheet that includes all students' final scores

10. A professor who knows a student had a bad week because of problems at home bumps the student's participation grade up a few points to compensate for his bad score on a quiz

11. At the beginning of the semester, a professor shares with students the rubrics for each task. The professor leads students in a discussion about the rubrics, makes changes to the rubrics according to students' feedback, and gives students the final versions to guide their completion of the course tasks

12. A professor is concerned that most students did not perform well on the class-level mid-term test. Based on the results, it has become mathematically impossible for about 70% of students to earn a passing grade. Thus, the professor adds 20 points to each student's mid-term score to make sure most students still have a chance to pass at the end of the semester

13. While administering a class-level mid-term test, a professor notices that most students missed the same question. The professor reminds all students to check their answers to that question one more time

14. An instructor uses only multiple-choice questions in the end-of-course exam. She justifies this practice by stating multiple choice questions can be graded objectively and efficiently

15. A college professor counts students' attendance as 20% of their final grades

Appendix

References

Airasian, P. W. (2000). *Assessment in the classroom: A concise approach*. McGraw-Hill.

Aldrich, R. S., Trammell, B. A., Poli, S., Potter, S., & Garringer, K. (2018). How age, gender, and class format relate to undergraduate students' perceptions of effective course assessments. *InSight: A Journal of Scholarly Teaching*, *13*, 118–129.

Alkharusi, H. A., & Al-Hosni, S. (2015). Perceptions of classroom assessment tasks: An interplay of gender, subject area, and grade level. *Cypriot Journal of Educational Sciences*, 205–217.

Alsarimi, A. M. (2000). *Classroom assessment and grading practices in the Sultanate of Oman* (Doctoral dissertation, University of Pittsburgh).

Beets, P. A. D. (2012). Strengthening morality and ethics in educational assessment through Ubuntu in South Africa. *Educational Philosophy and Theory, 44*, 68–83.

Bergman, D. J. (2013). Pre-service teachers' perceptions about ethical practices in student evaluation. *Issues in Teacher Education, 22*(1), 29–48.

Bergman, D. J. (2018). Pre-service teachers' pliable perceptions of ethical practices in student evaluation. *Issues in Teacher Education, 27*(3), 5–23.

Bergman, D. (2020). Pre-service teachers' perceptions of ethical student evaluation practices: Comparing impact of online and face-to-face coursework. *Journal of Technology and Teacher Education, 28*(4), 665–689.

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.

Biberman-Shalev, L., Sabbagh, C., Resh, N., & Kramarski, B. (2011). Grading styles and disciplinary expertise: The mediating role of the teacher's perception of the subject matter. *Teaching and Teacher Education, 27*(5), 831–840.

Biggs, J. (2003). Aligning teaching and assessing to course objectives. *Teaching and learning in higher education: New trends and innovations, 2*(4), 13-17.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university*. McGraw-Hill Education.

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., & van Haeringen, K. (2019). Contract cheating: A survey of Australian university students. *Studies in Higher Education, 44*(11), 1837–1856.

Brew, C., Riley, P., & Walta, C. (2009). Education students and their teachers: Comparing views on participative assessment practices. *Assessment & Evaluation in Higher Education, 34*(6), 641–657.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*(4), 399–413.

Brookhart, S. (2004). Grading. Pearson Education.

Brookhart, S. M., & Nitko, A. J. (2018). *Educational assessment of students*. Pearson Higher Ed.

Brown, G. T., & Hirschfeld, G. H. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice, 15*(1), 3–17.

Brown, G. T., & Wang, Z. (2013). Illustrating assessment: How Hong Kong university students conceive of the purposes of assessment. *Studies in Higher Education, 38*(7), 1037–1057.

Chappuis, S., & Stiggins, R. J. (2002). Classroom Assessment for Learning. *Educational Leadership, 60*(1), 40–44.

Cheng, L., Wu, Y., & Liu, X. (2015). Chinese university students' perceptions of assessment tasks and classroom assessment environment. *Language Testing in Asia, 5*(1), 1–17.

Cirlan, E. (2017). Ethical classroom assessment: differences for pre-service and in-service teachers in Finland (Master's thesis, University of Oulu Faculty of Education).

Cox, K. B. (2011). Putting classroom grading on the table: A reform in progress. *American Secondary Education*, 67–87.

Craddock, D., & Mathias, H. (2009). Assessment options in higher education. *Assessment & Evaluation in Higher Education, 34*(2), 127–140.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.

Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice, 21*(2), 149–166.

Dhindsa, H. S., Omar, K., & Waldrip, B. (2007). Upper secondary Bruneian science students' perceptions of assessment. *International Journal of Science Education, 29*(10), 1261–1280.

Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research, 53*(1).

Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.

Falchikov, N. (2013). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. Routledge.

Fan, X., Johnson, R ., & Liu, X (2017). Chinese university professors' perceptions about ethical issues in classroom assessment practices. *New Waves Educational Research & Development, 20*(2),1-18.

Fan, X., Johnson, R., Liu, J., Zhang, X., Liu, X., & Zhang, T. (2019). A comparative study of pre-service teachers' views on ethical issues in classroom assessment in China and the United States. *Frontiers of Education in China, 14*(2), 309-332.

Fan, X., Liu, X., & Johnson, R. L. (2020). A mixed method study of ethical issues in classroom assessment in Chinese higher education. *Asia Pacific Education Review, 21*, 183-195.

Fan, X., Johnson, R., Liu, X., & Gao, R. (2020). College students' views of ethical issues in classroom assessment in Chinese higher education. Studies in Higher Education.

Ferrer-Esteban, G. (2013). Rationale and incentives for cheating in the standardised tests of the Italian assessment system. *Programma Education FGA, Working paper*, (50).

Gao, M. (2012). Classroom assessments in mathematics: High school students' perceptions. *International Journal of Business and Social Science*, *3*(2).

Gao, R., Liu, J., Johnson, R., Wang, J., & Hu, L. (2019). Validating an ethical decision-making model of assessment using authentic scenarios. *Studies in Educational Evaluation, 62*, 187-196.

Gao, R., Liu, J., & Burgess, Y. (2021). An expanded ethical decision-making model in resolving ethical dilemmas in Assessment. *Studies in Educational Evaluation, 68*, 100978.

Gibbs, G. (2006). How assessment frames student learning. In C. Bryan & K. K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 23–36). Routledge.

Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects. In *Optimising new modes of assessment: In search of qualities and standards* (pp. 37–54). Springer, Dordrecht.

Göktürk Saglam, A. L., & Yalçin Duman, A. (2020). Exploring student perceptions of source-based writing assessment in a Turkish EAP context. *Novitas-ROYAL (research on Youth and Language), 14*(1), 25–38.

Green, S., Johnson, R., Kim, D., & Pope, N. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education, 23*(7), 999–1011.

Green, S., & Johnson, R. (2010). *Assessment is essential*. McGraw-Hill Higher Education.

Gronlund, N. (2003). *Assessment of student achievement*. Allyn and Bacon.

Guskey, T. R., & Jung, L. A. (2009). Grading and reporting in a standards-based environment: Implications for students with special needs. *Theory into Practice, 48*(1), 53–62.

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.

Harlen, W. (2007). Formative classroom assessment in science and education. In J. McMillan (Ed.), *Formative classroom assessment* (pp. 116–135). University Teachers College Press.

He, Y., Levin, B. B., & Li, Y. (2011). Comparing the content and sources of the pedagogical beliefs of Chinese and American pre-service teachers. *Journal of Education for Teaching, 37*(2), 155–171.

Hernández, R. (2012). Does continuous assessment in higher education support student learning? *Higher Education, 64*(4), 489–502.

Heritage, M. (2013). Gathering evidence of student understanding. In J. McMillan (Ed.), *SAGE handbook of research on classroom assessment.* Sage Publications.

Hosner, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. John Wiley & Son.

Johnson, R., Green, S., Kim, D., & Pope, N. (2008). Educational leaders' perceptions about assessment practices. *The American Journal of Evaluation, 29*(4), 520–530.

Johnson, R., Liu, J., & Burgess, Y. (2017). A model for making decisions about ethical dilemmas in student assessment. *Journal of Moral Education, 46*(2), 212–229.

Joint Committee on Standards for Educational and Evaluation. (2003). *The student evaluation standards: How to improve evaluations of students*. Educational Policy Leadership Institute/Corwin Press Inc.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education (revised)*. Science Directorate, American Psychological Association.

Joint Committee on Standards for Educational Evaluation. (2015). *The classroom assessment standards for PreK-12 teachers*. Kindle Direct Press.

Kennedy, K. J., Chan, J. K. S., Fok, P. K., & Yu, W. M. (2008). Forms of assessment and their potential for enhancing learning: Conceptual and cultural issues. *Educational Research for Policy and Practice, 7*(3), 197.

King, J. E. (2003). Running a best-subsets logistic regression: An alternative to stepwise methods. *Educational and Psychological Measurement, 63*(3), 392–403.

Kline, P. (2013). *Handbook of psychological testing*. Routledge.

Iannone, P., & Simpson, A. (2013). Students' perceptions of assessment in undergraduate mathematics. *Research in Mathematics Education, 15*(1), 17–33.

Lekholm, A. K. (2011). Effects of school characteristics on grades in compulsory school. *Scandinavian Journal of Educational Research, 55*(6), 587–608.

Liu, P., & Qi, C. (2005). Reform in the curriculum of basic education in the People's Republic of China: Pedagogy, application, and learners. *International Journal of Educational Reform, 14*(1), 35–44.

Liu, H., & Wu, Q. (2006). Consequences of college entrance exams in China and the reform challenges. *KEDI Journal of Educational Policy, 3*(1), 7–21.

Liu, J, Johnson, R., & Fan, X. (2016). A comparative study of Chinese and United States pre-service teachers' perceptions about ethical issues in classroom assessment. *Studies in Educational Evaluation, 48*, 56-66.

Lizzio, A., & Wilson, K. (2013). First-year students' appraisal of assessment tasks: Implications for efficacy, engagement and performance. *Assessment & Evaluation in Higher Education, 38*(4), 389–406.

Loyens, S. M., Rikers, R. M., & Schmidt, H. G. (2008). Relationships between students' conceptions of constructivist learning and their regulation and processing strategies. *Instructional Science, 36*(5–6), 445–462.

Ludlum, M., Conklin, M., & Tiger, A. (2021). Group projects in higher education: How demographic factors affect student perceptions of grading, leadership roles, assessment, and applicability. *Journal of Higher Education Theory & Practice*, *21*(1)

Lyon, E. G. (2013). Learning to assess science in linguistically diverse classrooms: Tracking growth in secondary science preservice teachers' assessment expertise. *Science Education, 97*(3), 442–467.

Marzano, R. J. (2000). *Transforming classroom grading*. Association for Supervision and Curriculum Development.

McGlory, D. W. (2013). *Assessing educators' perceptions of ethicalness of various assessment practices* (Doctoral dissertation, Wichita State University)

McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research, and Evaluation, 7*(1), 8.

McMillan, J. H. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction* (5th ed.). Pearson Education Inc.

National Council on Measurement in Education (NCME) (1995). *Code of professional responsibilities in educational measurement (CPR).* Available from http://ncme.org/resource-center/code-of-professional-responsibilities-im-educational-measurement/. Accessed August 2020

Niu, W. (2007). Western influences on Chinese educational testing. *Comparative Education, 43*(1), 71–91.

Oakland, T. (2004). Selected ethical issues relevant to test adaptations. In *Adapting educational and psychological tests for cross-cultural assessment* (pp. 77–104). Psychology Press.

O'connor, K. (2017). *How to grade for learning: Linking grades to standards*. Corwin Press.

Organisation for Economic Co-operation and Development. (2013). *Strong performers and successful reformers in education: Lessons from PISA 2012 for the United States*. OECD.

Oosterhof, A. (2009). *Developing and using classroom assessments* (4th ed.). Pearson.

O'Leary, M. (2008). Towards an agenda for professional development in assessment. *Journal in-Service Education, 34*(1), 109–114.

Özbek, O. (2013). Physical education teachers' types of analyzing professional ethical dilemmas. *Life Science Journal, 10*(1), 2670–2678.

Pan, Y. C. (2020). Taiwan university students' perceptions of summative and formative classroom assessment in English courses. *TESOL International Journal, 15*(2), 46–64.

Peterson, S. S., & Kennedy, K. (2006). Sixth-grade teachers' written comments on student writing: Genre and gender influences. *Written Communication*, *23*(1), 36–62.

Pope, N. S. (2006). *Do no harm to whom?* (pp. 25–31). South Atlantic Philosophy of Education Society Yearbook.

Pope, N., Green, S. K., Johnson, R. L., & Mitchell, M. (2009). Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education, 25*(5), 778–782.

Popham, J. W. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice, 1*(1), 12–15.

Popham, W. J. (2017). *Classroom assessment: What teachers need to know* (8th ed.). Pearson.

Rasooli, A., Zandi, H., & DeLuca, C. (2019). Conceptualising fairness in classroom assessment: Exploring the value of organisational justice theory. *Assessment in Education: Principles, Policy & Practice, 26*(5), 584–611.

Sambell, K., McDowell, L., & Montgomery, C. (2012). *Assessment for learning in higher education*. Routledge.

Skelton, C., Francis, B., & Smulyan, L. (Eds.). (2006). *The SAGE handbook of gender and education*. Sage.

Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice, 22*(4), 26–33.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice, 18*(1), 23–27.

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). Classroom assessment for student learning: Doing it right, using it well. Assessment Training Institute.

Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice, 44*(1), 11–18.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education, 30*(4), 325–341.

Taylor, K., & Nolen, S. (2005). *Classroom assessment: Supporting teaching and learning in real classrooms*. Pearson Education.

Textor, C. (2020, May 26). *Number of college and university graduates in China 2008–2018*. Statista. https://www.statista.com/statistics/227272/number-of-university-graduates-in-china/

Textor, C. (2020, Nov 13). *Number of college and in China 2019*. Statista. https://www.statista.com/statistics/1095134/number-of-colleges-and-universities-in-china-by-type/

Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). *Measurement and evaluation in psychology and education*. Macmillan Publishing Co Inc.

Tienken, C., & Wilson, M. (2000). Using state standards and tests to improve instruction. *Practical Assessment, Research, and Evaluation, 7*(1), 13.

Tierney, R. D. (2014). Fairness as a multifaceted quality in classroom assessment. *Studies in Educational Evaluation*, *43*, 55–69.

Tierney, R. D. (2015). Altered grades: A grey zone in the ethics of classroom assessment. *Assessment Matters*, *8*, 5–30.

Tierney, R. (2016). Fairness in educational assessment. In M. A. Peters (Ed.), *Encyclopedia of educational philosophy and theory* (pp. 1–6). Springer Singapore.

Wafubwa, R. N., & Ochieng, P. O. (2021). Students perception of teachers' use of formative assessment strategies in mathematics classrooms. *Elementary Education Online/ilkoegretim Online*, *20*(2), 123–132.

Waugh, C. K., & Gronlund, N. E. (2013). *Assessment of student achievement* (10th ed.). Pearson.

Weurlander, M., Söderberg, M., Scheja, M., Hult, H., & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: Students' experiences of different methods of formative assessment. *Assessment & Evaluation in Higher Education*, *37*(6), 747–760.

White, E. (2009). Are you assessment literate? Some fundamental questions regarding effective classroom-based assessment. *OnCUE Journal*, *3*(1), 3–25.

Winger, T. (2005). Grading to communicate. *Educational Leadership*, *63*(3).

Wurf, G., & Povey, R. (2020). "They show how smart you are": A mixed methods study of primary students' perceptions of assessment tasks. *Issues in Educational Research, 30*(3), 1162–1182.

Xu, X., Kauer, S., & Tupy, S. (2016). Multiple-choice questions: Tips for optimizing assessment in-seat and online. *Scholarship of Teaching and Learning in Psychology, 2*(2), 147.

Yazici, A., Yazici, S., & Erdem, M. S. (2011). Faculty and student perceptions on college cheating: Evidence from Turkey. *Educational Studies, 37*(2), 221–231.

Ye, L., Fei, Z., & Wang, T. (2007). *China: Five thousand years of history and civilization*. City University of HK Press.