

How Does Culture Influence Student's Assessments

Aihua Zhu*

Guangzhou College of Technology and Business, Guangzhou, Guangdong, China

*Corresponding Author.

Abstract:

Many studies focused on variables affecting assessment such as different assessors and proper use of scales, few research focused on cultural factors. This study aims to explore how Chinese culture affects assessment results. Taking a writing assignment as a case study, it collected 100 students' self-assessment and peer assessment scores, using a quantitative method to analyze their reliability. The result shows the inconsistency between different assessors. Face-to-face interviews are conducted, the findings indicate that some raters mark themselves higher to keep a high rank, for staying at the bottom is shameful. Peer assessments are mainly based on friendship, admiration, or maintaining harmonious interpersonal relationships. The results show that Chinese testing culture, face issues, relationship issues, the concept of harmonious coexistence, respect for labor, and worship of power have impacts on assessment behavior. Thus, assessments should take assessors' social bonds and culture into consideration.

Keywords: *Self-assessment, Peer assessment, Assessment results, Chinese culture.*

I. BACKGROUNDS

Educational assessment is an indispensable part of talent cultivation. With the in-depth study of student-centered education, great changes and innovations have taken place in educational assessment. The purpose of assessment shifted from "Assessment of Learning (Ao L)" to "Assessment for Learning" (A f L) and "Assessment as Learning" (Aa L), and assessment methods are shifting from teacher-centered assessment to students' assessment which aligns with student-centered education concept.

Mehrens (1992) defines students' assessment as a value judgment by using certain techniques and methods according to certain criteria, with students as the assessing object [1]. Different from grading, assessment is not only a part of the learning process but also strong evidence for judging the learning results, style as well as a learning strategy. So, its main function is reflected in diagnostic, summative, or formative assessment. As a systematic collection and analysis of information to improve student learning, it can identify, monitor, guide, or improve students' learning achievements or goals, and alleviate students' anxiety about assessment [2].

There are two methods of student assessment, one is peer assessment (Hereinafter referred to as PA) and the other is self-assessment (Hereinafter referred to as SA), both of which emphasize the enthusiasm of learners' participation in the assessment process. To promote learning, learners have a dual identity in the learning process, for they not only objectively analyze their learning but also identify and analyze their peers'. It can also be referred to as assessment literacy: a crucial ability to judge or evaluate students' own and their peers' performances on assessment tasks marked by criteria and standards, which affects students' learning in school and their ability to judge their performance in the work field in the long-term.

The theoretical framework of language assessment consists of testing knowledge, testing skills, and testing principles[3]. One of the important factors of the testing principle is that assessors should conduct tests or assessments based on local social and cultural values. Regarding the research on student assessment, scholars are constantly exploring the reliability and validity of student assessments, as well as their backwash effect on education. Studies of the interfering factors on their assessment behavior mainly focus on students' language ability and the validity of the assessment scale, and few researchers explored their cultural background. This study takes one writing task as an example to explore the characteristics of student assessment and analyzes the reasons for the characteristics from a cultural perspective.

II. LITERATURE REVIEW

2.1 Peer Assessment

Studies generally focus on the validity and reliability of peer assessment, feedback, and emotional perception as well as the effect of peer assessment.

Research had proved that the validity and reliability of peer assessment are generally high. Holistic scoring had adequate validity with 2nd-year students, sometimes it could be as reliable as teachers' assessment, sometimes it is more strict in language than teachers assessment [4],thus PA could be used as a trustworthy assessment method"[5]. However, the high reliability should be based upon anonymity to allow students to overcome inhibitions and improve their assessment skills[6]. Furthermore, PA is not as stable as teachers' assessment, for it showed developmental growth, it has higher validity for assessments of higher-level concerns across broad developmental levels, while lower-level students may require more support (Zhang, 2020).

Many studies also found that PA can promote language learning in different ways, and is even more helpful than self-assessment despite the stress and comfortlessness. Besides, it can play a role in assisting self-assessment [7]; Students receiving PA and teachers' assessments improved their writing in lexical complexity, accuracy, and some features of grammatical complexity and fluency than those who received only teacher assessment. In general, PA was conducive to improving writing quality [8].

Researchers also discussed feedback and perception of peer assessment. It was not only a good way to facilitate students' participation but also a reliable way to assess students' contributions to a group project

[9]. In addition to assessing function, PA is contributive to activating students learning motivation, attitudes. Studies proved that students involved in PA activities endorsed a stronger mindset and maintained a high level of motivation than those who did not engage [10]; Students who approbate PA believe that accepting and providing feedback are helpful to improve English skills, reduce stress and fear of opposition, and rebuild confidence. Providing feedback is more effective than receiving feedback. Generally, students perceive feedback as extremely useful, express more gratitude [11]. PA can significantly improve learners' self-efficacy.

2.2 Self-Assessment

The introduction of self-assessment was viewed as promising by many, especially in formative assessment contexts. On one hand, it could encourage students' learning awareness and confidence, on the other hand, it can give teachers feedback on the characteristics of students' learning styles, and areas needing remediation. There are two methods for students to conduct self-assessment, one is to provide a scale to rate their language proficiency in a specific field, and the other is to integrate SA into portfolio learning and assessment. Both methods have been proven effective in promoting students learning efficacy (Gu, 2019; Chen, 2019).

The current study of SA mainly focuses on the following two aspects: the function of promoting learning and the validity of self-assessment.

SA is proven to promote self-regulatory and metacognitive skills in learners, for it has a clear connection between teaching and learning and provides an excellent source of feedback [12]. In addition, it can also be effective in motivating students to think critically [13].

However, assessment results can be affected by some factors. Variables that affect the validity of students' assessment both (PA and SA) include assessors and SA scales. The simple comment from peers makes the assessment unsuggestive for modification, and peers' inexperience in assessing and efficiency in language knowledge can lead to unfair results. Sometimes 'disrespectful' behaviors and attitudes of peer reviewers can cause negative emotion in students being assessed. So, PA was resisted for summative purposes by some students. Besides, different language levels and motivations of learners also lead to completely different participation of assessment, making the results different[14]. Other variables such as gender, professional background, attitudinal factors such as interpersonal relations also contribute to the results of PA[15].

A study on the validity of the SA scale shows that inaccuracy and ambiguity in concepts of descriptors played an important role in the deviation of assessment ability. Moderators such as SA instruments, training, the total number of items in the SA instrument, and reliability of the SA instrument exert great effects on assessing results, external measures, languages skills did not influence SA. To achieve better results teachers should inform students about how to use SA in their classrooms to promote learner autonomy and independence.

Self-assessment is a reasonable measure, but the results of which cannot be trusted unconditionally, for psychological factors of self-assessors such as nervousness and anxiety influence assessment results[16]. In real assessing practice, even peer assessment may be inconsistent with the criteria for the variables listed above. Even the same student or student group in the same level might rate themselves and others differently with the same rubric. When focusing on the general pattern, some special cases should be considered. To take a holistic view of the underlying reasons, one should not only be confined to a linguistic or pedagogical domain, other attributing factors should be explored further. As is known that one's behavior is profoundly influenced by his culture. The same is true in students' learning and assessing behavior. However. The literature so far seldom considers cultural context. In this study, we aimed at exploring the cultural factors contributing to students' assessment behavior, instead of from a psychological, pedagogical, or linguistic perspective.

III. RESEARCH PROCESS

Research methods: a quantitative method is used to analyze students' rating reliability, and a qualitative method to explore the underlying cultural reasons affecting their rating results.

Research Questions:

- ①: what is the general pattern of students' assessment?
- ②: what cultural factors influenced their assessing behavior?

3.1 Quantitative Study

100 pieces of writing were collected from second-year English majors at a local college in China. The writing genre is a narration, finished within 30 minutes in a writing class according to the teacher's guidelines. Students did a self-assessment first, then their work was assessed face to face by the group members who are divided voluntarily, with 4 students in each. The final rating score is an average of the other three. Students were allowed to rebut or challenge the feedback or ratings from their peers.

3.1.1 Rating description

Students participating in PA and SA are already grouped voluntarily, according to their actual situation, such as ability level and close relationship at the beginning of the semester. Before this rating, they have collectively received rating training on writing, and have a certain understanding of scoring standards. In this study assessment of the teacher (with 13 years' experience of teaching and rating) was also included as a benchmark.

3.1.2 Rating scale

This writing rating criteria have been used for many years in this college, and its reliability and validity have been tested fine. As shown in TABLE I, the scoring scale is divided into four dimensions, namely task response, textual syntactic connection, vocabulary accuracy and richness, grammatical accuracy, and sentence pattern diversity. The writing ability is divided into five levels, 60 points Below, 60-70 points, 70-80 points, 80-90 points, 90-100 points, as shown in TABLE II. We set each gear to 1 point. And the rater will give an overall score, based on the description of the scale and also feedback. The data collected were analyzed by IBM SPSS Statistics 21.

TABLE I. Rating criterion

TASK RESPONSE	COHERENCE AND COHESION	LEXICAL RESOURCE	GRAMMATICAL RANGE AND ACCURACY
FULLY AND APPROPRIATELY EXPLAIN, PROVE OR DISCUSS THE THESIS WITH CLEAR, RELEVANT, AND WELL-SUPPORTED IDEAS AT THE REQUIRED LENGTH	INFORMATION AND IDEAS ARE WELL ORGANIZED, PRESENTED, AND LINKED	A WIDE RANGE OF VOCABULARY USED ACCURATELY AND APPROPRIATELY FOR THE TASK	A WIDE RANGE OF STRUCTURES USED ACCURATELY AND APPROPRIATELY FOR THE TASK

TABLE II. Rating scale

5. 90-100 POINTS: WRITE OUT ALL THE MAIN POINTS OF THE CONTENT, WITH CLEAR LEVELS, FLUENT LANGUAGE, SENTENCE CHANGES, COMPLEX STRUCTURES (AT LEAST 2-3), BASICALLY NO GRAMMATICAL ERRORS (1-2 ERRORS) OR A SMALL AMOUNT DUE TO COMPLEX USE GRAMMAR OR SPELLING ERRORS CAUSED BY STRUCTURE, BUT DO NOT AFFECT THE MEANING
4. 80-89 POINTS: BASICALLY, WRITE OUT ALL THE MAIN POINTS OF THE CONTENT, WITH CLEAR LEVELS, FLUENT LANGUAGE, RICH VOCABULARY AND SENTENCE PATTERNS, AND A FEW GRAMMATICAL ERRORS (3-4 ERRORS).
3. 70-79 POINTS: WRITE MOST OF THE MAIN POINTS, THE LANGUAGE IS FLUENT, THERE ARE SOME GRAMMATICAL AND SPELLING ERRORS, AND DO NOT AFFECT THE EXPRESSION OF MEANING.
2. 60-70PINTS.WRITE OUT MOST OF THE MAIN POINTS, THERE ARE MORE GRAMMAR AND SPELLING ERRORS, THERE ARE MORE THAN THREE WRONG SENTENCES BUT DOES NOT AFFECT THE GENERAL MEANING.

1. BELOW 60: WRITE ABOUT HALF OF THE MAIN POINTS OF THE CONTENT, THE LANGUAGE IS NOT SMOOTH, THE GRAMMATICAL STRUCTURE IS MONOTONOUS, THERE ARE MANY ERRORS, ONLY A FEW SENTENCES ARE READABLE, WHICH AFFECTS UNDERSTANDING

3.1.3 Findings

Frequency analysis showed that peer ratings range from 1 to 5, self-ratings vary from 2-5 and teacher-rated from 1-4. Self-raters did not rank their works as unqualified as their teacher and peers did. The teacher did not rate any work grade 5, while 2 peer ratings and 5 self-ratings are in grade 5. Peer rating comes mostly in grade 3, self-rating grade 4, and teacher's rating grade 3, as shown in TABLE III.

TABLE III. Frequency statistics

PEER RATING					
GRADE		FREQUENCY	PERCENT	VALID PERCENT	CUMULATIVE PERCENT
VALID	1	2	2.0	2.0	2.0
	2	7	7.0	7.0	9.0
	3	61	61.0	61.0	70.0
	4	28	28.0	28.0	98.0
	5	2	2.0	2.0	100.0
	TOTAL	100	100.0	100.0	
SELF- RATING					
VALID	2	6	6.0	6.0	6.0
	3	27	27.0	27.0	33.0
	4	62	62.0	62.0	95.0
	5	5	5.0	5.0	100.0
	TOTAL	100	100.0	100.0	
TEACHER'S RATING					
VALID	1	6	6.0	6.0	6.0
	2	34	34.0	34.0	40.0
	3	47	47.0	47.0	87.0
	4	13	13.0	13.0	100.0
	TOTAL	100	100.0	100.0	

The descriptive statistics in TABLE IV indicated the mean of the teacher's rating is the lowest with a Mean of 2.67, and self-rating the highest with a Mean of 3.66, peer rating stands at 3.21. The general standard deviation is <1, which is acceptable. Commonly, students rate themselves and their peers with higher scores than teachers. Hence, it can be concluded that teacher's assessment is more severe than students, self-assessment is the least strict one. Peer assessment stands in the middle. But this does not

necessarily mean that students' assessment is not trustworthy.

TABLE IV. Descriptive statistics

	N	MEAN	STD. DEVIATION
SELF-RATING	100	3.66	.670
PEER RATING	100	3.21	.686
TEACHERS' RATING	100	2.67	.779

As is shown in TABLE V, the reliability analysis showed consistency among teachers' assessments and PA, PA&SA with Cronbach's α respectively 0.719 and 0.747. The latter is quite close to that of these three items, i.e., 0.746. A generally accepted rule is that α of 0.6-0.7 indicates an acceptable level of reliability, and 0.8 or greater is a very good level. However, the α of self-assessment and teachers' rating is 0.508, which is unacceptable. It is suggested that peer assessment could be used as a credible result of assessment but not self-assessment.

TABLE V. Reliability

N of Items	Self-rating, peer & teacher's rating	Self-rating & Peer rating (2 Items)	Peer rating & Teacher's rating (2 Items)	Teacher's rating & self -rating (2 Items)
Cronbach's Alpha	.746	.747	.719	.508

3.2 Qualitative Study

Since there are gaps between students' and teachers' assessments, the following task is to find out the main courses of that. Face-to-face interviews were conducted: group interviews were initiated firstly, then 11 individuals were picked up to be interviewed for their rating score is different from the general pattern.

The first category: whose self-assessment and peer assessment scores were lower than the teacher's; The second genre: peer-assessment scores were 2 points higher than the teacher, while most peer assessments were 1 grade higher. In these situations, peer assessors would be interviewed. The third genre: 5 students who were rated as 1 by the teacher, which meant that their works were unqualified, although they rated themselves qualified, 4 of them even score themselves middle to high levels.

3.2.1 Findings of group interview

Students were asked questions about how they think of their self-assessment, peer assessment, and teacher's assessment.

Question1: Do you think it is justifiable to rate your peer's such grades?

Among the 25 groups, all of them thought that the peer rating was under careful consideration and fair. Although they rate higher than their teacher, they do not deem that result unreliable. On the contrary, they believed that the feedback was quite helpful. The major reason for a higher rating than the teacher was that they wanted to spare others' feelings, or tried to maintain a harmonious relationship with their peers, as students narrated below:

"We can see the writing problems raised by the teacher, but because of the face problem, we will not give others low scores to spare his feelings. Anyway, giving low marks will affect the harmony between us. We should encourage everyone by giving them affirmative scores."

"We keep some of the feedback because if you listed too many problems, the writers would be frustrated. Instead of helping them, we will demotivate them".

Question2: Do you think it justifiable to rate yourself in such a grade?

Most students responded that either they were ignorant of their mistakes or they cared much about the final score, for low marks and ranking was shameful. As some students told:

"I did not find the mistakes myself."

"I know that I did not meet the standard, but I do not want to rate myself lowly, for that is shameful while others' scores were much higher than yours, anyway, this assignment takes a certain percentage of our daily performances, and would ultimately affect our final assessment score".

"We have made more progress than last year when we did not know the rating criterion".

Question3: How do you think of the teacher's assessment

Although the teacher's assessment was helpful, a large sum of students thought their teacher was too harsh to their work, which had demotivated them to some degree. The most often talked about descriptors have been:

"Too severe; fare, but frustrating; too holistic but lack of details; demotivating".

3.2.2 Findings of some special individuals and groups.

For students1-3, whose self-assessment and peer assessment was lower than the teacher's rating, their response to self-assessment was quite simple: They are not confident.

“I think there is a long way to go on writing. I still need improvement. My English is poor, especially writing”.

But their peers listed different reasons, among which their attitude toward study and team cooperation is a key element in rating. As some students narrated:

“He is lazy and uncooperative; every time whenever we are assigned tasks, he contributed least to that.”

“He is a slacker, and failed in many tests last year. In this writing we cannot understand what he is narrating”.

Peer assessment scores for students 4, 5,6, two of them were monitors (No.4&5), male; Although she did not take any leadership in class, the other female student is very hard-working, winning a full scholarship last year. Interviews were initiated with their group members, and the result was quite surprising. Their peers appreciate their contribution to a team and their effort much more than their learning outcomes.

Comments on the rating of monitors:

“Our monitors did many things for us, not only did he work hard, but also cooperative and was willing to help everyone”. “Our monitor is smart and handsome; he is diligent and helpful”.

Comments on the scholarship winner:

“She is hardworking and she revised her essay before it was sent to us, she is a Straight-A student, she cannot be wrong, even if we find out the error, we don’t know if we or she is wrong.”

Students (NO.7-11) rated themselves much higher than the teacher who rated them grade 1 shared equal reasons as other students: ignorant of their mistakes or *face* problems.

IV. DISCUSSIONS

4.1 Major Findings

The statistics and interviews above indicated that:

① students tend to rate themselves and peer higher to keep face or maintain a harmonious relationship with their classmates;

② students tend to add more scores to those who are cooperative and contribute to group tasks;

③students are inclined to rate slackers and uncooperative students lower.

These assessment characteristics are largely influenced by the cultural environment in which they live. One of them is the long-standing testing culture in China, in which students are forced to rank at the top for college admission. The norm-referenced testing culture still exerts great influence on college students.

The second one is the profound impact of Confucian culture, which stresses the top priority of harmony. It is a state of congenial co-existence based on due respect for differences and diversity, also an important moral concept of the Confucian school in managing interpersonal relations. So, students mark their peers in a way that would not offend them.

The third one, facing issues. When taking others' feelings into account, Chinese people also try to maintain their faces, for losing face is a shameful thing. Therefore, in terms of self-assessment, no classmate is willing to give himself a low score because he is afraid of failing the exam and losing face.

The Fourth one is the Chinese philosophy of collectivism. Unlike western individualism, Chinese people pay attention to the power of the collective. Anyone who impairs or undermines the interests of the collective is despised, and whoever does not contribute his or her efforts for the well-being of a team is considered selfish and probably will be isolated and ostracized by other members. So, in peer assessment, those uncooperative students were given low marks, while cooperative and suggestive ones won more grades.

Last but not least, the worship of authority and appreciation for the spirit of hardworking. Chinese enjoys a long history of 5000 years in the process of which emperor dominates the society, people are deeply influenced by this authority dominant culture, even in modern society, Awe of authority inhibits people from challenging them. Furthermore, the hardworking spirit of the authority makes her work Unquestionable.

4.2 Limitations

However, there are some limitations to this study. Owing to the small sum of samples, there are only 11 exceptional cases to study. Besides cultural factors what are other weighing elements? For example, are there any gender differences in students' assessments? For this study is conducted in foreign language departments where 84% of the participants are female, the male makes up only 16%. The insufficient number of male students keeps this research from exploring gender differences, which assumes a critical part in Chinese culture.

From the analysis above, it is clear that assessments should be multi-dimensional, taking students' motivation, progress, social bond, and cultures into consideration. To help Chinese instructors achieve fairer students' assessments, more work should be done to achieve higher reliability, quantitative assessment

and qualitative assessment should both be applied in formative assessment. But to what extent, should teachers carry out to complement these diversified assessments to improve students learning, it is the journey ahead to be explored.

ACKNOWLEDGEMENTS

Funded by the school project of Guangzhou College of technology and business: Development and Validation of Business English Proficiency Scale. Fund NO.: ZL20211145

REFERENCES

- [1] W. A. Mehrens, "Using performance assessment for accountability purposes." *Educational Measurement: Issues and Practice*, 1992.
- [2] J. Tai, R. Ajjawi, D. Boud, P. Dawson, and E. Panadero, "Developing evaluative judgement: enabling students to make decisions about the quality of work," *High Educ*, vol. 76, no. 3, pp. 467–481, Sep. 2018, doi: 10.1007/s10734-017-0220-3.
- [3] L. F. Bachman, "What does language testing have to offer?" *TESOL quarterly*, vol. 25, no. 4, pp. 671–704, 1991.
- [4] A. Rezaei and K. Barkaoui, "Peer and teacher assessment of second-language writing in high-and low-stakes conditions," *ITL-International Journal of Applied Linguistics*, vol. 172, no. 2, pp. 199–228, 2021, doi: 10.1075/itl.20006.rez.
- [5] J. H. Yu *et al.*, "Assessment of medical students' clinical performance using high-fidelity simulation: comparison of peer and instructor assessment," *BMC Medical Education*, vol. 21, no. 1, pp. 1–6, 2021, doi: 10.1186/s12909-021-02952-w.
- [6] T. Seifert and O. Feliks, "Online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills," *Assessment & Evaluation in Higher Education*, vol. 44, no. 2, pp. 169–185, 2019, doi: 10.1080/02602938.2018.1487023.
- [7] M. Stančić, "Peer assessment as a learning and self-assessment tool: a look inside the black box," *Assessment & Evaluation in Higher Education*, vol. 46, no. 6, pp. 852–864, 2021, doi: 10.1080/02602938.2020.1828267.
- [8] E. Meletiadou, "Exploring the impact of peer assessment on EFL students' writing performance," *IAFOR Journal of Education*, vol. 9, no. 3, pp. 77–95, 2021, doi: 10.22492/ije.9.3.05.
- [9] S. J. Lee and K. Kwon, "Peer assessment as a facilitating and assessment strategy in online and face-to-face classes," *International Journal of Online Pedagogy and Course Design (IJOPCD)*, vol. 11, no. 3, pp. 36–48, 2021, doi: 10.4018/IJOPCD.2021070103.
- [10] Y. Yao, W. Wang, and X. Yang, "Perceptions of the inclusion of Automatic Writing Evaluation in peer assessment on EFL writers' language mindsets and motivation: A short-term longitudinal study," *Assessing Writing*, vol. 50, p. 100568, 2021, doi: 10.1016/j.asw.2021.100568.
- [11] K. Misiejuk and B. Wasson, "Backward evaluation in peer assessment: A scoping review," *Computers & Education*, vol. 175, p. 104319, 2021, doi: 10.1016/j.compedu.2021.104319.
- [12] K. M. Wong and P. Mak, "Self-assessment in the primary L2 writing classroom," *Canadian Modern Language Review*, vol. 75, no. 2, pp. 183–196, 2019, doi: 10.3138/cmlr.2018-0197.
- [13] M. H. Mohamed Jamrus and A. B. Razali, "Using Self-Assessment as a Tool for English Language Learning," *ELT*, vol. 12, no. 11, p. 64, Oct. 2019, doi: 10.5539/elt.v12n11p64.

- [14] S. Yu and I. Lee, "Understanding EFL students' participation in group peer feedback of L2 writing: A case study from an activity theory perspective," *Language Teaching Research*, vol. 19, no. 5, pp. 572–593, 2015, doi: 10.1177/1362168814541714.
- [15] Y. Zou, C. D. Schunn, Y. Wang, and F. Zhang, "Student attitudes that predict participation in peer assessment," *Assessment & Evaluation in Higher Education*, vol. 43, no. 5, pp. 800–811, Jul. 2018, doi: 10.1080/02602938.2017.1409872.
- [16] J. Šafranĳ and J. Zivlak, "Effects of big five personality traits and fear of negative evaluation on foreign language anxiety," *Croatian Journal of Education*, vol. 21, no. 1, pp. 275–306, 2019, doi: 10.15516/cje.v21i1.2942.