International Journal of Instruction *October 2026 ● Vol.19, No.4*
e-ISSN: 1308-1470 ● www.e-iji.net *p-ISSN: 1694-609X*
pp. 677-702

Article submission code: 20260201081508
Received: 01/02/2026 Revision: 01/05/2026
Accepted: 10/05/2026 OnlineFirst: 01/07/2026

Perspectives of English Language Teachers on Fairness and Justice in Assessing Writing

Jayakaran Mukundan
UNITAR International University, Malaysia, *jayakaranmukundan@yahoo.com*

Hui Geng
Corresponding author, English Department, College of Foreign Studies, Guilin University of Technology, China, *huiyuqiankun@gmail.com*

Hutheifa Yousif Turki
English Department, University of Anbar, Iraq, *ed.huzaifa.youssef@uoanbar.edu.iq*

Qu Tang
English Department, Guilin University of Technology, China, *627498339@qq.com*

Vahid Nimehchisalem
Faculty of Humanities and Arts, University of Economics and Human Sciences in Warsaw, Poland, *nimechie@gmail.com*

Mallika Vasugi
UNITAR International University, Malaysia, *mallika@unitar.my*

Neglect of issues in fairness and justice in writing assessment has far-reaching consequences on teachers and learners, giving way to consider the role of teacher training courses, professional development, and psychological state in scoring students' writings. The current qualitative study investigated Malaysian and Chinese English language teachers' perspectives of justice and fairness in assessing writing in response to different tasks. The participants were selected purposively. Multiple methods were followed to collect the data. The participants performed two tasks, which involved them in (i) individually reading two short stories, one related to fairness and the other to justice according to Kunnan's (2014) definition of the terms, (ii) writing their reflections about each task individually, and (iii) participating in group interviews. The findings indicated that participants were more confident about what constitutes justice than fairness. The participants' educational backgrounds and upbringing revealed differences between the Chinese and Malaysian participants. The study has pedagogical and methodological implications.

Keywords: fairness and justice, case stories, teacher interpretations, teacher professional development, writing assessment

Citation: Mukundan, J., Geng, H., Turki, H. Y., Tang, Q., Nimehchisalem, V., & Vasugi, M. (2026). Perspectives of English language teachers on fairness and justice in assessing writing. *International Journal of Instruction*, *19*(4), 677-702. https://doi.org/10.29333/iji.2026.19435a

INTRODUCTION

Numerous changes in writing instruction over the past five decades indicate a gradual movement from traditional, teacher-centered to learner-centered writing classes. In the late 1970s, a paradigm shift appeared in the writing instruction approach. The new writing classroom features urged Hairston (1982) to propose the beginning of The Revolution in teaching writing. The word "humanistic" was central in this new interpretation of what writing classrooms ought to resemble. In writing instruction of the past, teachers adopted a prescriptive view, provided guidelines, and taught learners models of good writing. The new writing classes were inspired by scholars like Elbow (1973), who recommended that teachers should free learner-writers in writing and remind them of their learning responsibilities, leading to the abandonment of product-based approaches. This new way of looking at writing instruction integrates learner-centered writing instruction, using activities such as brainstorming with peers and groups with ongoing assessment routines like peer reviews, teacher conferences, and reviews (Rashtchi & Beiki, 2015; Rashtchi & Porkar, 2020; Tran & Pham, 2023; Nair & Nesamany, 2021), focusing on the recursive process of writing. As a result, assessment and instruction were integrated, never meant to contribute to disharmony (Reid, 1993). In more process-oriented approaches, like portfolio writing, learners decide on the best writing pieces and offer them to the instructor, empowering learners in assessment. Recent research has also shown that teacher response, feedback practices, teacher encouragement, varied strategies, and the use of diverse learning resources can positively support writing development, learner engagement, and students' motivation in English language learning (Hosseini et al., 2025; Sukumaran et al., 2021).

Realistically, while teachers in learner-centered contexts have moved to view assessment as a "non-threatening" issue under most circumstances, teachers in countries like Malaysia and the People's Republic of China (PRC) have had reservations about incorporating excessive doses of humanistic approaches to assessment, which has its roots in their upbringing, traditional values, belief-systems, and their unique educational systems, all of which seem to be as important as Western notions of humanistic beliefs. Many believed it would be nearly impossible to adopt some very humanistic ways of teaching and assessing compositions promoted by the West (Mukundan, 2011). The complexities of implementing educational reforms inspired by Western models in non-Western contexts emphasize that such models often require appropriate adaptation to align with local educational settings and cultural practices (Alhashem & Alhouti, 2021).

Culturally, what assessments mean to people in Malaysia and China would determine how appreciative they are toward humanistic tendencies in teaching writing. Teachers who delay correcting learners' language errors (because they emphasize fluency rather than accuracy to encourage the uninterrupted flow of ideas) often have to abandon this formative and developmental purpose and move toward immediate error correction because of parental and societal pressure. According to Malaysian and Chinese cultural values, errors are like weeds that must be eradicated immediately. There is very little patience on the part of teachers, learners, and parents. Formative and developmental forms of writing assessment, which emphasize less stressful self-assessment, peer

assessment, and teacher conferences and require excessive time commitment on the part of the teacher, also have to be abandoned (Mukundan, 2011).

Literature Review

Humanising Teaching and Assessing Writing

Most of the literature on humanistic writing instruction (teaching and assessment) has been based on early notions of what the nature of humanistic education should feature. Gage and Berliner (1991) believe humanistic education embraces the following:

i. Students will learn what they want and need to know best.
ii. Knowing how to learn is more important than acquiring a lot of knowledge.
iii. Self-evaluation is the only meaningful evaluation of a student's work.
iv. Feelings are as important as facts.
v. Students learn best in a non-threatening environment.

As Reid (1993) puts forth, the principles proposed by Gage and Berliner (1991) follow humanistic views within the writing field. The non-prescriptive nature of teaching within classrooms has led teachers to re-learn their new roles as facilitators. Richmond (1999) discusses her problems as a writing teacher when it comes to student conferences. She tells one of her students who is having problems starting her composition, "I know exactly how you feel. Let me tell you what I do in those situations."

While most writing teachers would consider this a good reaction, Richmond now feels that, after reading up on empathy, this was not the case, and it may involve condescension and manipulation. She now believes that empathy should involve dialogue and that teachers must be listeners first (Bartholomae & Petrosky, 1986). The teacher needs, she says, to get into the writing world of the student writer to draw connections, not merely offer corrections. One humanistic assessment practice that aligns with this philosophy is self-assessment, which offers learners the opportunity to reflect on and regulate their written performance. Ratminingsih et al. (2018) found that implementing self-assessment increased both students' writing competence and learner independence, contrasting with traditional teacher-driven assessment.

Writing Instruction and Assessment in China and Malaysia

There was a revolution in the teaching of writing in the US, which eventually spread across most of Europe, including the 1st world in the Southern Hemisphere; countries in Asia, where English was a Second Language (SL) or Foreign Language (FL) struggled to make sense of the process revolution (Mukundan, 2011). For one, the traditional assessment systems and parental and societal beliefs would not accommodate something as laissez-faire and loose as the process approach and portfolio-based writing. The policymakers in Malaysia and China believed success in writing was achieved when learners obtained good scores in high-stakes public examinations (in Malaysia, the Malaysian Certificate of Education, *Sijil Pelajaran Malaysia* (the SPM -

the equivalent of O levels), and the National Matriculation English Test (NMET) in China).

Recent statistics and studies shed light on the current educational practices in these countries. In Malaysia, while the State schools have incorporated School-based assessments (which replaced two major high-stakes examinations, one at upper primary and one at the end of lower secondary), not much has been achieved trying to decentralize and the "system retains top-down features" (Nimehchisalem et al., 2023, p. 5) which meant the movement towards humanistic trends was hijacked by teachers who then went back to traditional assessment ways which included test preparation routines and learning from models (product based learning). In short, there was no sight of any process learning or formative developmental teaching and assessment, even though some studies have reported promising findings (Mohd Salleh et al., 2023). In China, traditional educational philosophies continue to constrain the implementation of Competency-Based Education (CBE) reforms, creating a significant challenge in aligning classroom practices with educational policy objectives. As detailed by Deng et al. (2024), while teachers may outwardly agree with the principles of CBE, many lack a substantive understanding of the critical competencies required and default to traditional methods of knowledge transmission. Ernawati et al. (2022) also found that multiple factors, including instructional quality, technological readiness, and teacher preparedness, significantly influence students' competence in curriculum implementation, suggesting that competence-based reforms alone are insufficient without supportive classroom and infrastructural conditions. The frequent use of paper-and-pencil tests compounds this adherence to conventional teaching. It reflects the influence of high-stakes assessments like the *Gaokao*, reinforcing an exam-oriented educational culture. Parental pressure further amplifies teachers' reluctance to adopt innovative educational methodologies, thus widening the gap between policy intentions and actual classroom practices. This resistance to embracing new teaching approaches in China mirrors educational challenges in Malaysia, where entrenched assessment methods and educational practices impede the adoption of more holistic approaches despite policy efforts to shift towards these humanistic methods. In both countries, the discrepancy between educational reform goals and the reality in classrooms is starkly highlighted in the findings of Kigwilu and Mokoro (2022). Their research across these contexts reveals that, despite the advocacy for more learner-centered and progressive educational methods, most teachers continue to rely on traditional, teacher-centered instruction. The persistent use of written tests, class exercises, and homework instead of engaging students through inquiry-based learning, project-based work, cooperative learning, and interdisciplinary approaches underscores the misalignment between policy and practice.

The prevalent 'Double Reduction' educational policy, initiated by the Chinese government and authorities in 2021, has attempted to lessen the students' academic burden. Gao et al.'s (2022) research offers an important educational context to better understand the broader potential influences of these policy reforms on students' learning experiences and academic achievements. Their findings suggest the practical implications that, to prevent lowering the educational quality and elevating pressures at

other educational levels, it is necessary to consider the specific actions taken to reduce academic workloads when implementing the 'Double Reduction' policy.

Writing Teachers as Humans

The human condition seems to have consequences not only on test takers but also on teachers. Teachers are not flawless in behavior, especially in decision-making. Part of the reason for using the two cases is that the teachers portrayed in them were deemed uniquely unstable. Teacher attributes, which include their psychology, have been investigated for the past three decades. Researchers have been evaluating the relationships between fairness and numerous associative factors that could influence learning and teaching. Would that mean that aspects such as the emotional stability of teachers can affect fairness? Ghanizadeh and Moafian (2010) investigated the relationship between Emotional Quotient (EQ) and success in teachers as professionals. They found significant correlations in 12 of the 15 components (self-regard, self-actualization, independence, empathy, interpersonal relationships, social responsibility, reality testing, flexibility, stress tolerance, impulse control, happiness, and optimism) that formed the EQ Test. This finding confirms the work of researchers like Goleman (1996), who hypothesized that individuals with higher-than-average EQ are usually more successful.

In contrast, Emotional Intelligence (EI) deficiency can lead to a lack of success in tackling emotional problems, affecting job performance. Higher EQ indicates one's potential for academic and professional success. Ghanizadeh and Moafian (2010) suggested incorporating skills associated with EI into teacher professional development programs so that younger teachers can handle their emotions and "shift undesirable emotional states to more productive ones" (p. 431).

Teachers are like nurses in a job sector where experiencing stress is part of a typical day. Teachers not only teach, but they also have to prepare lessons, grade homework, and do administrative work. Mukundan et al.'s (2015) study on burnout among school types (a national type with Malay as a medium of instruction and two vernacular schools, including Chinese and Tamil, involving 714 Primary school teachers) showed that teachers from Chinese Medium schools (SKJC) suffered from burnout in the domain of reduced personal accomplishment when compared to those in the SK and SKJT schools. The researchers have attributed the finding to the culture of Chinese Medium Schools and Chinese education, revealing excessive dependence of learners on teachers, which resulted in teachers experiencing pressure. Besides, dealing with excessive feedback on learners' homework could also be a source of teachers' burnout.

The Contribution of Newer Explorations on Assessment and Validity

Assessment experts have redefined validity as they believed it lacked aspects of humanism and ethics. Kunnan (2004) introduced his ethic-inspired Test of Fairness Framework (TFF), incorporating two principles: (1) The Principle of Justice (The test should treat everyone with equal respect) and (2) The Principle of Beneficence (Tests should bring about good and harmony in society). He stressed that the Principle of Fairness (first) is before the Principle of Justice (second) and that a discussion on the

second would be irrelevant if the first principle were not satisfied. Kunnan (2014) stated that new elements, such as fairness and justice, would take the concept of test validity to another level. However, ethics-based approaches, inclusive of fairness and justice, have been slow in progression, with Spolsky (1995) stating that for most of the twentieth century, social, economic, and political concerns (rather than ethics) dominated discussion among testing professionals.

However, assessment literacy, which provides teachers and test developers with issues such as fairness and justice, cannot be overlooked in how they should score students. Most students receiving training for teaching at the Diploma, B.Ed./BA, M.Ed./MA/PhD levels are usually exposed to assessment literacy. There are generally three approaches: the Traditional-Engineering Curriculum, the Innovative Design Curriculum, and the Ethical-Critique Curriculum. In terms of assessment, for courses developed, the focus is mainly on the first approach, the Traditional-Engineering Curriculum, with influences from the second, the Innovative Design Curriculum. The third approach, the Ethical-Critique Curriculum, which is the newest of the three approaches, is generally ignored (Kunnan, 2019). This ignorance of ethical aspects of assessment practice has revealed weaknesses in learning-teaching- assessment.

The present study explored teacher awareness of aspects of fairness and justice in a critique of two cases (short stories), the evaluation of the participants' narratives using the Principle of Fairness (with four sub-principles), and the Principle of Justice (with two sub-principles). Writing is a field where teachers have shown a deep concern for issues related to fairness and justice, as many aspects of subjectivity are involved. Fairness is the first virtue of writing assessment (Elliot, 2016), meaning that validity is at stake if teachers are unaware of subjectivity issues.

As used in this study, case studies are gaining momentum in EFL contexts, and not surprisingly, many teachers find them valuable (Roell, 2019). Teachers need to be deeply involved in cases while at Teacher Professional Development (TPD), something not very common in the profession when compared to the vast extent of its use in the legal and medical professions. Cases are stories, and in this study, while both stories are classified as fiction, they are seen as simulations of what goes on in real life. Shulman (1992) illustrates cases as accounts of practical or strategic dilemmas that confront a teacher; however, he is cautious about their contribution by engaging teachers. As Shulman argues, the narrative must represent a class or type of dilemma or quandary that is common in situations where teachers work. Another interesting point about cases is that they show little respect for disciplinary boundaries (there is a potent mix of education, sociology, psychology, and many other areas), an issue that researchers believe is excellent in modern-day education. Such characteristics amount to a complex tapestry of issues, interwoven, messy, and recalcitrant, and rarely admitting a single correct answer. In the context of this research, both the Catherine Lim stories, The Teacher and Adeline Ng Ai Choo (Lim, 1978), set in Singapore writing classrooms, fulfill all the requirements of what is deemed effective and engaging cases. First, these writing classes are set in Singapore classrooms, where learners prepare for their Singapore-Cambridge General Certificate of Education (GCE). Writing for the GCE has been widely seen around the world as harshly assessed, especially in the context of

colonized countries like Malaysia and Singapore, where the demands placed on FL learners of English were unrealistic (Mukundan, 2011). This view was accurate as teaching did not complement assessment. Secondly, these stories have come at a time (in the 1970s) when there were significantly large numbers of suicide cases among the student population in Singapore, especially between October and November, corresponding to the examination months (Ung, 2003).

The lack of focus on TPD and the lack of concern for enhancing teacher competencies (Nguyen, 2015) were a few reasons for the current study. While substantial literature on the theory and practice of humanistic writing instruction exists, there is a lack of research investigating how these practices are perceived and implemented in non-Western contexts. Therefore, this study aimed to fill this gap by exploring Malaysian and Chinese English language teachers' perceptions of fairness and justice in assessing writing. This empirical study offers insights into the dynamic connection between cultural values, pedagogy, and writing assessments in educational settings. The researchers proposed the following research questions to meet the study's objectives:

1. To what extent did the Narratives (N) and subsequent Group Interviews (GIs) indicate Malaysian and Chinese English language teachers' awareness of Fairness and Justice?

2. How did the participants feel they benefited from the two cases in terms of TPD?

Ethics approval and consent to participate (Exemption Statement)

This project involved a small group of adult postgraduate teachers in classroom settings and posed no more than minimal risk; accordingly, it qualified for exemption from formal institutional review under the authors' university guidelines. All seven participants provided written informed consent before participation. Participation was voluntary, non-graded, and could be withdrawn at any time without penalty. Pseudonyms are used throughout, and all data were securely stored on password-protected devices accessible only to the research team.

METHOD

The study benefited from qualitative methods to collect and analyze data. The participants, data collection, and data analysis procedures are explained as follows.

Participants and Materials

Participants

This research was conducted in a public university in Malaysia. Seven postgraduate students, all female, were purposively selected from an intact group of 26 on a master's course. The seven selected were practicing teachers with the greatest number of years of experience. The teachers had enrolled in LHE 5256 Professional Development in TESL, a three-credit course in the postgraduate program. Only one had a B.Ed. TESL degree and had taken a course on the Teaching of Writing. The rest had undergraduate degrees in Applied Linguistics. The seven were Jan, Cindy, Karen, Randy, Suzy, Bonnie, and Sandy – (all pseudonyms). Jan, Sandy, Suzy, and Bonnie were from the People's

Republic of China (PRC), while Randy, Karen, and Cindy were Malaysians (MY). All seven teachers had taught writing for at least five years. This minimal-risk study, conducted in regular classroom settings, met the criteria for ethics exemption at the authors' institution. All participants volunteered and signed individual written consent forms before data collection; participation was not linked to course assessment.

Materials (Case Stories)

The materials were two short stories (case stories), Adeline Ng Ai Choo and The Teacher by Catherine Lim (1978) in her book Little Ironies: Stories from Singapore. While the book was used as a set text for literature in the GCE in Singapore, ironically, there has been no mention of it being used for teacher training, especially in professional development. Appendix A shows the summaries of the two stories.

The selection of the two stories (considered as Cases) would be in line with the recommendations of Kunnan (2019, p. 331), who illustrates what teaching-learning activities in the Ethical-Critique should include; i) making moral or ethical judgment while working on hypothetical scenarios or case studies; ii) evaluating assessment using ethical principles; and iii) writing critiques of assessment on assessment practices. The selection of the two cases as material for the study and the prompts fulfilled all three of Kunnan's requirements. The material for the study and the prompts fulfilled all three of Kunnan's requirements.

Data Collection

The first part required all seven participants to read the two short stories, Adeline Ng Ai Choo and The Teacher (considered cases in this study), and then respond to the task: Comment on both stories: Was there Fairness and Justice? They were given 48 hours to work on their narratives and were told not to exceed 1500 words.

Then, the participants were invited to three Group Interviews (GIs) that lasted two hours each. The first author was the moderator. During the GIs, the researchers took field notes. The prompts for the GIs were structured along the questionnaire items used by Wallace and Qin (2021), which were based on their Model of Language Assessment Fairness and Justice Judgment of Social Entity. Two experts validated the procedures for extracting the main themes, and data gathered from Group Interview sessions underwent member checks. All narrative and interview data were de-identified at source, stored on password-protected drives accessible only to the research team, and reported using pseudonyms to preserve confidentiality.

Data Analysis

Content analysis was carried out on the data with a thematic approach. The researcher used deductive coding to address the first research question, and the narratives were analyzed to check participants' awareness of the Principles of Fairness and Justice. Kunnan's (2014) Principles of Fairness (with four sub-principles) and Principles of Justice (with two sub-principles) were used for coding purposes, and the researchers analyzed the narratives to find out if the participants were aware of the two fundamental principles. The coding process was based on Kunnan's (2014) framework. The

Principle of Fairness has four sub-principles: 1) An assessment ought to provide adequate opportunity to learn the knowledge or abilities that are to be assessed; 2) An assessment ought to be consistent and meaningful in terms of its test-score interpretation; 3) An assessment ought to be free of bias against any test taker groups, in particular by assessing construct-irrelevant matters; and, 4) An assessment ought to use appropriate access, administration and standard setting so that decision-making is equitable to test-taker groups. The Principle of Justice has two sub-principles: 1) An assessment institution ought to benefit society by making a positive social impact, and 2) An assessment institution ought to advance justice through public reasoning of their assessment. The coding done by the researchers was then validated by two university lecturers who taught assessment. Data collected for the second research question was analyzed through inductive thematic analysis, progressing from open coding (textual data to phrases), then to axial coding (which involved linking the codes, and finally to selective coding (selection of the core themes that emerge).

FINDINGS AND DISCUSSION

The seven postgraduate participants wrote a 1500-word narrative, each responding to the prompt: Comment on both stories: Was there Fairness and Justice? The researchers used the data from the analyses of the narratives and the subsequent group interviews to answer the research questions:

Research Question 1: To What Extent Did the Narratives (N) and Subsequent Group Interviews (GIs) Indicate Participant Awareness of Fairness and Justice?

Table 1
Shows the frequency of references made by the participants and coded accordingly to the principles of fairness and justice

Frequency of References								
Participants	Principle of Fairness (PF)					Principle of Justice		
	PF 1	PF2	PF3	PF4	Total	PJ1	PJ2	Total
1. Jan (PRC)	0	1	0	2	3	4	2	6
2. Cindy (MY)	0	1	0	3	4	2	2	4
3. Karen (MY)	0	1	3	0	4	1	2	3
4. Randy (MY)	0	1	2	0	3	2	1	3
5. Suzy (PRC)	0	2	1	0	3	1	4	5
6. Bonie (PRC)	3	2	1	2	8	2	1	3
7. Sandy (PRC)	0	2	2	1	5	2	0	2
Grand total	3	10	9	8	30	14	12	26

Kunnan (2004) states that there is a difference between fairness and justice - only a test and its administration may be evaluated for being fair or not, while justice judgments are reserved for the organization that delivers that assessment and uses its scores. The characteristics of all the narratives resembled this.

The narratives were analyzed to identify the PFs (Principles of Fairness) and the PJs (Principles of Justice). PF1 recorded three occurrences; PF2, a total of 10; PF3, a total of 9; and PF4, a total of 8. In total, there were 30 occurrences representing the PFs (across four Principles). PJ1 recorded a total of 14 occurrences, while PJ2 recorded a total of 12. In total, PJ had 26 occurrences (across only two principles). The data

revealed that the participants had a greater awareness of the Principles of Justice (PJ) (mean of 13) as opposed to the Principles of Fairness (PF) (mean of 7.5).

Data from Group Interview 1:

The Group Interviews (GIs), conducted over three two-hour sessions each, revealed that, like in the narratives, references toward PFs lagged PJs among the Chinese (PRC) and the Malaysians (MY). The moderator probed them further on this (especially on teacher incompetence) as found in the story The Teacher. The participants, especially the four from the People's Republic of China (PRC), felt there was nothing wrong with the teacher (which meant they did not detect most of the teacher's lack of fairness) across all principles (PF1-PF4). They also viewed assessment primarily as rule-bound, with a clear line between pass and fail. For example, some participants defended strict grading and the unchanging application of assessment criteria:

Suzy: I think the teacher was correct in failing her – how can you use bad language and still pass?

Jam: What's wrong with being strict? The stricter, the better!

Sandy: Suicides are common in China, where the weak take their lives. But life goes on; school rules and examination results are to be respected, and people forget these things.

Bonnie: The criteria used for assessment must be maintained and cannot be changed at any cost.

While the PFs were less in focus in the narratives, the PJs were sufficiently detected, and overall, all seven participants were very explicit in their views regarding the changes that must be in place. All seven indicated in their narratives that institutions must be more transparent and willing to engage learners and parents so that tragedies can be averted in the future:

Randy: Schools must involve parents and teachers in assessment practices.

Sandy: Schools in China have never considered parental voice important. They should.

Most references to PJ1 and PJ2 were present in the narratives when participants debated the teachers' actions in *Adeline Ng Ai Choo*. At the special meeting, the teachers and the school administrators were engaged in discussion to "justify their actions and suggest improvements to current practices" (Karen, MY), while "some of the other teachers were against returning marked scripts to learners and making scoring clearer (a clear pass or a clear fail, without borderline scores)" (Cindy, MY). At GIs, all the participants expressed their disgust at the teachers featured in *Adeline Ng Ai Choo,* who they claimed was "more interested in defending themselves than being sorry or remorseful for the death of Adeline" (Randy, MY).

Suzy: It is surreal. It is clear that this must be considered a period of mourning, but the teachers at that extraordinary emergency meeting looked at it as an opportunity to discredit the dead girl. Where is the compassion you must show as a teacher?

Sandy: None of the teachers, in fact, showed any "soft side" – they looked like prison warders and did not show any sympathy.

Data from Group Interview 2:

The second GI was focused on some probes, with the moderator (lead researcher) attempting to gauge teacher attitudes towards teacher actions in the two stories to determine if they were aligned to such major principles of acquisition/learning as themes: "learner-centredness," considered as "humanistic," or if learning environments were "non-threatening."

These Group Interviews (because there were three) saw changes in the perceptions of the seven participants. They started very strongly in Group Interview 1 (the teacher was correct in failing her) and mellowed in Group Interviews 2 and 3, where they began to see the negative traits of the teachers (Yes, only reinforcement of the negatives).

Detailed analysis based on the four Principles of Fairness and two Principles of Justice:

Principle of Fairness 1 (PF1):

An assessment ought to provide an adequate opportunity to learn the knowledge or abilities to be assessed.

Bonnie was the only one among the postgraduate participants to register references to PF1 in her narrative (3 instances). This was clearly due to her background (her degree was in TESL, and she studied a component of Teaching Writing). According to Bonnie:

Bonnie: One of the goals of classroom assessment is that teachers should effectively use the results to make appropriate adjustments to teaching and give reasonable feedback to students so that they benefit from it.

Similarly, in the Group Interview, Bonnie elaborated on this by saying that the protagonist Tan Geok Peng in The Teacher never had "the opportunity to get proper feedback, and this, in turn, made her even more depressed – she was merely getting confirmations on her lack of ability by getting an F grade continually" (GI). Bonnie declared feedback was "the closest assessment got to teaching; in fact, it linked assessment to teaching" (GI, Appendix B).

In her narrative, Bonnie also believed that fluency training was vital to writing instruction and that teachers should "aim to establish a learning environment that motivates learners to take risks in their writing" (Bonnie, N). In the Group Interview, Bonnie argued that the focus on accuracy (which was all about grammatical accuracy) showed that the part in writing, fluency, which is equally important, was "totally neglected" (GIs):

The findings of this study echoed Fulcher's (2013) thoughts that assessment can be viewed as a controlling or gate-keeping tool (very top-down and serves the needs of harsh and threatening learning environments run by the powerful). The lack of references that connected teaching and learning to assessment was significant because only one participant could conceptualize learning as associated with assessment. Similarly, studies by Elshawa et al. (2017) and Nimehchisalem et al. (2019) also

revealed that Fulcher's classification of test purpose (test as a gate-keeping tool) was the topmost in teachers' priorities, partly because of their lack of commitment to the use of process-based approaches in teaching writing, which promoted learner-empowerment (Rashtchi et al., 2019; Walker, 2017). Such strategies like non-direct instruction (Rogers & Freiberg, 1970), student engagement through peer and teacher conferences in social learning (Vygotsky & Cole, 1978), and the creation of workshop-like class settings (Murray, 2004) were not offered by the participants as alternatives either in their narratives or at GIs.

Principle of Fairness 2 (PF2):

An assessment ought to be consistent and meaningful regarding its test-score interpretation.

We recorded ten instances of this principle, most of which referred to the teachers' dilemma in the story of *Adeline Ng Ai Choo*. All seven participants supported the teacher and believed that giving extra marks so that students who are borderline cases can pass brings about a lack of consistency in grading, and this is not acceptable, and that scoring regulations and rubrics have to be followed.

Sandy: I agree that marks should not be given simply due to a student begging… on the one hand, it would defeat the original purpose of having the exam… it would have been really unfair to the other students who were seriously studying.

Moderator: Let's go back to what we discussed earlier. Was the teacher's conduct when scoring the compositions correct? Let's not talk about the teacher giving free marks for a passing mark, but about the teacher's lack of emphasis on content.

Bonnie: I believe Tan Geok Peng did well in her content, but the teacher seemed to have ignored her strengths.

Jan: I agree. This was never a learner-centered class, with no positive reinforcement by the teacher when the learner did well.

Randy: Yes, only reinforcement of the negatives.

Cindy: Yes, only the negatives – all about weak language use.

Moderator: So, do you think PF2 was misplaced here? Do you think the assessment was valid?

Suzy: I guess not. I cannot remember everything about scoring rubrics, but I believe equal or almost equal allocations are given for language and content, but this (content) was never in the teacher's mind.

Moderator: If the total mark was ten and Tan Geok Peng scored 2 for Language, would you pass her by giving her a 3 for content?

(Participants agreed.)

Randy made one reference (f=1) that implied associations to PF2 and had different beliefs concerning teacher actions in the two stories. She believed Adeline Ng Ai

Choo's teacher was right in not awarding extra marks for her to pass: "I agree with the teacher as scores must indicate what a student learned, understands, and is competent in doing, and this must be universally accepted and respected" (N). Randy, however, was against Tan Geok Peng's Teacher, whom she describes as "wrong in only exaggerating on his student's grammatical mistakes" (N). Randy believed the validity of the assessment was compromised as the content was not given the weightage it deserved, and grammar seemed to be assessed rather excessively and exclusively. In the Group Interviews, Randy, like all the other participants, believed that the suicide of both protagonists in the two stories did not sway her into believing that if these tragedies happened, the teachers were at fault. However, Randy believed Tan Geok Peng's teacher was utterly wrong regarding scoring. She believed content could have almost the same weightage as language, as indicated in writing assessment instruments like the Wong Qualitative Writing Scale (Wong, 1989).

In some cases, she believed that the allocation of marks for content would be more prominent for content than for language use (Jacobs, 1981). Furthermore, current studies have continued to explore the relative importance of content and language in writing assessment. For example, the Cambridge English guidelines on assessing writing performance at level B1 (Cambridge English Assessment, 2024) emphasize that content, organization, and language are equally indispensable elements in marking writing. They promote a balanced approach to writing assessment, ensuring students' ability to communicate relevant content effectively is as essential as their grammatical accuracy. This approach aligns with contemporary educational standards that aim for holistic evaluations of a student's writing ability, reflecting a comprehensive perspective of language proficiency.

Principle of Fairness 3 (PF3):

An assessment ought to be free of bias against any test-taker groups, particularly by assessing construct-irrelevant matters.

In total, nine references were made to PF3. Two in the group, Jan and Cindy, did not show any focus on PF3. Karen, with three (3) associations with PF3 in her narrative, stated explicitly that "Ai Choo was not a weak student from the beginning. She became unstable when she entered Secondary 3" (N). Karen then states that Ai Choo "got progressively weaker and weaker" (N) and that this slow decline, which the teacher admitted she had witnessed all along two grade levels, did not trigger any alarm within her. During one Group Interview, Karen read from the story Adeline Ng Ai Choo (page 65), the part where the teacher (Mrs. Tan Cheng Kwee) tries to detach herself from any blame at the inquiry (See Appendix C).

Karen claimed that Ai Choo should have been categorized "at risk" (GIs) by the class teacher and the school, and proper counseling procedures should have been put in place two grade levels ago. The teacher, Karen claimed, then went on to say that Ai Choo "didn't seem to be able to concentrate" and was "nervous and distracted" and that the teacher claimed she was doing "irrelevant things in class like scribbling in her diary" (GIs). Karen stated that the final observations of Ai Choo, as commented by the teacher, showed that "she possibly was already clinically depressed" (GIs) and that she would

have developed suicidal tendencies. "Clinical depression, as far as I know, set in after a long period" (GIs), she said, and for the teacher and the system not to have found her "at risk" (GIs) defies logic. Karen clearly showed how "Ai Choo being at risk should not be subjected to tests as life was already testing her" (GIs).

All the references to PF3 were linked to the protagonists in both stories, something all participants at GIs considered at risk. However, no strong associations were made by the seven participants to related aspects in teaching/assessing writing, such as writer's block (Bastug et al., 2017; Boice, 1993; Cohen, 1998), writing apprehension (Elif & Yayli, 2019; Hassan, 2001) or writing anxiety (Cheng, 2002; Choi, 2013) although both the stories were set in writing classrooms which were considered threatening to learners, which revealed that content knowledge about the teaching of writing and assessment was found to be lacking in participants (except Bonnie) (See Appendix D).

Principle of Fairness Four (PF4):

An assessment should use appropriate access, administration, and standard setting to make decision-making equitable for test-taker groups.

There were, in total, eight references identified. Bonnie stated that teachers need continuous professional development and "need to be specialists in subject matter" (N). In the Group Interview, she illustrated this point by saying that "while Tan Geok Peng's teacher was possibly to be faulted for rating failure (ignoring the content and penalizing the grammar errors excessively)" (GIs), the bigger flaw in the teacher is in his "inadequacy in the knowledge of composition" (N). Bonnie stated, "If we had teachers like him marking scripts for public examinations, bright students might fail" (GIs). Tan Geok Peng's teacher gave her a straight F9, stating that her essay was "out of point" (Bonnie, N), stating that the title he gave for the class to write was *The Stranger,* but Tan Geok Peng had written a "great deal of trash about her father" (Bonnie, GIs). Bonnie pointed out that the writer "made a mockery out of the teacher," Tan Geok Peng manipulated the title to reveal her unique ability to think and compose metaphorically (GIs).

> *Moderator: Do you think inadequate knowledge in content areas like composition could be detrimental to learners? Would you have preferred your Public Examination Compositions to be marked by Tan Geok Peng's teacher?*
>
> *All: No!*
>
> *Bonnie: Tan Geok Peng's teacher could not even appreciate the learner's manipulation of the title, The Stranger. She had made the stranger her father (reflecting her real-life problems at home).*
>
> *Randy: I think I read too fast and failed to spot this at first. Yes, this is clever manipulation. Shows that Tan Geok Peng did have unique ideas.*

Administration of assessment at the school level seemed to be discussed by the participants in their narratives mainly from the perspective of teacher bias, especially with regard to the discrimination by the teacher (in the story The Teacher), where excessive focus on assessment seemed to be on aspects of language rather than content.

However, none of the participants, through their narratives, discussed aspects of PF4 that had importance to this principle: the inadequacy in teacher knowledge, the lack of competence (Bourgonje & Tromp, 2011), teacher limitations in terms of Teacher EQ (Goleman, 1996) or Teacher EI (Ghanizadeh & Moafian, 2010). The participants expressed none of these. The 3rd and final GI did see enlightenment among the participants, who started believing that perhaps the administration of writing programs in schools and the administration of assessment was flawed and that teachers probably need retraining.

All seven postgraduate participants of the study were more expressive when discussing aspects that were associated with justice. All seven participants made at least one reference associated with PJ1, while six out of the seven had references associated with PJ2.

Data from Group Interview 3:

Principle of Justice 1 (PJ1):

An assessment institution ought to benefit society by making a positive social impact.

There were 14 references to PJ1 in total. Generally, the seven writers wrote rather enthusiastically on justice, recommending that the quest for grades alone should not be a priority in life. Almost all the writers mentioned parents contributing to their children's despair regarding assessment. Jan, who made four associations to PJ1, mentioned that a single score cannot "show everything" and that parents must realize their children's priorities by asking themselves: "Their marks or growth?" (Jan, N). At Group Interview 1, Jan was very emotional, stating that schools in her home country, China, should try to engage parents and discuss the disasters that have struck children when "marks are the only thing that mattered, and happiness sacrificed" (GIs). Bonnie, in her narrative, said that suicides were common in China among school-going children and that just before she left for studies, there was a case in her town "where a student took her own life after being reprimanded by the teacher for not doing her homework before the summer break" (N). While "there are now advertisements on TV in China which want schools and parents become more aware of creating balance in learners' lives by having more recreation and sport, these views develop slowly, and even summer holidays are not spared as kids are given homework" (Suzy, GIs).

> *Suzy: Parental involvement (especially in assessment matters) is important. In China, parent-teacher meetings mainly add more anxiety to learners.*
>
> *Sandy: Yes, for weaker students, it actually punishes them more. Teachers magnify learners' weaknesses in front of their parents.*
>
> *Karen: In Malaysia, it isn't that bad, but most teachers do not even provide suggestions as to how parents can help with improvement, like perhaps getting the learner to read books he/she likes.*

Principle of Justice 2 (PJ2):

An assessment institution ought to advance justice through public reasoning.

In total, 12 references were made to PJ2. All seven participants mentioned the need for assessments to be more transparent. The participants (three Malaysians and four Mainland Chinese) believed that the lack of transparency in assessment in China and Malaysia led to many misunderstandings about assessment. Bonnie and Suzy (China) believed that "rating scales for composition are seldom made public, and this would result in suspicion about scoring" (GIs). Karen (MY) was also critical of the "secrecy surrounding marking schemes" (N), which "is rampant" and "scales used for scoring are usually classified" (GIs). Karen admitted that unless "one was an examiner for public examinations, one would never visualize the instruments used for scoring" (GIs). Suzy believed that in the story *Adeline Ng Ai Choo*, "the school raised the issues of what had to improve (with regards to assessment) only after Adeline Ng Ai Choo's suicide" (GIs) only because "to deflect blame that was placed on the school" (N). They took a very defensive posture, and the Head, Ramalingam, "only thought of the bad name" it brought to the school (N). During the Group Interview, Suzy expressed that even during the school meeting, the school staff made further attempts to make tests "become even more cloaked in secrecy" (GIs). The school Head suggested not returning marked scripts to students (to solve the problem of disputes), which Suzy thought "contributed even more to the lack of transparency in assessment" (GIs).

Suzy: No one should hide Composition Evaluation instruments from learners.

Moderator: Do you think it is an excuse on the part of schools and the Ministry to hide these instruments so that they do not invite disputes?

Cindy: That's so true. So only teachers who are examiners of public exam scripts have access to knowledge (which by right should be accessible to all, especially the exam candidates).

Bonnie: When it is about exams, secrecy is encouraged, which is discriminatory. This does not reflect learner-centeredness at all.

Research Question 2: How Did the Participants Feel They Benefitted from the Two Cases in Terms of TPD?

The participants' experiences reading and responding to the two case stories, *The Teacher* and *Adeline Ng Ai Choo,* and their perceptions of how they benefited were done using inductive thematic analysis. Recurring ideas and notions were identified, grouped, and classified into themes.

The following themes emerged from data obtained from the Group Interview (GI):

Stories Used as Cases Bring New Meaning to Teacher Professional Development

The participants believed the two stories were a perfect fit – both complemented each other and were simulations of reality (Shulman, 1992). Both were tragedies "induced by many factors," which also made "reasoning complicated, as there were dilemmas" (Karen, GIs). When discussed in groups, the members hardly agreed on who was wrong as they "were constantly challenging each other's stand" (Randy, GIs) since no correct answer existed in the cases (Shulman, 1992). Both cases were also about teaching writing, complementing each other. *The Teacher* focussed on teachers' and learners'

frustration with the challenges in a writing class, while *Adeline Ng Ai Choo* was all about the aftermath from assessment failure "as the staff meeting that followed the death of the protagonist was all about individual teachers deflecting blame from themselves" (Sandy, GIs) and "justifying actions they and the school authorities took" (Cindy, GIs).

> *Moderator: All of you seem to have moderated your views concerning the teachers in both stories.*
>
> *Jan: Like most Chinese teachers, I initially felt that "assessment principles cannot be messed around with." They are clear and "written in stone by the powers that lead us."*
>
> *Randy: I started changing my perceptions of the learners involved. I developed more sympathy for them after reading and analyzing more of them. I discovered increasingly more inadequacies regarding the teachers, the schools, and the system.*

Assessment in Writing Is Complex (When It Involves the Process); Thus, Teachers' Clinging to Traditional Ways Must Change.

Jan suggested that Teacher Professional Development must include cases through which participants realize that "assessment within a process approach was a different kettle of fish" (GIs). She stated that assessment and teaching "seemed mutually connected" (Jan, GIs). Bonnie confirmed that "TPD activity like this never existed in China (PRC) and that even if it did, writing instruction and writing assessment were "never viewed as connected" (Bonnie, GIs). She believed that timed writing, as a product-oriented approach, was emphasized rather than process writing, which is more formative and developmental. Bonnie believed that the sufferings of the protagonist Tan Geok Peng, who was "clever and had good ideas but was let down by poor grammar" (GIs), were why Bonnie thought about how she, as a teacher, should change. She stated, "Ignoring the process, which included making writing recursive and involved peer review and teacher conferences with students, benefitted teachers (because it saved them time), but it discriminated against learner-writers" (Bonnie, GIs), especially since writing was a difficult skill to acquire. Bonnie believed that because writing assessment in China was "utterly detached from the teaching" (GIs) there was "bigger emphasis on judgment rather than diagnostics" (GIs).

Cindy, Karen, and Randy, Malaysian teachers, believed that after more than four decades of trying "to force-fit" (Cindy, GIs) the Process Model into Malaysian schools, all they ever faced was the frustration of failure. All preparations for public examinations were "in the form of practice involving timed writing which simulates the actual examination" (Randy, GIs). There was no process in place as the teachers claimed they were "pressed for time, dealing with mixed ability classrooms and teaching loads which are heavy" (Cindy, GIs). The failure of the writing curriculum, which disconnects assessment from learning (and which ought to be humanistic and process-oriented), especially so in SL /FL situations, would flout all the Principles of Fairness and Justice (Kunnan, 2014).

CONCLUSION

While the researchers were analyzing the data, they expected to come across some promising responses and themes. Some themes that did not emerge were 'learning-oriented assessment,' 'assessment for learning,' 'self-assessment,' and 'collaborative assessment.' As evident from the results, the participants regarded the testing contexts in the stories as summative assessments; however, they could have also considered the setting formative. One of their justifications for this view was the lack of time and the pressure on teachers to complete the syllabus on time and prepare their students for exams, indicating that teachers have a misconception of classroom-based assessment and their failure to realize that encouraging self-feedback and peer feedback, for example, can be a game-changer in the process of assessment. Practicing self-feedback, for instance, would clearly show the success criteria to the student writer and would leave little need for unnecessary teacher comments.

Another theme that did not emerge was the notion of differentiated instruction, which is tightly intertwined with the concept of formative assessment (Nimehchisalem, 2024). In a student-centered, inclusive language classroom, students are not expected to learn from the teacher; rather, teachers and students learn together and with one another. Students are considered individuals with diverse personalities, skill sets, interests, and learning preferences. Therefore, differentiated instruction and individualized, learning-oriented assessment need to be addressed, in which self-feedback, peer feedback, and teacher feedback provide a feed-forward loop to scaffold students' next zone of proximal development.

Related to differentiated instruction and individualized assessment, a fair assessment of students does not mean considering them equal since all students are not equal (Denbo & Williams, 2016). Diversity is undeniable. Two students may have similar proficiency levels but have different opinions about the difficulty level of a specific test type. The current study participants also failed to provide any data addressing this concept, which is also missing in the findings of the previous research (Nimehchisalem et al., 2019; Nimehchisalem & Nur Izyan Syamimi, 2018). In these studies, the data from association tasks showed that the term 'assessment' is frequently associated with 'tests,' 'marks,' and 'anxiety'. At the same time, themes that did not occur with an extremely low frequency included 'learning,' 'feedback,' and 'improvement.' It seems to the researchers that the research settings are far from the learning-oriented, inclusive, democratic, non-threatening, and empathetic assessment practices designed to engage learners and prompt collaboration and self-regulation (Masuku et al., 2021).

Theoretically, studies similar to the present one would confront teachers and researchers with more challenges in understanding fairness and justice. It would also lead them to determine if there is unfinished business within fairness and justice. For example, the investigation of fairness in classroom assessments by Rezai (2022) stressed the dynamic principles of equity and interactional fairness, which could be linked to the challenges discussed in this reflection section. Rezai's (2022) research provided empirical evidence that fairness is perceived differently based on contextual factors such as cultural and educational backgrounds. This finding can align with the present study regarding

fairness and justice, suggesting that integrating empathy into assessment validity could address such complexities, thereby enriching learners' experience. This insight into diverse perceptions of fairness can help guide future educational policy assessment practices to be more humanistic, fair, and empathetic.

A final point to add is the theme of 'sympathy' that emerged from the GIs. In the assessment literature, sympathy has been associated with 'rater leniency' (e.g., Dewberry et al., 2013). The construct that is closely related to sympathy but is slightly different is empathy. Empathy has been widely researched and well-defined in the area of educational psychology. For example, Lam et al. (2011) break empathy down into its: (i) cognitive component, "one's ability to take the perspective of others, and see the world through their perspective"; (ii) affective component, "involves experiencing the feelings of another person"; and (iii) behavioral component, "involves verbal and non-verbal communication to indicate an understanding of an emotional resonance with the other person" (Lam et al., 2011, p. 163). Empathy in education is defined by Boyer (2010) as "the ability to interpret signals of distress or pleasure with effortful control" (p. 313), which underscores the unique challenges teachers face in recognizing subtle signs of distress or pleasure among many students while maintaining control to minimize additional stress. Empathy emerges as a crucial yet often overlooked variable in assessment validity studies. Without empathy, teachers will fail to understand the extent to which caring for students is vital. The aspect of care is more prominent in classroom empathy (White, 1999), highlighting its importance in creating supportive and effective assessment environments. By incorporating empathy, teachers can better address individual student needs, fostering a more inclusive and supportive learning atmosphere. Thus, empathy may arguably be a significant feature of fair and valid assessment and has potential for further research.

Several conclusions can be drawn from this research pertaining to teachers' awareness of the perspective of the Principles of Justice and Fairness, the foremost being that awareness is more outstanding in aspects of justice than Fairness, which can be attributed to the lack of knowledge on assessment itself, especially in the assessment of writing or composition. The four Principles of Fairness features require teachers to have competencies in writing assessment – distributive, procedural, and interactional (Wallace & Qin, 2021), which the present study participants lacked. Most teachers with applied linguistics degrees (like in this research) admit that assessment is not the focus of their programs and believe that assessment is not given the attention it deserves. This view was made worse by the fact that while ongoing teacher professional development was scarce, courses on assessment did not feature at all in TPD agendas. New teacher professional development programs would need to address these inadequacies.

It would also be necessary to view writing assessment as a priority and encourage educators and policymakers to foster teacher training courses through which teachers practice incorporating formative humanistic methods instead of summative assessments. Such training is necessary for countries like Malaysia moving toward School-Based Assessment (Nimehchisalem et al., 2019). The shift in assessment would lead teachers to apply aspects of learner-centered process approaches with more enthusiasm.

One limitation of the present study should, however, be acknowledged. The participants in this study were selected from an intact group of postgraduate students in an MA Professional Development in TESL course. However, even with purposive sampling, selecting participants with general teaching and assessment content knowledge and comprehensive experience in the teaching and assessment of writing was challenging. The study only benefited from one participant with a TESL background, who, in undergraduate study, had to do a three-credit course in the Teaching of Writing.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of UNITAR International University, Malaysia. The publication of this research was made possible through research grants provided by UNITAR International University.

REFERENCES

Alhashem, F., & Alhouti, I. (2021). Endless education reform: The case of Kuwait. In A. W. Wiseman (Ed.), *Annual Review of Comparative and International Education 2020* (Vol. 40, pp. 345–367). Emerald Publishing. https://doi.org/10.1108/S1479-367920210000040019

Bartholomae, D., & Petrosky, A. R. (1986). *Facts, artifacts, and counterfacts*. Boynton/Cook.

Bastug, M., Ertem, I. S., & Keskin, H. K. (2017). A phenomenological research study on writer's block: Causes, processes, and results. *Education+ Training*, *59*(6), 605-618. https://doi.org/10.1108/ET-11-2016-0169

Boice, R. (1993). Writing blocks and tacit knowledge. *The Journal of Higher Education*, *64*(1), 19-54. https://doi.org/10.1080/00221546.1993.11778407

Bourgonje, P., & Tromp, R. (2011). *Quality educators: An international study of teacher competencies and standards*. Oxfam Novib.

Boyer, W. (2010). Empathy development in teacher candidates. *Early Childhood Education Journal, 38*, 313-321. https://doi.org/10.1007/s10643-010-0419-8

Cambridge English Assessment. (2024). *Assessing writing for Cambridge English qualifications: A guide for teachers - B1 Preliminary for schools.* Cambridge University Press & Assessment. Retrieved from https://www.cambridgeenglish.org/images/231794-cambridge-english-assessing-writing-performance-at-level-b1.pdf

Cheng, Y. s. (2002). Factors associated with foreign language writing anxiety. *Foreign Language Annals*, *35*(6), 647-656. https://doi.org/10.1111/j.1944-9720.2002.tb01903.x

Choi, S. (2013). Language anxiety in second language writing: Is it really a stumbling block? *Second Language Studies*, *31*(2), 1-42. Retrieved from http://hdl.handle.net/10125/40722

Cohen, M. R. (1998). *The pause that represses: The experience of working through writer's block during the dissertation writing process*. Massachusetts School of Professional Psychology.

Denbo, S., & Williams, B. (2016). All Students Are Not Equal. In P. M. Flynn, K. Haynes & M. A. Kilgour (Eds.). *Overcoming challenges to gender equality in the workplace: Leadership and innovation* (pp. 148-150). Routledge.

Deng, L., Wu, Y., Chen, X., & Peng, Z. (2024). Pursuing competencies or pursuing scores? High school teachers' perceptions and practices of competency-based education reform in China. *Teaching and Teacher Education, 141*(1), 104510. https://doi.org/10.1016/j.tate.2024.104510

Dewberry, C., Davies-Muir, A., & Newell, S. (2013). Impact and causes of rater severity/leniency in appraisals without postevaluation communication between raters and ratees. *International Journal of Selection and Assessment, 21(*3), 286-293. https://doi.org/10.1111/ijsa.12038

Elbow, P. (1973). *Writing without Teachers*, pp. 12–75. Oxford University Press.

Elif, G., & Yayli, D. (2019). The second language writing anxiety: The perceived sources and consequences. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, *45*(45), 235-251. Retrieved from https://creativecommons.org/licenses/by-nc-nd/4.0/

Elliot, N. (2016). A theory of ethics for writing assessment. *Journal of Writing Assessment*, *9*(1). Retrieved from https://escholarship.org/uc/item/36t565mm

Elshawa, N. R., Abdullah, A. N., & Rashid, S. M. (2017). Malaysian instructors' assessment beliefs in tertiary ESL classrooms. *International Journal of Education and Literacy Studies*, *5*(2), 29-46. https://doi.org/10.7575/aiac.ijels.v.5n.2p.29

Ernawati, H., Hidayat, H., Primandari, S. R. P., Ferdian, F., & Fitria, R. (2022). The empirical study of factors affecting students' competence of fashion design education in the industrial revolution 4.0 era. *International Journal of Instruction*, *15*(4), 259–276. https://doi.org/10.29333/iji.2022.15415a

Fulcher, G. (2013). *Practical language testing*. Routledge.

Gage, N., & Berliner, D. (1991). *Educational psychology* (5th ed.). Houghton, Mifflin.

Gao, R., He, T., Liao, Y., Liu, X., Fan, Y., Su, Y., Zuo, H., & Mo. L. (2022). An investigation on the academic burden of Chinese students ranging from primary schools to universities based on a Word Association Test in Guangdong Province. *International Journal of Environmental Research and Public Health, 19*(4), 2481. https://doi.org/10.3390/ijerph19042481

Ghanizadeh, A., & Moafian, F. (2010). The role of EFL teachers' emotional intelligence in their success. *ELT Journal*, *64*(4), 424-435. https://doi.org/10.1093/elt/ccp084

Goleman, D. (1996). *Emotional intelligence: Why it can matter more than IQ*. Bloomsbury Publishing.

Hairston, M. (1982). The winds of change: Thomas Kuhn and the revolution in the teaching of writing. *College Composition and Communication*, *33*(1), 76-88. https://doi.org/10.2307/357846

Hassan, B. A. (2001). The relationship of writing apprehension and self-esteem to the writing quality and quantity of efl university students. *Mansoura Faculty of Education Journal*, *39*, 1-36. https://doi.org/10.32038/ltrq.2020.18.01

Hosseini, H. M., Derakhshesh, A., & Mehraein, S. (2025). Effects of teacher response on learners' writing development, ideal L2 writing self, and writing engagement. *International Journal of Instruction*, *18*(2), 615-636. https://doi.org/10.29333/iji.2025.18233a

Jacobs, H. L. (1981). *Testing ESL composition: A practical approach. English Composition Program*. ERIC.

Kigwilu, P. C., & Mokoro, D. K. (2022). Teachers' assessment practices in implementing competence-based curriculum in secondary schools in Arumeru District of Tanzania. *East African Journal of Education and Social Sciences*, *5*(2), 792. https://doi.org/10.37284/eajes.5.2.792

Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. J. Weir (Eds.). *European language testing in a global context* (27-48). Cambridge University Press.

Kunnan, A. J. (2014). Fairness and justice in language assessment: Principles and public reasoning. *Alternative Pedagogies In The English Language and Communication Classroom*, *36*. Retrieved from https://bpb-us-w2.wpmucdn.com/blog.nus.edu.sg/dist/a/15400/files/2021/12/Alternative-pedagogies-in-the-English-language-Communication-classroom-4th-CELC-Symposium_2014.pdf#page=41

Kunnan, A. J. (2019). An agenda for language assessment development, research and pedagogy. *Journal of Asia TEFL*, *16*(1), 327. Retrieved from https://scholarworks.gvsu.edu/cgi/viewcontent.cgi?article=1285&context=mrj

Lam, T. C. M., Kolomitro, K., & Alamparambil, F. C. (2011). Empathy training: Methods, evaluation practices, and validity. *Journal of Multidisciplinary Evaluation*, *7*, 162–200. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/314

Lim, C. (1978). *Little Ironies: Stories of Singapore*. Heinemann.

Masuku, M. M., Jili, N. N., & Sabela, P. T. (2021). Assessment as a pedagogy and measuring tool in promoting deep learning in institutions of higher learning. *International Journal of Higher Education, 10*(2), 274-283. https://doi.org/10.5430/ijhe.v10n2p274

Mohd Salleh, N. F. A., Nimehchisalem, V., Jalaluddin, I., & Mukundan, J. (2023). Washback effects of School-Based Assessment (SBA) on Malaysian secondary school students' English language learning. *PASAA: Journal of Language Teaching and Learning in Thailand, 66*, 168-201.

Mukundan, J. (2011). Developed world influences on ESL/EFL writing situations: Differentiating realities from fantasies. In P. L. Ha & B. Baurain (Eds.). *Voices, identities, negotiations, and conflicts writing academic English across cultures*. Emerald Group Publishing.

Mukundan, J., Zare, P., Zarifi, A., Manaf, U. K. A., & Sahamid, H. (2015). Language Teacher Burnout and School Type. *English Language Teaching*, *8*(9), 26-32. DOI:10.5539/elt.v8n9p26

Murray, D. M. (2004). *A writer teaches writing* (revised second edition). Cengage Learning.

Nair, S. D., & Nesamany, S. S. (2021). Teacher Perception of Online Teaching Approaches During the Pandemic Covid-19. *Asian Journal of Social Science Research*, *3*(2), 20-38. doi:10.5281/zenodo.8063275.

Nguyen, C. (2015). Training high qualified teachers in Vietnam: Challenges and policy issues. University of Education.

Nimehchisalem, V. (2024). Differentiated assessment in a postgraduate language assessment course. In S. Xia & C. A. Hafner (Eds.). *Asia-Pacific LSP & Professional Communication Association News*, *6*, 16-19.

Nimehchisalem, V., & Nur Izyan Syamimi Binti Mat Hussin (2018). Postgraduate students' conception of language assessment, *Language Testing in Asia, 8*(11), 1-14. https://doi.org/10.1186/s40468-018-0066-3

Nimehchisalem, V., Hosseini, M., Cortazzi, M., & Jin, L. (2023). Multiple perspectives of stakeholders towards young learners' language assessment in an international school in Malaysia. *Language Teaching Research*, 13621688231154440. https://doi.org/10.1177/13621688231154440

Nimehchisalem, V., Kai, E. F. S., & Nowrouzi, S. (2019). English as a Second Language Learners and Teachers' Conceptions of Language Assessment. *Journal of Asia TEFL*, *16*(4), 1348. http://dx.doi.org/10.18823/asiatefl.2019.16.4.18.1348

Rashtchi, M., & Beiki, M. (2015). The effect of teacher-generated cooperative brainstorming versus learner-generated brainstorming on activating EFL learners' background knowledge in essay writing classes. *Indian Journal of Fundamental and Applied Life Sciences, 5*(S2), 1218-1227.

Rashtchi, M., & Porkar, R. (2020). Brainstorming revisited: Does technology facilitate argumentative essay writing? *Language Teaching Research Quarterly, 18*, 40-52. Retrieved from: https://eric.ed.gov/?id=EJ1270008

Rashtchi, M., Porkar, R., & Ghazi Mir Saeed, F. (2019). Product-based, process-based, and genre-based instructions in expository writing: Focusing on EFL learners' performance and strategy use. *European Journal of Education Studies, 6*(6), 115-136. http://dx.doi.org/10.5281/zenodo.3427842

Ratminingsih, N. M., Marhaeni, A. A. I. N., & Vigayanti, L. P. D. (2018). Self-assessment: The effect on students' independence and writing competence. *International Journal of Instruction*, *11*(3), 277–290. https://doi.org/10.12973/iji.2018.11320a

Reid, J. (1993). *Teaching ESL writing*. Prentice-Hall.

Rezai, A. (2022). Fairness in classroom assessment: Development and validation of a questionnaire. *Language Testing in Asia, 12*(1). https://doi.org/10.1186/s40468-022-00162-9

Richmond, K. J. (1999). The ethics of empathy: Making connections in the writing classroom. *The Journal of the Assembly for Expanded Perspectives on Learning*, *5*(6), 37-46. https://doi.org/10.7290/jaepl56foy

Roell, C. (2019). Using a case study in the EFL classroom. *English Teaching Forum, 57*(4), 24-33.

Rogers, C. R., & Freiberg, H. (1970). *Freedom to learn.*Charles Merrill.

Shulman, J. (1992). *Case methods in teacher education*. Teachers College Press.

Spolsky, B. (1995). *Measured words: The development of objective language testing.* Oxford University Press.

Sukumaran, R. D., Nasir, F. D. M., Rahman, N. I. A., Che Ani, S., & Zuki, N. H. M. (2021). Students' motivation towards English language learning during the pandemic COVID-19 in Malaysia. *Asian Journal of Social Science Research*, *3*(2), 1-10. doi:10.5281/zenodo.8063215.

Thi Thuy Oanh, T., & Pham, V. P. H. (2023). The effects of online peer feedback on students' writing skills during corona virus pandemic. *International Journal of Instruction*, *16*(1), 881-896. https://doi.org/10.29333/iji.2023.16149a

Ung, E. (2003). Youth suicide and parasuicide in Singapore. *Annals of the Academy of Medicine, Singapore*, *32*(1), 12-18.

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Walker, D. (2017). Empowering English writing students: Reflecting on aspects of the process that helped me most? *LLT Journal: A Journal on Language and Language Teaching*, *20*(2), 85-93. https://doi.org/10.24071/llt.v20i2.735

Wallace, M. P., & Qin, C. Y. (2021). Language classroom assessment fairness: Perceptions from students. *LEARN Journal: Language Education and Acquisition Research Network*, *14*(1), 492-521. Retrieved from https://so04.tci-thaijo.org/index.php/LEARN/article/view/248700

White, W. F. (1999). What every teacher should know about the functions of emotions in children and adolescents. *Education, 119*(1), 120-125.

Wong, H. (1989). *The development of a qualitative writing scale*. Penerbit Universiti Kebangsaan Malaysia.

Appendix A

Summaries of the Two Stories (Case Stories)

Text 1*: Adeline Ng Ai Choo*

The protagonist, Adeline Ng Ai Choo, sees her teacher after she fails the end-of-year English language test. She asks for extra passing marks, but the teacher dismisses her request. Adeline takes her own life as a result of this. Most of the story describes the debate in the staff meeting where teachers debate the issues that have emerged, all related to assessment, like whether marks should be awarded to borderline cases or whether scripts after evaluation should be returned to students. All the arguments in the staff room concerned teachers trying to shift the blame away from themselves or figuring out excuses to justify their actions and those of the school. Her class teacher claimed that Adeline was unstable, although she had been excellent in Secondary One and Two, but seemed ignorant of the fact that her inaction over two years probably was a big factor to consider in Adeline's death. Adeline's parents, when meeting the teacher, claimed that she had the habit of constantly scribbling in her diary, something that they were against. Ironically, the scribblings in her diary were what writing experts recommend (freewriting) to learners seeking fluency in writing. The parents did not even allow her to watch television. She only studied. Upon her death, they found a note addressed to her parents, apologizing for being a disappointment. Her diary entries showed her mental state – she was terribly disturbed in her mind.

Text 2: *The Teacher*

Tan Geok Peng, the protagonist, a Secondary Four student, commits suicide as she is unable to cope with her studies. Her father wanted her to work rather than study to supplement the household income. Eventually, her mother convinces her father to allow her to continue her studies. Much of the story focuses on staff room scenes where Tan Geok Peng's teacher is discussing the language errors in the girl's essays. At the end of the story, however, the reader is made aware that Tan Geok Peng is indeed a clever student. She had language problems in her compositions, but she never lacked ideas. She had written "The Stranger," but the teacher failed her as he thought she had gone completely out of point by describing the stranger as her own father. The teacher had missed the clever manipulation of the title – the metaphorical description of her father as a stranger. The story's writer made a fool out of Tan Geok Peng's teacher, who had no clue about how potent metaphors are.

Appendix B

Interactions on Assessment Feedback and Teaching Approaches at Group Interviews

Bonnie: I believe the only thing the teacher did was to provide negative written feedback on Tan Geok Peng's essay. This is also what teachers in China do.

Karen: I believe this is also a type of teaching. Providing written feedback in assessment can also be considered a form of teaching.

Moderator: But is this "form of teaching, giving written feedback" is it adequate? Especially when it concerns an introverted and timid learner like Tan Geok Peng.

Jan: But this was never supposed to be about the Process Approach, which it is almost non-existent in the Chinese context. The traditional writing class is about this – written feedback from the teacher, and if at all learners go and see the teacher after getting their exercise books, it would be to dispute some aspect of feedback or claim for more marks.

Moderator: But PF1 says, and I repeat, that an assessment ought to provide an adequate opportunity to learn the knowledge or abilities that are to be assessed. Was this truly accomplished?

Suzy: I'm afraid not. The Chinese teacher is not at all prepared for this. While teacher editing of language in learners' exercise books is common, that is the language aspect, but I personally believe that is the most important thing in a composition. Examiners, when they see too many errors, usually have the impression that it is a bad essay and will be influenced very early on to give it a fail.

Moderator: Does everybody agree? I have seen so many writing evaluation instruments where marks for content are almost the same as language. Sometimes, content has greater weightage.

Bonnie: That's true. There was no teaching for the content section.

Moderator: You mean idea generation, etc.?

Suzy: yes, if at all, teachers occasionally taught learners how to outline by getting the whole class to contribute points (ideas) for the essay.

Moderator: So, you do believe that PF1 was never fulfilled?

(Everyone agrees)

Appendix C

Interactions on Recognising and Supporting At-Risk Students at Group Interviews

(Karen, reading an extract from the story): Perhaps we shouldn't blame the school or examination system. As Ai Choo's class teacher, I must say that she was a very unstable girl – she used to be good and stable in Secondary 1 and 2, but since last year, she suddenly changed. I spoke to her a few times and asked her what the matter with her was, and why her work was going down, but she was very evasive and didn't want to tell me anything, so I had to leave it at that. (*Adeline Ng Ai Choo*, p. 65)

She is obviously an "at risk" learner and she never got the attention she needed.

Moderator: I have known of some learners who were going through depression. Their condition already makes them very insecure – they shun company, they keep to themselves, etc. Do you think teachers need to be aware of these conditions (because the teacher in Adeline Ng Ai Choo seems to know there is a downward slide but does nothing about it)?

Karen: The system failed Adeline because the teacher had very little knowledge about these human conditions (like depression), which hinder learning.

Bonnie: Of course, this is made worse when the teacher also has very little knowledge of learner-centered writing strategies. Adeline kept a diary in which she wrote a lot (although her parents disagreed with her writing). The teacher could have sanctioned the use of diaries by getting them to write in official exercise books and called them journals. This would have supported Adeline's writing.

Appendix D

Interactions on Fairness and Bias in Language Assessment at Group Interviews

Bonnie: But not everything about the Chinese education is bad. I believe like most writing teachers in China believe, that classroom assessments and homework tasks are practice sessions for the same thing. They are meant to help the learner-writer to prepare for exams as well. They are meant to provide practice.

Moderator: But how about the "supportive" features (gestures with fingers as he expresses) like peer reviews, idea-generation strategies (collaborative learning in class), and teacher conferences

Bonnie: But teachers support their learners by providing written feedback (comments in the students' exercise books).

Moderator: It's never the same, is it? What if the teacher had face-to-face conferences with a student like Tan Geok Peng (in the story The Teacher)? Wouldn't she have told her the other problems as well?

Bonnie: That's the only thing Chinese Writing Teachers can do. They don't have the time and energy to do most other things carried out in Western-based Process Approaches. They teach large classes. It's impossible.

Moderator: Do you believe the tragedy would have been avoided if there had been one-to-one conferences?

Cindy: Yes, I believe so. But even in large Malaysian classes, the Process Approach is almost impossible to implement.