

RESEARCH

Open Access



Postgraduate students' conception of language assessment

Vahid Nimehchisalem*  and Nur Izyan Syamimi Mat Hussin

* Correspondence: nimechie@gmail.com;
vahid@upm.edu.my
Universiti Putra Malaysia, Serdang,
Selangor, Malaysia

Abstract

Background: Assessment is an essential part of any learning process; however, it is often disliked by most students. A wealth of research is available on issues on assessment, such as washback effect and test anxiety, and yet we know little about what students conceive of assessment. The objective of this case study was to explore a group of postgraduate students' conception of language assessment.

Methods: The participants of this qualitative study were an intact group of postgraduate students ($n = 21$) who completed two tasks. The first task was creating a metaphor by completing the sentence "A world without assessment is..." and then explaining the metaphor. To fulfil the second task, the participants wrote a short story each related to language assessment. The collected data were analyzed thematically.

Results: Interesting metaphors emerged. The metaphors covered four purposes of assessment, including "to guide," "to motivate," "to empower," and "to control" with frequencies of 9, 6, 3, and 3, respectively. For example, assessment was viewed as a map (guide), reward (motivation), driver's license (empowerment), and iron fist (control). In the short stories, the most frequent themes that recurred were "cheating," "test anxiety," and "motivation tool" with frequencies of four each. The themes which were the least frequent were "fair assessment," "promoting learner collaboration," "teaching to the test," and "biased assessment" each occurring only once.

Conclusion: The results show that collectively students are well aware of the purposes of assessment but their experiences with assessment illustrate cases of language assessment which indicate room for improvement in the way assessment is understood and administered.

Keywords: Language assessment, Assessment literacy, Metaphors on assessment, Short stories on assessment

Introduction

Assessment is an inseparable part of English language learning-teaching. It is defined as "any act of interpreting information about student performance, collected through any of a multitude of means" (Brown and Hirschfeld 2008, p. 4). Language educators need to depend on assessment for several reasons. Most importantly, they should assess their learners to check their learning uptake and to revise their teaching methods or materials if necessary. Therefore, as an inseparable component of the learning-teaching process, assessment should help the learner and the teacher to ensure that learning is actually taking place. Assessment can also be used for placement purposes. At certain stages, educators

may need to make decisions about students' language proficiency levels. Sometimes assessment is also necessary to diagnose students' particular difficulty areas of language learning. While some forms of assessment are high-stakes and important decisions are made based on their results, other types of assessment may be low-stakes and their results may not have significant consequences in the test-taker's academic life. Whatever the purpose of assessment, it plays an important role throughout the learning-teaching process. By examining how assessment is practiced in an educational setting, it is possible to evaluate the quality of the entire system. In a low-quality educational system, assessment is never paid its due respect and attention. For most of the learners in such a system, assessment typically connotes scores and tests. Some may even remember pain, fear, punishment, and competition when thinking of the term "assessment." The teachers in such a system are reminded to assess their learners only at the end of courses and may practice it only when they have to.

There is empirical research evidence that indicates for most teachers assessment still means tests given to students and scores that determine their passing or failing courses, as a means of disciplining or threatening them (López 2008). Research findings also indicate that students are not well-trained in terms of assessment knowledge and skills. For example, in Colombia, López and Bernal (2009) reported that only about 25% of the undergraduate programs in their study offer a course on assessment. Another example could be offered from Iran where the managerial, institutional, and professional barriers seem to be standing in the way of the reforms in the new English curriculum and language assessment (Razavipour and Rezagah 2018). In the context of Malaysia, the educational system has been undergoing a drastic change in the past few years. With regard to assessment, currently the focus has turned from students' summative evaluation to their ongoing formative assessment (Ministry of Education, Malaysia 2013). Therefore, centralized exams are gradually being replaced by localized, "school-based assessment." The shifts in educational policies in the country are in line with the globally accepted paradigms, but the question that remains to be answered is whether the Malaysian community is ready for such a shift. Previous literature is available from the local context of Malaysia regarding English language teachers, instructors, and lecturers' points of view. The importance of teachers' assessment skills has been recognized especially in school-based assessment; however, teachers are uncertain about their role in school-based assessment (Majid 2011).

Theoretical background

Constructivists see a learner as a member of a community of learners that construct knowledge based on their interaction with their environment. Constructivism supported the idea of learner-centered classrooms where learners constructed knowledge rather than receiving it passively, where learners analyze and think rather than memorize and where learners actively understand and apply rather than repeat back (Marlowe and Page 2005). From the viewpoint of social interactionism, learning is an ongoing process in which the individual interacts with the environment (Vygotsky 1978). Social interactionists developed the idea of the zone of proximal development (ZPD) which differentiates between (i) what learners can do with no aid and (ii) what they cannot do yet. From such a perspective, learners will find themselves in one of the following three conditions in the process of learning:

1. Comfort zone: There will always be conditions in which learners can handle it independently with no assistance.
2. Zone of proximal development: In such conditions, individuals can learn only with guidance which often happens in the form of meaningful interaction with the teacher or other learners who have already mastered the skill
3. Danger zone: This encompasses what is not yet learnable even with assistance.

This means that learning will always proceed as learners as well as their teachers, peers, and parents continuously assess what has already been learned, what can be learned, and what will have to wait to be learned. Therefore, in a learner-centered classroom, assessment is a constant process. The teacher makes every attempt to move beyond summative assessment, which only focuses on learners' achievement after experiencing some instruction. The focus is shifted to the process of learning whereby assessment occurs continuously and formatively while the learner is still learning (Stiggins 2002). Thus, assessment is there *for* learning as the students are in the process of learning. Admittedly assessment *of* learning will always remain an essential component of learning as it can show whether the learning objectives have been achieved. However, it should not be the only form of assessment that is practiced.

Significance of this study

In a student-centered language classroom, learners are supposed to participate actively in the process of learning-teaching-assessment. Learners frequently assess their own performance and progress with or without self-assessment checklists. They are also expected to assess their peers' works as well as their teachers' performance. Unlike the traditional teacher-centered classroom, the learner has to be a capable assessor. Thus, the way students understand assessment influences their quality of learning. If learners carry misconceptions about assessment, their learning and their peers' learning will be at stake.

Depending on how it is practiced and therefore what the learner conceives it to be, assessment can result in both adaptive and maladaptive learner behavior. It can promote adaptive learning behavior, when learners believe that it provides useful information on their learning needs and progress; however, it may cause maladaptive behavior if learners fail to see its relevance to their learning and development (Brown et al. 2014). From this viewpoint, it seems logical to argue how learners behave before, during, and after assessment depends on how they conceive of it (Segers, Dochy, & Cascallar, 2006).

Therefore, it is crucial to understand students' conception of assessment. Research is needed to investigate what assessment means to learners. The outcome of such studies can provide invaluable information based on which more effective activities and teaching plans can be developed. Hence, this study is significant as its findings will provide useful implication for teachers, schools, parents, and education policymakers.

Objective and research questions

The present study aims to explore a group of postgraduate students' conception of language assessment. More specifically, the following research questions are addressed:

1. What do the postgraduate students conceive to be the purpose of language assessment based on the analysis of their metaphors?

2. What major themes emerge as a result of analyzing the short stories written by the postgraduate students about language assessment?

Past studies

In the related literature, numerous studies are available which focus on teachers' conceptions of assessment (Brown 2004, 2008; Fives and Buehl 2012). Several studies indicated that students' conceptions of assessment affect the quality of learning (Entwistle and Entwistle 1991; Marton and Säljö 1997; Ramsden 1997) underlining the significance of learners' conceptions of assessment. Studies have shown that teachers' conceptions influence the assessment outcomes and practices (Barnes et al. 2015; Fulmer et al. 2015).

In New Zealand, using the structural equation modeling (SEM), Brown and Hirschfeld (2008) studied secondary school students' ($n = 3469$) conceptions of assessment. The source of data of this large-scale study was the students' scores from a standard reading comprehension test and their responses to a self-report inventory. Four conceptions were identified, including (i) assessment makes schools accountable, (ii) assessment makes students accountable, (iii) assessment is ignored, and (iv) assessment is fun. The results indicated that lower reading achievement scores were predicted for non-majority ethnicity students, male students, those who conceived of assessment as a measure of school accountability, those who ignored assessment, and those who conceived of assessment as fun. On the other hand, it was reported that there is a positive relationship between the students' conception that "assessment is about grading students" and their achievement scores (Brown and Hirschfeld 2008).

In another study, Brown et al. (2014) examined 760 Iranian university students' conceptions of assessment, using confirmatory factor analysis (CFA). They adapted an instrument, called the students' conceptions of assessment (SCoA) inventory, which addresses the four domains of improvement (assessment resulting in learning), affect (social and emotional influences of assessment), external attributions (factors beyond the control of the learner), and irrelevance (inappropriate assessment interfering with learner development). The results showed that students had both positive and negative conceptions of assessment. They regarded it as what can both improve learning-teaching and impede learner development.

In their investigation of the impact of the current language assessment reform (LAR) on English teachers' assessment practices in Iran, Razavipour and Rezagah (2018) conducted a focus group discussion with four teachers and also examined several teacher-made tests. Their results indicated that (i) teachers have not taken ownership of the reform, (ii) schools lack adequate resources, (iii) accountability demands have resulted in grade inflation, and (iv) teachers tend to be unable to assess consistently with the LAR demands. More specifically, Razavipour and Rezagah (2018) found that teachers appeared to be unclear about what and why they assessed; the teachers also found it difficult to align their assessment practices with principles of communicative language teaching method. These findings indicate the need for developing language teachers' language assessment literacy to ensure that they comprehend the idea of language assessment reform and adjust the way they assess.

Hakim (2015) investigated the assessment literacy of a group of English language instructors ($n = 30$) at a university language center in Saudi Arabia. Her results revealed

that all the participants had clear understanding of the use of assessment tools. Another noteworthy finding was that the instructors' understanding of assessment tools was little influenced by their teaching experience. However, although these instructors were well aware of the assessment techniques, poor practicing techniques were reported among them.

In Hong Kong, Tong and Adamson (2015) conducted a survey on 423 and semi-structured interviews on 45 secondary school students' perceptions of school-based assessment. They found that most of the students provided negative conceptions of the assessment and were dissatisfied with the insufficient feedback from their teacher. However, they agreed that feedback helped their learning.

In the Malaysian context, past studies have commonly focused on language teachers' assessment beliefs, perceptions, and skills. These studies indicate that Malaysian English language teachers generally lack the necessary assessment skills. To offer an example, in a survey of Malaysian secondary school teachers ($n = 73$), Zainuriyah (2012) reported that the teachers were incompetent in all aspects of language assessment literacy. In another study, Malaysian secondary school teachers' assessment literacy was found to be inadequate, especially in reporting results (Mohamed, Kamis, & Ali, 2016). A survey on the assessment practices of Malaysian teachers also showed that they still prefer traditional types of assessment (Suah and Ong 2012). Studies in the context of Malaysian universities seem to show slightly different outcomes. English language instructors in Malaysia indicated relatively high levels of language assessment beliefs (Rezvani Kalajahi and Ain Nadzimah 2016). Research has also shown that Malaysian English language instructors view assessment as a vital tool to improve students' learning and instructors' teaching (Niveen et al. 2017).

As this brief review of literature illustrates, studies on language learners' conceptions of assessment are lacking in Malaysia. Furthermore, most of the previous studies followed quantitative methods. More qualitative research in the area may be able to contribute new insights to the area.

Methods

In this case study, qualitative methods were followed to collect and analyze the data. In this section, we describe the participants as well as the data collection and analysis methods.

Participants

An intact group of postgraduate students ($n = 21$) in the area of Applied Linguistics from a public university in Malaysia was selected as the participants of this study. Most of them were females ($n = 15$). They came from different countries, including Malaysia, Nigeria, Iraq, Syria, Libya, and Turkey. These students had enrolled in a master's degree course in the area of language testing. About one third of the students were language teachers who were furthering their MA on a part-time basis. Most of the participants who were not teaching currently had been language teachers before they moved to Malaysia to further their studies.

Data collection

The data were collected from an intact group of MA students who had enrolled in a course on language assessment. The participants completed the first task in the first

week. The task prompted the participants to complete the sentence “A world without assessment is...” by creating a metaphor. Then, they were requested to explain the logic behind the metaphors they had created (Appendix 1). The participants were given a week to complete this task. In the related literature, metaphors have been analyzed by previous researchers to explore what individuals conceive of a phenomenon. In a study of Chinese learners’ expectations of teachers, Cortazzi et al. (2009) analyzed about 3000 metaphors which they collected from 496 university students. In another study that sought to describe the external and internal landscape of education in China, Jin and Cortazzi (2011) collected and analyzed a number of related metaphors. In the present study, the metaphor data had been collected and analyzed before collecting the data of the second task.

The second task involved writing short stories which presented the participants’ personal experiences of language assessment (Appendix 2). The participants were given 4 weeks to complete this task. They read Cathrine Lim’s *The Teacher*, as a sample short story about English language assessment. Then they created the plot of their short story, and based on the feedback they received from their peers, they drafted the first version of their work. Revising their initial drafts based on their self-assessment and peer feedback would result in the second and third drafts of the short stories, respectively. It was this third draft which was used for analysis. In the related literature, collecting and analyzing learners’ narratives is an established method. Richmond (2002) provides an illustrative example by analyzing an adult learner’s story to establish the significance of stories in eliciting rich and in-depth data. Pavlenko (2007) discusses the strengths and weaknesses of narrative methods (such as language memoirs, linguistic autobiographies, and learners’ journals and diaries) in qualitative data collection. Finally, the collected data consisted of 21 samples of metaphors and short stories written by the participants.

Data analysis

The data were analyzed qualitatively via thematic analysis. To analyze the metaphors and to address the first research question, deductive coding method was followed. The metaphors were analyzed based on Fulcher (2010), who outlines four broad purposes for language tests:

1. To control: assessment as a gatekeeping tool
2. To empower: vehicle by which society can implement equality of opportunity or learner empowerment
3. To motivate: to attract the interest of a pupil and directing it into the desired channels
4. To guide: a tool directing us to make decisions
 - i. To select: to render information to aid in making intelligent decisions about possible courses of action
 - ii. To place: to place learners into classes
 - iii. To evaluate achievement: to discover how much they have achieved
 - iv. To diagnose: to diagnose difficulties that individual learners have

Tests can be used for four interrelated purposes of gatekeeping, empowering, motivating, and guiding. A very common example for tests as gatekeeping tools would be the

International English Language Testing System (IELTS) test which has to be taken by individuals interested in emigrating to English-speaking countries. In addition to their purpose to help institutions to control their policies, employment, intake, etc., tests may also be used to empower or create equity in the community. For example, to avoid institutes of higher education to be monopolized by only the high-income class, unlike a few centuries ago, tests are given to the whole population and only those who perform better are given the opportunity to further their studies. Tests also motivate students to overlearn and internalize the material. Finally, tests can also guide stakeholders, which may happen in four different ways. First, tests can provide useful information which enables decision makers to select the right individuals for the suitable positions. Second, the right individuals can be placed in classes with different levels based on their performance in placement tests. Third, tests can also show how much the learner has achieved after some treatment. Finally, tests can also help teachers diagnose their learners' specific needs and areas of difficulty.

This information provided the present researchers with a useful analytical framework according to which the metaphors written by the participants could be analyzed and classified. Furthermore, this study also aimed at investigating the participants' experiences with language assessment by analyzing their short stories. Inductive coding method was followed for this purpose. The researchers read the stories and extracted their themes. The emerging themes were listed as the participants' experiences. Then, the stories with similar themes were identified and the results indicated the frequency of occurrence of each theme across all the short stories.

Results

In this section, the results of the study are presented in the order of the two research questions.

Metaphors and purpose of language assessment

The metaphors were analyzed based on Fulcher (2010). Table 1 presents the purposes of assessment indicated by the metaphors:

As can be seen in Table 1, a large proportion of the metaphors (9 in 21) regarded assessment as a guide. Assessment was introduced as a "guiding star" that gives direction and meaning to learning. For one of the participants, assessment was the "steps of a ladder." Such a metaphor rightfully emphasizes the fundamental role of assessment (steps) in education (ladder) and it also views assessment as a gradual and ongoing process. Another interesting metaphor was "life of a free bird but forever a wanderer," which emphasized that world without assessment may mean freedom for a bird (learner) that wishes to fly freely, and yet the problem is that without assessment, this bird does not know where to fly since it has no sense of directions.

The next assessment purpose which recurred in the metaphors was "to motivate." Based on the results, there were six metaphors which viewed assessment as a reward that reinforces and motivates learning. These metaphors regarded assessment as the taste or salt without which food would be bland. The two assessment purposes with the lowest occurrences were "to empower" and "to control" with relatively low frequencies of three each. The metaphors which presented assessment as a means of control

Table 1 Metaphors and purposes of assessment

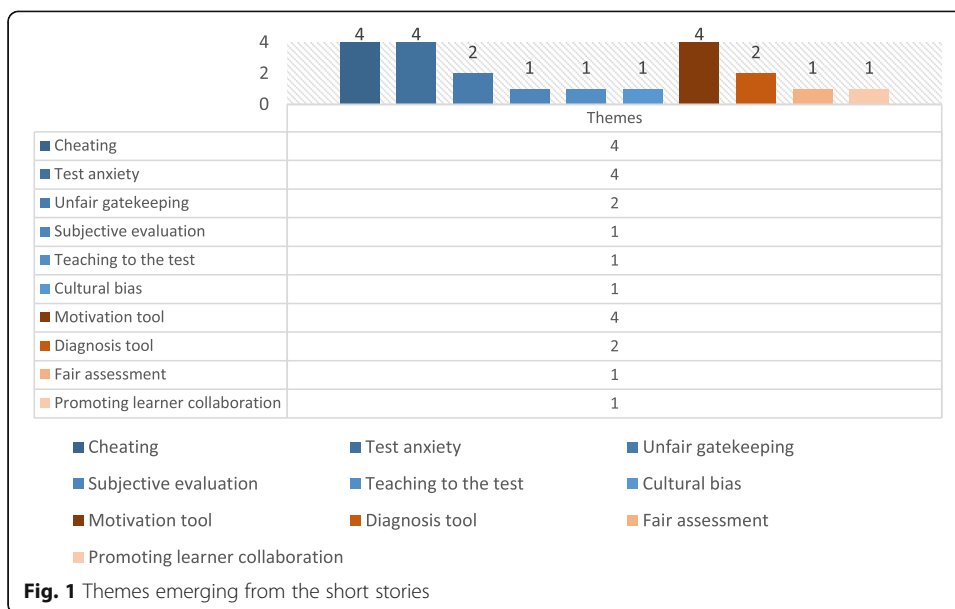
Purpose	Metaphor	Frequency
To guide	1. A life without sun	9
	2. A ladder without the steps	
	3. Life of a free bird but forever a wanderer	
	4. A road without directions	
	5. A map without a route	
	6. A human without brain	
	7. Vision problem without eyeglasses	
	8. A life with no meaning	
	9. A journey with no ends	
To motivate	1. Food without salt	6
	2. Work with no reward	
	3. A food without taste	
	4. Chocolate without taste	
	5. A job with no salary	
	6. A world with rest (grin emoticon)	
To empower	1. Driving without license	3
	2. A world without improvement	
	3. Adventure without risk	
To control	1. Doomsday without hell	3
	2. A state ruled with no iron fist	
	3. A lion with no cuspid	

likened it to “hell,” “iron fist,” and “lions’ cuspid” without which God, politicians, and lions (or teachers) would lose power and would fail to control (their learners). Assessment as a source of empowerment recurred only three times in the metaphors; one participant wrote world without assessment would be “a world without improvement.”

Experiences of language assessment

The short stories written by the participants were read by the researchers. The repeating ideas that emerged were categorized into themes that would indicate the experiences that the participants had of language assessment. Figure 1 presents the themes that emerged from the short stories:

Two different color themes have been used in the figure to differentiate the positive from negative themes. As shown in the figure, negative themes, such as “cheating” and “test anxiety,” occurred twice as much as positive themes, such as “motivation tool.” The most common themes which emerged in the short stories were “cheating,” “test anxiety,” and “motivation tool,” each occurring four times. What recurred in most of the stories was unfair assessment practices which urged students to cheat or created deliberative anxiety in them. One of the stories was about a good student who is taking a centralized exam. Even though she leaves early and takes a taxi, she arrives at the exam venue too late because the taxi driver is lost and his car breaks down. The story rightfully points out the drawback of centralized and summative assessment which can cause a great deal of anxiety and pain in the learner. The most common (4 out of 21) positive theme in the stories was related to tests spurring students to study. Although



the positive theme of assessment as a motivating tool emerged from these four stories, motivation was often created in a painful and coercive manner; therefore, assessment was presented as a necessary evil in these stories which had to be practiced to discipline the student and/or to ensure that learning had taken place.

The second most frequently recurring themes were “unfair gatekeeping” and “diagnosis tool,” each of which occurred in two of the short stories. One of the stories was about two friends one of whom comes from a minority group and fails to achieve her desired mark because of unfair assessment practices and prejudice against minority groups. She fails a test and cannot have her dream job because of unfair discriminations in her community. On the other hand, assessment as a tool for diagnosing students’ specific problems also recurred in two of the stories. In one story, a student finally understands why he cannot achieve better marks in English writing when his lecturer provides specific feedback on his writing performance and he notices how he can improve. In the second story, a group of students have problems in oral presentations but with the help of teacher’s assessment and support they learn how to present more effectively.

Other themes which emerged each only in a single story were “subjective evaluation,” “teaching to the test,” “cultural bias,” “fair assessment,” and “promoting learner collaboration.” One of the short stories was about a student whose essay is marked unfairly by her writing teacher who seems not to appreciate her style; the problem is resolved as an independent rater marks the script but only after the student and her family have gone through a lot of pain and anxiety. Another story relates the case of a teacher who prepares his students for a centralized test and teaches to the test; his students perform badly on the test as they have not learned and they rely only on their teacher’s shortcuts. There was another story in which a native speaking teacher is unaware of his culturally biased test content which causes serious problems for his Muslim students. In the only story which represented fair assessment practices, a novice teacher who is also a novice test designer constructs a test with vague items and inappropriate level of

difficulty but the school principal finds out and provides timely support. Finally, in another story, students learn how they can learn and perform better on tests when they collaborate and learn together.

Discussion

There is empirical research evidence that the way learners conceive assessment determines their final achievement (Brown and Hirschfeld 2008). Thus, it is important to explore learners' conception of assessment. In the current study, a group of postgraduate students' conception of language assessment was investigated via the analysis of their metaphors and short stories. As the analysis of the metaphors showed, the participants in the current study covered all the purposes outlined by Fulcher (2010). Therefore, it seems logical to conclude that these participants are well aware of the purposes of language assessment. One possible reason could be these participants' relatively high assessment literacy as postgraduate candidates in Applied Linguistics. Another explanation for the participants' relatively high level of language assessment literacy could be attributed to the fact that apart from being postgraduate students most of them were or had been language teachers and had experienced teaching and assessing English as a second or foreign language.

The purpose of assessment which was most frequently pointed out in the metaphors was "to guide." This finding is similar to the Malaysian English language instructors' viewpoint in Niveen et al.'s (2017) study where assessment is regarded as a vital tool which guides and improves students' learning and instructors' teaching. The results also showed that one of the purposes that had the lowest frequency was "to empower." It can be inferred from this result that the participants come from communities in which assessment is conceived as a testing rather than a learning tool. Our findings are in line with those of previous studies by Cheng et al. (2008), Remesal and Brown (2015), and Azis (2012, 2015) where grades were perceived as the indicator of students' success, and assessment made students so competitive that they made every attempt to obtain higher grades rather than trying to improve their learning and understanding of the subject matter. These results also are in line with the themes which emerged from most of the short stories in which assessment was conceived often as an ends that triggers maladaptive behavior such as cheating, anxiety, and prejudice, and yet rarely does it function as a means that promotes collaboration and improvement. Our results also corroborated those of Tong and Adamson's (2015) survey in Hong Kong, who reported students' negative conceptions of and dissatisfaction with assessment practices in their respective educational system. Their respondents also agreed that feedback enhanced their learning. The analysis of short stories in our study also indicated the usefulness of formative and ongoing assessment which accounts for particular areas of difficulty faced by learners.

The comparison of the analysis of the metaphors and short stories shows that very few (5 out of 21) metaphors created by the participants presented negative conceptions toward assessment. On the contrary, as the short story data indicated, most (2 out of 3) of the short stories represented negative themes. This result could be attributed to the effect of the model short story since the participants had performed the metaphor task before they started the second task and before they read the model short story, *The Teacher*, which also happened to have a negative theme. Another possible explanation

could be the effect of the language testing course that the participants were experiencing. As the course continued, they gained better understanding of assessment and how far assessment practices that they had experienced were far from ideal. Therefore, their views toward assessment practices could have changed and they could have become more critical about the way assessment is typically practiced in their communities.

Conclusions

The objective of the present study was to explore a group of postgraduate students' conceptions of language assessment by analyzing the metaphors and short stories that they wrote about assessment. Based on the analysis of the metaphors created by the participants of this study, it was found that assessment helps teachers guide, motivate, control, and empower their learners. When assessment was viewed as a means of accountability and improvement, it often triggered adaptive behavior in learners and guided their learning. As it can be inferred from our results, very few participants considered assessment an empowering tool. This implies that steps need to be taken to promote assessment *for* learning rather than assessment *of* learning so that learners' viewpoints change about assessment and more learners start seeing assessment as a tool in service of learning.

The results of analysis of the short stories lead us to conclude that process-oriented and progressive assessment are more likely to lead to adaptive behavior and success stories while product-oriented and summative assessment will result in maladaptive behavior. Based on constructivist approach to learning, presented earlier in this paper and based on the current results, it can be concluded that effective assessment:

1. Is ongoing and progressive;
2. Motivates learners to excel;
3. Promotes collaborative learning; and
4. Helps learners and teachers diagnose areas of difficulty.

In addition, it can also be concluded that inadequate assessment:

1. Is unfair;
2. Is culturally biased;
3. Urges learners to cheat;
4. Causes fear and anxiety;
5. Causes failure and suicide;
6. Results in biased judgment;
7. Only considers the end-product;
8. Becomes an unfair gate-keeping tool;
9. Encourages the teacher to teach to the test; and
10. Leaves the teacher blind to what the learner is going through.

Surprisingly, in four stories, cheating was the result of unfair assessment practices. Teachers' malpractice entails and, interestingly, justifies students' unacceptable behavior. The prime goal of any educational system must be developing individuals with high

integrity levels. When our assessment lacks fairness, reliability, and validity, it will encourage cheating. In all the success stories, assessment motivated and spurred the learner to excel. This provides evidence for positive washback effect of assessment when it is practiced properly. As it was the case in most of these stories, assessment was practiced formatively in an ongoing manner.

In some stories, very interesting issues were addressed. In one particular story, a good student, who is under extensive pressure to succeed, fails a test, and therefore, she decides to commit suicide. This suggests that more attempts need to be made to help the community cope with failure. With the fast pace of life these days, people have become less tolerant of failure and they expect (or are expected to) accomplish too much within a short period. Students and parents need to loosen up a bit and must learn that failure is an invaluable indicator of our way to success. More support should be provided by the educational systems for the community to deal with its misconception of failure.

The present study has some limitations. One is its small sample size. A sample of about 20 students from one faculty in a public university cannot be representative of the whole respective population of postgraduate student population in Malaysia. Therefore, the results of this study cannot be transferable for the whole context of Malaysia. Another limitation is the methods of data collection. A study which uses multiple methods of data collection and data triangulation would result in more dependable outcomes. Therefore, further studies using multiple data collection methods, such as interviews, focus group discussions, or reflective journals, will result in interesting findings. Also, different forms of tasks could be used for investigating assessment beliefs and conception of students from different age groups. For example, it would be very interesting to see young learners' conception of assessment by analyzing their paintings of assessment. More qualitative studies comparing conceptions and experiences of different groups of people will result in more interesting insights. Some examples would be students from diverse educational levels or varying majors, parents with different educational levels, teachers with different experience levels, and participants from diverse nationalities. As the data of the present study were collected by the course lecturer, they could have been affected by social desirability bias. In future research, data could be collected through games to elicit more natural and real data.

Appendix 1

Task 1 (Metaphors about assessment)

Instructions: Complete the sentence, "A world without assessment is..." with your own metaphors and then explain it.

Appendix 2

Task 2 (Stories of language assessment)

Instructions: Write a short story of no more than 1000 words on the theme of assessment. The story should address an issue in the area of language evaluation. You may follow the steps below to write your story:

Plot > Peer-feedback > Draft 1 > Self-assessment > Draft 2 > Peer-feedback > Draft 3 > Marking and selection for publication.

Marking scheme: The short stories will be scored based on their:

- Task fulfillment (meeting the deadline, writing a short story within the word limit, and covering an issue in language evaluation) 3%
- Clarity (Is the story able to present the issue in a simple and comprehensible way?) 3%
- Accuracy (Is the language accurate?) 2%
- Engagement (Does the story attract and engage the audience?) 2%

Here is an example: <https://teachershelp.ru/the-teacher-by-catherine-lim/>.

Abbreviations

CFA: Confirmatory factor analysis; IELTS: International English Language Testing System; LAR: Language assessment reform; SCoA: Students' conceptions of assessment; SEM: Structural equation modeling; ZPD: Zone of proximal development

Acknowledgements

The authors would like to thank the reviewers for their useful comments.

Funding

This project was funded by the Research Management Center, Universiti Putra Malaysia (project code: UPM/700-2/1/GBP/2017/9550500).

Availability of data and materials

The data of the metaphors have been shared in this manuscript while the short stories have been published in a book which is available at <https://goo.gl/Q4t1dC>.

Authors' contributions

VN made contributions to the idea of the study, data collection, and data analysis. NISMH helped VN in writing the manuscript. Both authors read and approved the final manuscript.

Authors' information

Vahid Nimehchisalem holds a PhD in Teaching English as a Second Language and is a Senior Lecturer in the Faculty of Communication and Modern Languages, UPM. His main areas of research interest include ESL writing, language assessment and testing, and ELT material evaluation. His research activity revolves around English language learning-teaching. He is Chief Editor of the International Journal of Education and Literacy Studies and editorial team member of several journals.

Nur Izyan Syamimi Binti Mat Hussin is currently an MA English Language student from the Faculty of Modern Languages and Communication, UPM. Her areas of interest include ELT textbook evaluation, ESL writing, and language assessment.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Received: 23 January 2018 Accepted: 27 June 2018

Published online: 01 August 2018

References

- Azis, A. (2012). Teachers' conceptions and use of assessment in student learning. *Indonesian Journal of Applied Linguistics*, 2(1), 41–51.
- Azis, A. (2015). Conceptions and practices of assessment: a case of teachers representing improvement conception. *TEFLIN Journal*, 26(2), 129–154.
- Barnes, N, Fives, H, Dacey, CM (2015). Teachers' beliefs about assessment. In H Fives, MG Gill (Eds.), *International handbook of research on teacher beliefs*, (pp. 284–300). New York: Routledge.
- Brown, GTL. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11, 301–318 <https://doi.org/10.1080/0969594042000304609>.
- Brown, GTL (2008). *Conceptions of assessment: understanding what assessment means to teachers and students*. New York: Nova Science Publishers.
- Brown, GTL, & Hirschfeld, GHF. (2008). Students' conceptions of assessment: links to outcomes. *Assessment in Education: Principles, Policy & Practice*, 15(1), 3–17.
- Brown, GTL, Pishghadam, R, Sadafian, SS. (2014). Iranian university students' conceptions of assessment: using assessment to self-improve. *Assessment Matters*, 6, 5–33.
- Cheng, L, Rogers, T, Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classroom. *Assessment & Evaluation in Higher Education*, 33(1), 9–32.
- Cortazzi, M, Jin, L, Zhiru, W (2009). Cultivators, cows and computers: Chinese learners' metaphors of teachers. In *Internationalising the university*, (pp. 107–129). London: Palgrave Macmillan.

- Entwistle, NJ, & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: the student experience and its implications. *Higher Education*, 22(3), 205–227.
- Fives, H, & Buehl, MM (2012). Spring cleaning for the “messy” construct of teachers’ beliefs: What are they? Which have been examined? What can they tell us? In KR Harris, S Graham, T Urdan (Eds.), *APA educational psychology handbook: individual differences and cultural and contextual factors*, (vol. 2, pp. 471–499). Washington, DC: APA.
- Fulcher, G (2010). *Practical language testing*. London: Hodder Education.
- Fulmer, GW, Lee, ICH, Tan, KHK. (2015). Multi-level model of contextual factors and teachers’ assessment practices: an integrative review of research. *Assessment in Education: Principles, Policy, and Practice*, 22, 475–494. <https://doi.org/10.1080/0969594X.2015.1017445>.
- Hakim, B. (2015). English language teachers’ ideology of ELT assessment literacy. *International Journal of Education & Literacy Studies*, 3(4), 42–48.
- Jin, L, & Cortazzi, M (2011). The changing landscapes of a journey: educational metaphors in China. In J Ryan (Ed.), *Education reform in China: Changing concepts, contexts and practices*, (pp. 113–131). London: Routledge.
- López, AA (2008). *Potential impact of language tests: examining the alignment between testing and instruction*. Saarbrücken: VDM Publishing.
- López, MAA, & Bernal, RA. (2009). Language testing in Colombia: a call for more teacher education and teacher training in language assessment. *Profile: Issues in Teachers’ Professional Development*, 11(2), 55–70.
- Majid, FA. (2011). School-based assessment in Malaysian schools: the concerns of the English teachers. *Journal of US-China Education Review*, 8(10), 1–15.
- Marlowe, BA, & Page, ML (2005). *Creating and sustaining the constructivist classroom*, (2nd ed.,). Thousand Oaks: Corwin Press.
- Marton, F, & Säljö, R (1997). Approaches to learning. In F Marton, D Hounsell, N Entwistle (Eds.), *The experience of learning: implications for teaching and studying in higher education*, (2nd ed., pp. 39–59). Edinburgh: Scottish Academic Press.
- Ministry of Education, Malaysia (2013). *Malaysia Education Blueprint, 2013–2025*. Putrajaya: Ministry of Education Malaysia Retrieved from shorturl.at/fjs13.
- Mohamed, S, Kamis, A, Ali, N (2016). Gauging the assessment literacy of Malaysia’s home economics teachers: an empirical study. *Geografia - Malaysian Journal of Society and Space*, 12(3), 130–138.
- Niveen, RE, Ain Nadzimah, A, Sabariah, MR. (2017). Malaysian instructors’ assessment beliefs in tertiary ESL classrooms. *International Journal of Education & Literacy Studies*, 5(2), 29–46.
- Pavlenko, A. (2007). Autobiographic narratives as data in applied linguistics. *Applied Linguistics*, 28(2), 163–188.
- Ramsden, P (1997). The context of learning in academic departments. In F Marton, D Hounsell, N Entwistle (Eds.), *The experience of learning: implications for teaching and studying in higher education*, (2nd ed., pp. 198–217). Edinburgh: Scottish Academic Press.
- Razavipour, K, & Rezagah, K. (2018). Language assessment in the new English curriculum in Iran: managerial, institutional, and professional barriers. *Language Testing in Asia*, 8(1), 1–18.
- Remesal, A, & Brown, GT. (2015). Conceptions of assessment when the teaching context and learner population matter: compulsory school versus non-compulsory adult education contexts. *European Journal of Psychology of Education*, 30(3), 331–347.
- Rezvani Kalajahi, SA, & Ain Nadzimah, A. (2016). Assessing assessment literacy and practices among lecturers. *Pedagogika*, 124(4), 232–248.
- Richmond, HJ. (2002). Learners’ lives: a narrative analysis. *The Qualitative Report*, 7(3), 1–14 Retrieved from <https://nsuworks.nova.edu/tqr/vol7/iss3/4>.
- Segers, M, Dochy, F, Cascallar, E. (2006). *Optimising new modes of assessment: in search of qualities and standards*. Dordrecht: Kluwer Academic Publishers.
- Stiggins, RJ. (2002). Assessment crisis: the absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758–765.
- Suah, SL, & Ong, SL. (2012). Investigating assessment practices of in-service teachers. *International Online Journal of Educational Sciences*, 4(1), 91–106.
- Tong, SYA, & Adamson, B. (2015). Student voices in school-based assessment. *Australian Journal of Teacher Education*, 40(2), 2.
- Vygotsky, LS. (1978). *Mind in society: the development of higher psychological processes*. Cambridge: Harvard University Press.
- Zainuriyah, AK. (2012). *A Study on English Teachers’ Assessment Practices on the School-based Assessment for English Language in Malaysia*. Retrieved from <https://goo.gl/5RCm3X>.