

## Challenges and Potential Solutions for Assessment Practice in Pentaksiran Tingkatan Tiga (Pt3)

Muhamad Firdaus Mohd Noh

*Faculty of Education, Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia.*

Mohd Effendi @ Ewan Mohd Matore\*

*Research Centre of Education Leadership and Policy, Faculty of Education, Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia*

### **Abstract**

*The objective of high-stakes assessment is to appraise students' achievement at the end of an academic calendar. Students' result in the assessment is impactful in determining their future life such as a qualification for enrolment in university or job interviews. In operationalizing high-stakes assessment, standardization is an important element to ensure students are assessed with fairness and only based on their abilities. High-stakes assessment is normally administered by a centralised body without the involvement of students' respective schools. However, the decision to assign students' teachers as the sole examiner of high-stakes assessment is contentious whether teachers can adhere to standardization and measurement principles. Therefore, this paper aims to discuss the challenges faced by teachers when they are tasked to assess students in high-stakes assessment particularly in the context of Pentaksiran Tingkatan Tiga (PT3) in Malaysia and propose several potential solutions. The main idea underpinning this paper is to enhance the quality of assessment practice in education system which is salient to assure validity and reliability of results received by students. Such discussion is also needed to acknowledge teachers as a professional player in the arena of educational assessment. More importantly, the discussion in this paper puts forward an urge for educational stakeholders to redesign assessment system based on research-based practice.*

*Keywords: school-based assessment, high-stakes assessment, teachers' involvement, lower secondary school assessment, standardization*

### **1.0 INTRODUCTION**

Teachers, as the main actor in cascading the aspiration of any country's education system to the young generations, are expected to be expert in assessing their own students. Assessment in education plays an important role to recognize the extent to which learning process has engrained impact on students in terms of their cognitive and behavior. In understanding the aims of educational assessment, it is best to look at how the assessment results are used which can be discussed in three kinds. Firstly, assessment can be either formative or summative (AERA et al., 2014). Formative assessment happening in classroom or within school context accentuates the utilization of information on students' progress to help teachers and students to further enhance their teaching and learning. Meanwhile, summative assessment which is normally implemented at the end of a period of learning sessions functions to segregate students based on their achieved result. Next, school-based assessment versus centralized assessment. The former can be executed either through formative or summative assessment, acknowledges teachers as the main assessor of students' learning without interference from external parties. On the contrary, centralized assessment refers to one that is administered by a central body in the education system upon all schools. On top of that, educational assessment can also be discussed through how the results are used in determining high-stake decision in students' life for instance in order to enroll for university programs or to apply for a job interview (Coniam & Falvey, 2018). On the other hand, low-stakes assessment such as monthly test does not involve any important decision in students' future life.

The different kinds of assessment are indispensable in any educational institutions globally including in Malaysian schools. A recent revamp of educational assessment in Malaysian secondary schools is the change from Penilaian Menengah Rendah (PMR), a centralized high-stakes assessment, to Pentaksiran Tingkatan 3 (PT3), a school-based high-stakes assessment in 2014. This revamp has elevated teachers' roles from merely a teacher to an assessor in the context of high-stakes assessment which clearly creates a total new working culture among teachers. It also means that teachers are now expected to balance between assessing students for the sake of students' learning within school context and assessing students for high-stakes purpose.

## 2.0 LITERATURE REVIEW

High-stakes assessment (HSA) refers to any assessment which the result is used to make significant decision about candidates and impactful in their lives for instance to screen for admission to tertiary education and examination for professional certification (Coniam & Falvey 2018). Due to its important use of the result, standardization is a critical element in HSA. Standardization is when norm-referenced assessment happens to compare candidates' scores with that of a standard candidate depending on the context of the assessment (Kempf 2016) either school, national or international level. As the context gets bigger, standardization is more valued as to promise accountability, fairness and equity to all candidates. HSA is normally centralized in respective countries and administered by external parties as the main assessors. However, there are also countries that employ candidates' teachers to be the assessors including Malaysia. Such effort is because HSA is claimed to be isolated from the reality in classroom in addition to appreciate teachers' expertise as the key players in pedagogical routines and to be transparent in any examination reform towards practitioners (Xerri & Briffa, 2018).

### 2.1 Comparison of Lower Secondary School Assessment among Leading Countries

Every country has their own assessment system based on their aspiration and culture. Table 1 portrays a comparison between nine leading countries in education on how assessment for lower secondary school students are designed. The comparison includes whether these countries implement centralized high-stakes assessment (CHSA), engage students' teachers in the assessment procedures, incorporate school-based assessment (SBA) and device entrances exam for special schools in the upper secondary schools.

**Table 1. Implementation of lower secondary school assessment compared by countries**

Countries	Centralized high-stakes assessment	Teachers' involvement	School-based assessment	Entrance exam for special schools
Singapore	x	x	x	x
Hong Kong	x	x	x	x
Japan	√	x	x	x
Taiwan	√	x	x	x
India	√	x	x	x
South Korea	√	x	x	√
China	x	√	√	√
Finland	x	√	√	√
Malaysia	√	√	√	√

Both Singapore and Hong Kong do not implement centralized high-stakes assessment (CHSA), teachers' involvement, school-based assessment (SBA) and even entrance examination for lower secondary students. In Singapore, the streaming of students starts immediately after they complete primary school with two options, namely normal academic stream and normal technical stream (Singapore Examination and Assessment Board 2019). Then, upon completion of secondary school years, students sit for Singapore-Cambridge General Certificate of Education (GCE) to enroll in tertiary education. However, school-based assessment is nowhere a part of centralized high-stakes assessment. Meanwhile in Hong Kong, lower secondary school students do not need to select any specific stream for their upper secondary school years. Elective subjects are offered and open for any students to take in accordance to their interest and consideration (Hong Kong Examination and Assessment Board 2019). The only centralized HSA is when they finish secondary school to enter colleges or universities in which SBA has 20% to 50% contributing to the final grade (Hong Kong Examination and Assessment Board 2019).

Japan, South Korea, Taiwan and India administer CHSA but teachers and school-based assessment do not contribute to any part of the final grade. In Japan, the level of high-stakes in the assessment is highly competitive as it distinguishes students for special programs and involves paper-and-pencil test as well as follow-up interview (Ministry of Education, Culture, Sports, Science and Technology of Japan 2006). Similarly, examination result is the determinant of future schooling of the students in Taiwan and India. On the contrary, the centralized assessment in South Korea is only a requirement for entrance to any general high schools but special entrance examination to special and autonomous schools devised by these respective schools are competitive as there is a minimum number of intakes every year (Mani & Trince 2018).

The implementation of lower secondary school assessment in China and Finland is similar that these two countries do not opt for CHSA, but the assessment takes place through SBA and teachers' involvement. Students in China who wish to pursue in general (academic) track must pass the entrance examinations for general senior secondary schools. Meanwhile in Finland, the selection criteria used by vocational institutions is mainly based on students' grades in their basic education certificate, working experience and aptitude tests.

## **2.2 General Certificate of Education in England**

Interestingly in England, the percentage of SBA in the General Certificate of Education (GCE) has witnessed a major reduction. Since 1980s, GCE was designed to have a huge portion of SBA named "coursework". Then, due to reliability and validity issues and the increase of report on plagiarism, in 2009 coursework has been replaced by "controlled assessment" in which control levels are set for each stage of assessment process. Controlled assessment is slightly different from coursework in which higher level of control including from external parties is demanded over the operationalization of assessment (Cuff, Meadows & Black 2018). Recently in 2016, the percentage of final examination has taken over overall assessment in some subjects like English Literature, Physics, Geography, Business and History leaving no percentage for SBA and only few subjects retain the element of SBA with small percentage (Oppos 2016). Non-examination assessment is only allowed for certain subjects with grade given is either pass or fail, clear segregation of marks (Oppos 2016) and the combination of consistency marking from teachers and external moderation process is highly emphasized (Crisp 2018).

## **2.3 Malaysia's Practice of Lower Secondary School Assessment**

CHSA in Malaysian schooling system starts right after the end of primary school named as Ujian Penilaian Sekolah Rendah (UPSR), followed by Pentaksiran Tingkatan Tiga (PT3) at the end of lower secondary school, Sijil Pelajaran Malaysia (SPM) at the end of form five and Sijil Tinggi Pelajaran Malaysia (STPM) for form six students. In particular, only PT3 has both teachers' involvement and SBA element contributing to final grade. The history of assessment for lower secondary school students in Malaysia started with Sijil Rendah Pelajaran (SRP) which was then changed to Penilaian Menengah Rendah (PMR) until recently the government has decided to revamp it to PT3. The commonality of SRP, PMR and PT3 is that the assessment results are used for high-stakes decision in streaming students into either science, technical, arts or business streams in upper secondary school. It is also utilized to screen students for enrolment in high-performing schools such as Sekolah Berasrama Penuh (SBP) and Maktab Rendah Sains Mara (MRSM) as these schools accept only excellent students.

Among the rationale behind the change from PMR to PT3 is to reduce testing anxiety among students and to incorporate the school-based assessment as part of the contributing marks. Previous research by Muhamad Firdaus and Mohd Effendi (2019) also mentioned about raters and not only students. The students need to be resilience and avoid anxiety on themselves as covered by the previous studies (Mohd Effendi Ewan Mohd Matore, 2019; Mohd Effendi Ewan Mohd Matore et al., 2018; Mohd Effendi Ewan Mohd Matore & Ahmad Zamri Khairani, 2016, 2019). It has also elevated the role of teachers from only the disseminator of knowledge to assessment examiners. Teachers are responsible in the process of assessment taking, marking students' answers and reporting to the ministry. The comparison among different countries made earlier reveals that Malaysia has a totally different practice in assessing lower secondary school students. It is therefore salient to draw our focus on challenges faced by teachers in executing PT3 as a school-based high-stakes assessment.

## **2.4 Teachers as Assessors in High-stakes Assessment**

A growing body of literature has investigated the effects of employing teachers as assessors in high-stakes assessment. One of the concerns is the subjectivity and reliability of teachers' assessment in assessing their own students. Among the issues faced by teachers in implementing high-stakes assessment are their failure to sustain standardization throughout the assessment process (Kremmel et al., 2018; Sundqvist et al., 2018), teachers' lack of skills rendering them to employ idiosyncratic approach, schools' lack of supporting system and facilities (Brookhart, 2013) and also students' inconsistent performance between in-class interaction and during assessment taking (Alm & Colnerud, 2015).

Other studies have examined the extent to which teachers are capable to produce scores that are similar with external or expert assessors. The findings reported that teachers' assessment quality in high-stakes assessment has been variable depending on many factors such as their experience, assessment training and

familiarity with students (Attali, 2016; Davis, 2016; Kang et al., 2019; Tanriverdi-Koksal & Ortactepe, 2017). However, research on operational assessment in Malaysian' education system especially in the context of high-stakes assessment has been very limited because the issue is sensitive. Nevertheless, to be competitive with other leading countries, discussion should be made on how to enhance assessment practice in the education system.

Thus, this paper aims to discuss issues aroused among Malaysian teachers when they are assigned to assess their students in high-stakes assessment, particular in the context of PT3 for lower secondary school students. Such discussion is noteworthy as to ensure that the decision to revamp PMR and to sustain PT3 as a high-stakes assessment is worth with the changes in teachers' roles. Discussion in this paper hopefully can shed some light into how to best assess students based on research-driven practice without compromising elements of quality assessment such as reliability and validity.

### **3.0 CHALLENGES OF PT3**

#### **3.1 First Challenge: The Tension between School-Based Assessment and High-Stakes Assessment**

As compared to PMR, PT3 requires teachers to not only execute teaching to prepare students for standardized examination but also to administer the assessment procedures from conducting the assessment, marking students' answers, finalizing students' final mark, involving in moderation process to finally submit report. However, it does not mean teachers are excused from implementing assessment in classroom which they have been doing for all this while including assessing students formatively, appraise students to recognize their current level, use the information on students' mastery level to enhance teaching and also report in an offline system (previously was an online system).

In other words, PT3 teachers are now expected to administer two kinds of assessment: high-stakes assessment and school-based assessment. As a result, this situation has created tension among teachers to juggle between the two types of assessment as they are of different end on a spectrum of assessment type. At one end, teachers collect authentic evidence to support and prove how students make their progression throughout the schooling year, while at the other end, the utilization of 'real-life' evidence contradict with the notion of standardization, reliability and validity accentuated in HSA (East 2015).

This irreconcilable notion of the two types of assessment anticipates teachers to excel in their teaching practice as well as to yield quality assessment. From time to time, authority either from external body or school administrators pushes teachers to strive for both expectations. For example, teachers' teaching is observed to ensure that formative assessment is happening in their teaching scenes, files are inspected to see if each individual student's progress is recorded and teachers are required to engage in reflective projects such as action research and innovation projects to improve their pedagogical practice. This procedure has the tendency to be inimical towards the execution of HSA during which, teachers on the other hand, are also obliged to meet the standard requirement as an assessor. They should be free from any bias or unrelated inclination in appraising students to maintain reliability and standardization of the HSA even though each school has different settings.

How can we expect standardization when these two types of assessment collide between each other? This is a serious matter that should be under scrutiny because to possess the capability to distinguish the two types of assessment and make it happen according to the standard is not an easy task. Teachers are burdened to accomplish the two assessment and eventually their core business, teaching can be greatly affected. In fact, it can also affect the reliability and validity of assessment that teachers generate.

#### **3.2 Second Challenge: The Lack of Trust in Teachers' Assessment Practice**

PT3 is the only high-stakes examination administered by schools in Malaysia as compared to other centralized high-stakes testing such as SPM, STPM and MUET. Using internal assessor to mark HSA indeed brings about the question of reliability and validity of scores that teachers reward to students' answers because they know the students inside out through everyday interaction in classrooms. The key feature highlighted here is quality assurance including the procedures of assessment setting, assessment taking and assessment marking. When the awarding body of students' marks have control over the assessment procedures, it may diminish the quality of assessment provided (Vitello & Williamson 2017).

Put differently, teachers' assessment quality is questioned and teachers are not trusted to generate reliable and valid score. Hence, the current practice employs pre-assessment intervention, which is training

compulsory for all teachers before they engage in assessment procedure and post-assessment intervention, which is moderation process (internal and external). This argument is definitely valid because one cannot deny the utmost importance of standardization in high-stakes assessment setting, yet it is ironic since teachers' involvement in HSA aims in giving authority to teachers and to incorporate authentic real-life evidence from students into assessment. Furthermore, why are teachers assigned to assess in the first place if there is a need to doubt teachers' assessment at the end? As Johnson (2013) pointed out, if the objective of appointing teachers in high-stakes assessment is to respect their professionalism, then how is this objective attained?

On top of that, the need to compare teachers' judgement across teachers within schools and across schools implies the same issue which is lack of trust for teachers. Teachers are demanded to allocate enormous time and effort to gather evidence and keep records to prove that their judgment is valid (Qian 2014). The intention to sustain validity of marks teachers award is sacred but somehow it has made teaching profession full of scepticism.

Next, the lack of trust for teachers infers that status of reliability and validity of assessment only come from teachers whereas standardization does not solely come from teachers as assessor. How can we expect standardization occurs when even the number of students per classroom is different? How can we aim standardization to happen when even the facilities and amenities available in different schools are not standardized? How can we anticipate internal moderation to happen effectively when the number of teachers teaching particular subject is not sufficient? These irrelevant variance does influence the process of assessment conducted by teachers, yet they are ignored as if teachers are superhero that can produce reliable and valid scores without other noise or disturbance.

### **3.3 Third Challenge: Workloads**

Recapitulating on teachers' workload during PMR, teachers are only responsible to assess students for the purpose of classroom use and within school context during which HSA is centralized by Lembaga Peperiksaan Malaysia (LPM). During that years, only appointed teachers acting as examination markers were involved with the marking and moderation process and they were paid for each paper they marked. At present, all form three teachers are assessor. Obviously, it increases teachers' workload as now they are involved in the administration of the assessment from the beginning until report is submitted.

Furthermore, it results in imbalanced distribution of assignment among teachers. Teachers' workload is tripled if they are assigned to large class and the worst comes to the worst is when schools do not have enough teacher supply. It is also unfair that form three teachers have more workload than other form teachers. If one claims that workload is normal in any profession, the underlying issue in this problem is the extent to which this new innovation in assessment support the purpose of teaching and learning rather than adding unnecessary tasks for teachers to accomplish (McMahon & Jones 2015). In a recent study, Qian (2014) found that teachers' major concern in implementing new assessment system either for the purpose of high-stakes or low-stakes interest is the increment of their workload. It is worrisome that if this circumstance continues, teachers are just doing their job for the sake of not getting complaints and skip the intended purpose of the current practice in educational assessment. Workload does not only affect teachers but also students which in turn influence their motivation to learn (East 2016). Students compare subjects they learn based on the amount of workload they need to accomplish and obviously subjects with less assignment win over the others. This situation is alarming as it may impact students' academic performance especially those with low motivation.

## **4.0 POTENTIAL SOLUTIONS**

Two potential solutions are proposed to extricate teachers from the three challenges they grapple with in the execution of PT3 as discussed earlier. These solutions are comprehensive that they do not respond to any individual challenge but each of the solution can contribute to the minimization of the three challenges altogether.

### **4.1 First Potential Solution: Redesign PT3 as either HSA or SBA**

Indeed, the involvement of teachers as internal assessor offers some benefits as they are familiar with their students but there is an issue with standardization. A study in Sweden by Sundqvist et al. (2018) on the standardization of assessment led by teachers as internal examiner recommended the policy maker to either retain the assessment by abolishing the standardization element or redesign the assessment and its

operationalization to uphold the purpose of standardization. In essence, teachers as internal assessors cannot be the only body to assess students' product.

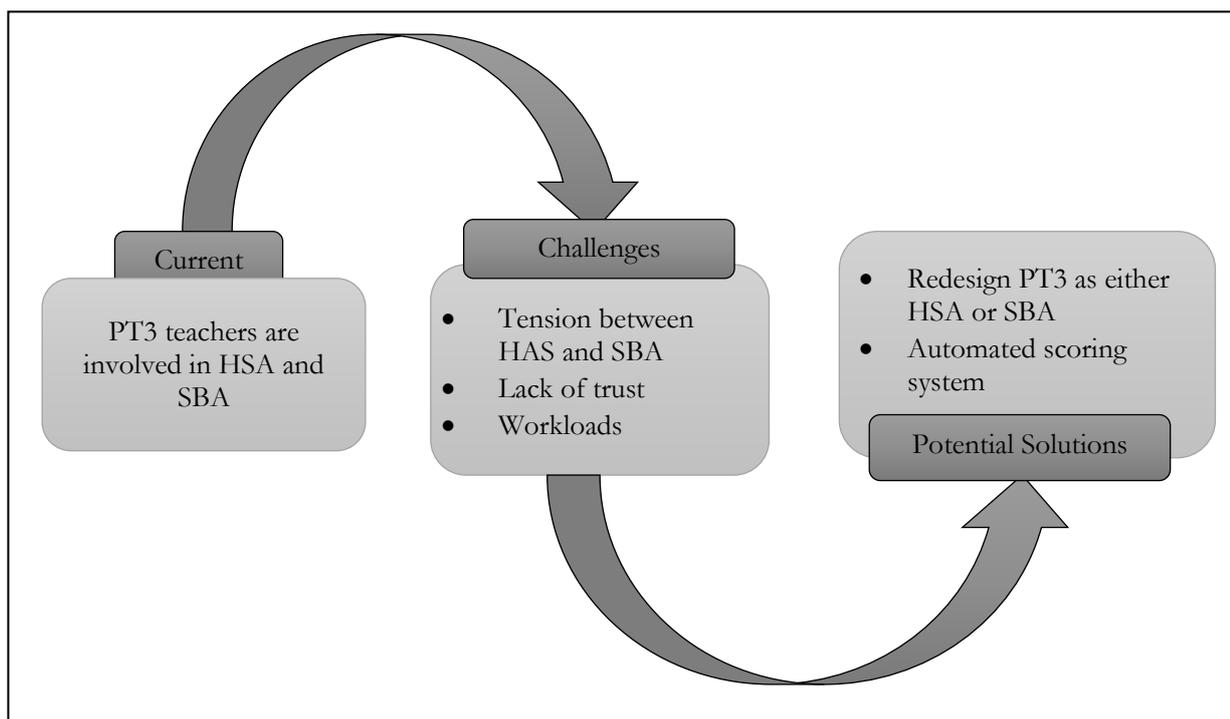


Figure1. Current practice of lower secondary school assessment in Malaysia, its challenges and potential solutions

The similarity between PMR and PT3 is that both possess high-stakes interest, so it is not appropriate to abolish the element of high-stakes in PT3 as the result is still utilized in screening students for admission to the next level of schooling. However, it is problematic to have the two kinds of purposes and operationalization in one assessment thus a decision should be made to redesign PT3. Therefore, if the element of high-stake is to be retained, PT3 should be divided into two parts: centralized final examination administrated by the examination board and marked by external assessors and also SBA that involve only assessment like project, assignment and oral test monitored by teachers.

The redesign of PT3 will greatly benefit our teachers whom now facing tension between SBA and standardized assessment by creating a crystal-clear border between the two types of assessment run by two different parties. In fact, the amount of workload that teachers are supposed to accomplish will be reduced and teachers' assessment practice will be more focused and effective. Eventually, teachers' professionalism is respected and trust is given back to them as they are only in charge of assessment within school setting.

#### 4.2 Second Potential Solution: Automated Scoring System

The use of automated scoring system (ASS), or also known as automated essay scoring (AES) and automated writing evaluation (AWE), in HSA especially for subjective scoring item is not happening in Malaysian context but it has become the practice in other countries. In addressing the challenges of recruiting, training, monitoring and moderating human raters for high-stakes assessment to generate reliable and valid score, ASS can offer a good solution (Raczynski & Cohen 2019). ASS is "a computer technology that evaluates and scores written prose using algorithms that focus on syntax-, content-, or style-related features of text" (Shermis & Burstein 2013).

Utilizing a computer program to automatically mark and grade students' examination answer is definitely faster and cheaper as compared to using teacher-assessors (Hockly 2019). It also offers scoring consistency that sometimes human raters are not capable to achieve (Ramineni & Williamson 2013). Investigating on a number of ASS software, Shermis (2014) discovered that ASS is able to attain the same level of agreement with adjudicated human ratings. Research has been carried out in different countries to find out the feasibility, validity and reliability of ASS in assessment practice (shown in Table 2) especially in comparison

to human raters. These studies denote that automated scoring is beneficial as an alternative scorer for assessment settings and in fact manages to maintain reliability, accuracy and consistency of the scored generated.

**Table 2. Research of automated essay scoring system in other countries.**

Author	Year	Country	Result
Qurratul Aini, Achmad Eko & Dwijoko	2019	Indonesia	AES has the potential to increase the reliability of the scoring (no significant difference from teachers' scoring), increase speed and reduce teachers' workload.
Cohen, Levi & Ben-Simon	2018	Israel	High correlation between automated scoring and human rater scoring but AES is more consistent.
Ranalli	2018	USA	Automated written corrective feedback (AWCF) is effective when the feedback provided is explicit and specific as compared to generic feedback.
Kieftenbelda & Boyer	2017	USA	Inter-rater agreement between automated raters and human raters is high for essay items but slightly lower for short-answer items.
Rababah & Al-Taani	2017	Jordan	ASS is capable to score answer scripts with high rates of success and consistent with human rater scoring with high correlation.

However, is it feasible and practical in Malaysian school context? Sing Yii et al. (2016) investigated the feasibility of automated essay scoring feedback (AESF) in Malaysian University English Test (MUET) classes and discovered that AESF offers high accuracy level of feedback and scores to students' work. However, they emphasized that automated scoring is beneficial to be used in real-life classroom scenes rather than nation-wide scale testing due to insufficient resources. As a matter of fact, since 2018, Malaysian University English Test (MUET) has offered candidates to choose sitting for the test using computer-based test (CBT), a kind of ASS. Even though it is still in the beginning of its implementation, this suggests that ASS is feasible in Malaysian educational assessment system with proper planning and operationalization.

As noted by Sing Yii et al. (2016), automated scoring can be utilized in classroom-based assessment settings, thus suggesting that SBA can be implemented with the aid of AES. Today, there are computer programs that offer feedback on students works immediately after they type in their answers on computers. For example, Cambridge English's 'writeandimprove' website is available for students to write their essays according to appropriate level, receive feedback from the system and improve their writing. Teachers can manipulate this software to monitor and facilitate students' academic progress without needing to manually mark students' answer script. The feedback generated from the software can be used to discuss with students on how to improve their learning. Resource management is also improved as teachers no longer need to use files to records students' progress and their development can be easily traced in the system.

## 5.0 CONCLUSION

To be competitive with developed countries, assessment in education system should be of utmost priority by ensuring educational assessment to achieve its objective and adhere to its standards. The decision to retain PT3 as high-stakes assessment comes with consequences that the standardization should be the highlighted desiderata. The discussion above has given some ideas towards the challenges of PT3 and raised some interesting questions for contemplation. Different kind of operationalization of HSA in this country as compared to other countries gives us reasons to doubt if there is any hidden agenda underlying such decision. If practices in other countries appreciate SBA as only in low-stakes assessment, why does Malaysia incorporate it in high-stakes assessment? On top of that, teachers' increasing workload may suggest that teachers contribute only a minor portion to any policy making. It is time for the authorized body to decide on only one operationalization of PT3 so that standardization is maintained, assessment quality is not sacrificed and teachers are not inundated with undue workload. This study has several limitation including it only identifies the challenges in the implementation of PT3 from teachers' perspectives in addition to differences between PMR and PT3 in terms of students' achievement are not discusses due to confidentiality issue and the political influence of the decision on changing PMR to PT3. Future research should endeavor to investigate the challenges faced by students in PT3, standardization issue in teachers' judgement potentially using statistical analysis and comparison of different HSA in Malaysian context.

## ACKNOWLEDGEMENT

We gratefully acknowledge financial support from Universiti Kebangsaan Malaysia (UKM) under Dana Pecutan Penerbitan Tahun 2020 FPEND and Dana Penyelidikan FPEND (Grant number: GG-2019-034). We thank to everyone who provided insight and expertise that greatly assisted the research. We thank all the experts for assistance with constructive comments that greatly improved the manuscript.

## REFERENCES

1. AERA, APA, & NCME. (2017). *Standards for Educational and Psychological Testing: National Council on Measurement in Education*. Washington: American Educational Research Association.
2. Alm, F., & Colnerud, G. (2015). Teachers' Experiences of Unfair Grading. *Educational Assessment*, 20(2), 132–150. <https://doi.org/10.1080/10627197.2015.1028620>
3. Attali, Y. (2016). A comparison of newly-trained and experienced raters on a standardized writing assessment. *Language Testing*, 33(1), 99–115. <https://doi.org/10.1177/0265532215582283> [ltj.sagepub.com](http://ltj.sagepub.com) Abstract
4. Brookhart, S.M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy and Practice*, 20(1), 69–90. <https://doi.org/10.1080/0969594X.2012.703170>
5. Cohen, Y., Levi, E., & Ben-Simon, A. (2018). Validating human and automated scoring of essays against “True” scores. *Applied Measurement in Education* 31(3): 241-250.
6. Coniam, D., & Falvey, P. (2018). *High-Stakes Testing: The Impact of the LPATE on English Language Teachers in Hong Kong*. Singapore: Springer.
7. Crisp, V. (2018). Insights into teacher moderation of marks on high-stakes non-examined assessments. *Research Matters: A Cambridge Assessment Publication* 25(2018): 14-20.
8. Cuff, B.M.P., Meadows, M., & Black, B. (2018). An investigation into the Sawtooth Effect in secondary school assessments in England. *Assessment in Education: Principles, Policy & Practice* (2018): 1-19.
9. Davis, L. (2016). The influence of training and experience on rater performance in scoring spoken language. *Language Testing*, 33(1), 117–135. <https://doi.org/10.1177/0265532215582282>
10. East, M. (2015). Coming to terms with innovative high-stakes assessment practice: Teachers' viewpoints on assessment reform. *Language Testing* 32(1): 101–120.
11. East, M. (2016). *Assessing Foreign Language Students' Spoken Proficiency: Stakeholder Perspectives on Assessment Innovation*. Singapore: Springer.
12. Hockly, N. (2018). Automated writing evaluation. *ELT Journal* (2018): 82-88.
13. Hong Kong Examination and Assessment Board. 2019. Hong Kong Diploma Secondary Education Examination. <http://www.hkeaa.edu.hk/en/hkdse/introduction/> [7 April 2019].
14. Johnson, S. (2013). On the reliability of high-stakes teacher assessment. *Research Papers in Education* 28(1): 91-105.
15. Kang, O., Rubin, D., & Kermad, A. (2019). The effect of training and rater differences on oral proficiency assessment. *Language Testing*, 36(4), 481–504. <https://doi.org/10.1177/0265532219849522>
16. Kempf, A. (2016). *The Pedagogy of Standardized Testing: The Radical Impact of Educational Standardization in the US and Canada*. New York: McMillan.
17. Kieftenbeld, V., & Boyer, M. (2017). Statistically comparing the performance of multiple automated raters across multiple items. *Applied Measurement in Education* 30(2): 117-128.
18. Mani, D., & Triance, S. (2018). Education in South Korea. <https://wenr.wes.org/2018/10/education-in-south-korea> [7 April 2019].
19. McMahon, S., & Jones, I. (2015). A comparative judgement approach to teacher assessment. *Assessment in Education* 22(3): 368-389.
20. Ministry of Education, Culture, Sports, Science and Technology of Japan. (2006). Basic Act on Education. <https://web.archive.org/web/20160331133057/http://www.mext.go.jp/english/lawandplan/1303462.htm> [6 April 2019].
21. Mohd Effendi Ewan Mohd Matore. (2019). The impact of EQ , SQ and IQ towards AQ using path analysis and Rasch logit for modelling purpose among mechanical engineering students. *Proceedings of Mechanical Engineering Research Day 2019, August*, 225–226.
22. Mohd Effendi Ewan Mohd Matore, & Ahmad Zamri Khairani. (2016). Correlation between Adversity Quotient (AQ) with IQ , EQ and SQ Among Polytechnic Students using Rasch Model. *Indian Journal of*

- Science and Technology*, 9(47). <https://doi.org/10.17485/ijst/2016/v9i47/108695>
24. Mohd Effendi Ewan Mohd Matore, & Ahmad Zamri Khairani. (2019). Perhubungan antara Adversity Quotient (AQ) dengan Pencapaian Akademik Pelajar. *Global Journal Al-Thaqafah (GJAT), Special Is*(May), 63–78. <http://www.gjat.my/gjat2019si/SI2019-06.pdf>
  25. Mohd Effendi Ewan Mohd Matore, Ahmad Zamri Khairani, Siti Mistima Maat, Nor Adila Ahmad, & Effa Rina Mohd Matore. (2018). The Influence of Intellectual Quotient (IQ), Emotional Quotient (EQ) and Spiritual Quotient (SQ) against Adversity Quotient (AQ) on polytechnic students in Malaysia. *Journal of Engineering Science and Technology, Special Is*(2018), 83–91.
  26. Muhamad Firdaus Mohd Noh, & Mohd Effendi Ewan Mohd Matore. (2019). Brunswiks Lens Model : This Is How to Inspire Accurate Raters. *Creative Education*, 10(20), 2859–2868. <https://doi.org/10.4236/ce.2019.1012212>
  27. Oppos, D. (2016). Whatever happened to school-based assessment in England’s GCSEs and A levels? *Perspectives in Education* 34(4): 52-61.
  28. Qian, D.D. (2014). School-based English language assessment as a high-stakes examination component in Hong Kong: insights of frontline assessors. *Assessment in Education: Principles, Policy & Practice* 21(3): 251-270.
  29. Qurratul Aini, Achmad Eko, & Dwijoko, P. (2018). Development of a Scoring Application for Indonesian Language Essay Questions. *Proceedings of the 2018 2nd International Conference on Education and E-Learning*, hlm. 6-10.
  30. Rababah, H., & Al-Taani, A.T. (2017). An automated scoring approach for Arabic short answers essay questions. *2017 8th International Conference on Information Technology (ICIT)*, pp. 697-702.
  31. Raczynski, K., & Cohen, A. (2018). Appraising the scoring performance of automated essay scoring systems - Some additional considerations: Which essays? Which human raters? Which scores? *Applied Measurement in Education* 31(3): 233-240.
  32. Ramineni, C., & Williamson, D.W. (2013). Automated essay scoring: Psychometric guidelines and practices. *Assessing Writing* 18: 25–39.
  33. Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning* 31(7): 653-674.
  34. Shermis, M.D. (2014). State of the art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing* 20: 53–76.
  35. Shermis, M.D. & Burstein, J. (2013). *Handbook of Automated Essay Evaluation Current Applications and New Directions*. New York, NY: Routledge.
  36. Singapore Examination and Assessment Board. (2019). Examinations. <https://www.seab.gov.sg/home/#> [7 April 2019].
  37. Sing Yii, N., Chih How, B., Nung Kion, L., & Kian Sam, H. (2016). Automated essay scoring feedback (AESF): An innovative writing solution to the Malaysian University English Test (MUET). *International Journal on E-Learning and Higher Education* 4(2016):130-143.
  38. Sundqvist, P., Wikstrom, P., Sandlund, E., & Nyroos, L. (2018). The teacher as examiner of L2 oral tests: A challenge to standardization. *Language Testing* 35(2): 217-238.
  39. Tanriverdi-Koksal, F., & Ortactepe, D. (2017). Raters knowledge of students proficiency levels as a source of measurement error in oral assessments. *Hacettepe University Journal of Education*, 32(3), 1–19. <https://doi.org/10.16986/HUJE.2017027583>
  40. Vitello, S., & Williamson, J. (2017). Internal versus external assessment in vocational qualifications: A commentary on the government’s reforms in England. *London Review of Education* 15(3): 536-548.
  41. Xerri, D., & Briffa, P.V. (2018). *Teacher Involvement in High-Stakes Language Testing*. Switzerland: Springer.