

Developing an Automated Essay Scorer with Feedback (AESF) for Malaysian University English Test (MUET): A Design-based Research Approach

Sing Yii Ng¹, Chih How Bong², Kian Sam Hong¹ and Nung Kion Lee^{1*}

¹*Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia*

²*Faculty of Computer Sciences and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia*

ABSTRACT

This paper presents the development of an automated essay scoring mechanism based on the Malaysian University English Test essay marking criteria using the Design-based research (DBR). It is a learning intervention to facilitate students in their essay writing process and at the same time, serves as a tool for teachers to mark essay. DBR is the most commonly used method for conducting research in technological enhanced learning context especially for solving real classroom problem. The development of the automated scoring system is presented step by step following the four phases in DBR model. In each phase, data collection procedure, research instrument and the lessons learnt that lead to further iterations are discussed in order to produce a workable and effective automated essay grader. The

outcome resulted from the five iterations lead to the present intervention, Automated Essay Scorer with Feedback (AESF). This system allows teachers to collect samples of marked essays to be trained to grade newly entered essays. Then the teacher can set task and keep track of students' progress and provide additional feedback as well as rectify the scores generated. For students, they can practice writing essays and demand for feedback at any point of their essays writing process for the system to provide scores by paragraph as well as the whole

ARTICLE INFO

Article history:

Received: 26 October 2017

Accepted: 19 February 2019

Published: 13 September 2019

E-mail addresses:

ngsingyii@gmail.com (Sing Yii Ng)

chbong@unimas.my (Chih How Bong)

hksam1964@gmail.com (Kian Sam Hong)

nklee@unimas.my (Nung Kion Lee)

*Corresponding author

essay. The system was tested by 24 teachers from 5 schools in real-classroom context with favorable comment.

Keywords: Automated essay scorer with feedback (AESF), design-based research, Malaysian University English Test (MUET)

INTRODUCTION

Writing is an important productive skills that students of all disciplines need to master (Graham, et al., 2013). This skill is often under-developed due to time constraints (Weigle, 2007). Writing an academic piece of writing involve time and similarly, more time is needed for the teachers to read, grade and provide feedback (Kellogg et al., 2010). It is necessary to provide timely feedback to let students to have better understanding of the given task before they forget or lose interest of the topic written (Ahmad & ul Hussnain, 2012; Lipnevich & Smith, 2009). A technology-based system, in this case the Automated Essay Scorer with Feedback presented in this study can help teacher score and provide immediate language feedback to students. At the same time, this automated essay scorer with feedback provides students a platform to write essays, get feedback on demand as well as access to immediate scores based on paragraph and complete essay as a whole.

In order to design a technologically rich teaching and learning experience, design-based research (DBR) approach has gained popularity as this research model calls for improving an intervention based on the context of occurrence by seeking

help from expert as well as practitioners in reality (Reeves & McKenney, 2015). This is a pragmatic approach that utilises both the quantitative and qualitative data to solve a real classroom problem (Reeves & McKenney, 2015).

DBR is the most commonly used methodology when conducting research in technological enhanced learning context (Kennedy-Clark, 2013). In line with the pragmatic worldview, DBR “seeks to increase the impact, transfer and translation of education research into improved practice” (Anderson & Shattuck, 2012), a term synonym with developmental research that focuses on solving complex real world problem critical to education while at the same time leads to theory construction and explanation (Reeves, 2006).

Since learning is a complex phenomenon that cannot be solved by only one discipline, DBR allows researchers to derive important characteristics about the messiness of natural condition (Bell, 2004) so that we can systematically understand and predict how learning occurs, then attempt to create and sustain educational innovation in actual everyday classroom setting that is not merely workable in the laboratory (Barab & Squire, 2004). In terms of sustenance, it requires the understanding of how and why an innovation may have work and vice versa so that on-going improvement can be made over time and across setting (Joseph, 2004). It usually attempts to connect the relationship between the theory, designed innovation and practice where the innovation may even lead to new teaching

and learning theory (The Design-Based Research Collective, 2003)

The process in DBR does not stop at merely testing the innovation in a particular experiment. Its iterative nature requires cyclic processes where improvement is made at every level of testing from its prototype to actual testing grounds with teachers and students from various settings. The reflection on why the innovation works will strengthen the theory proposed while failures will indicate more rooms for improvement and the generation of more validated data (Reeves, 2006). In short, the fundamental principles of DBR are listed below:

- Addressing complex problem in real contexts in collaboration with practitioners;
- Integrating known and hypothetical design principles with technological advances to render plausible solutions to these complex problems; and
- Conducting rigorous and reflective inquiry to test and refine innovative learning environment as well as to define new design principles. (Reeves, 2006)

DBR serves as the main approach for the current study on designing and developing an Automated Essay Scorer with Feedback (AESF) environment to facilitates students essay writing and teachers essay marking because it is “a systematic but flexible methodology aimed to improve educational practice through iterative analysis, design, development and implementation based

on collaboration among researchers and practitioners in a real-world setting and leading to contextual sensitive principles and theory.” (Anderson & Shattuck, 2012).

The AESF is based on the behaviourist theory which espouses more practice leads to better performance (Mitchell, 2013) and the humanist theory that higher motivation leads to more the satisfaction in improving the targeted skills (Mitchell, 2013). It is believed that with this new innovation, the immediacy of feedback to the extent of paragraph by paragraph leads to new method in assessing writing in general, where commonly teachers will mark the finish product instead of unfinished paragraphs.

Figure 1 illustrates the DBR model methodology adopted for this study. It has 4 phases that is aimed at refinement of problem, solutions, methods and design principles. Each phase illustrates all the research procedures and instruments involved to suggest the next step for the refinement of AESF.

METHOD

This section illustrates the details in the design and development of AESF based closely to the DBR model.

Analysis of Practical Problems and Practitioners in Collaboration

The close collaboration between the practitioners as main informers, researcher as literature reviewer and mediator, and the technical support team for the technological invention, as shown in Figure 2, have successfully produced the first Automated

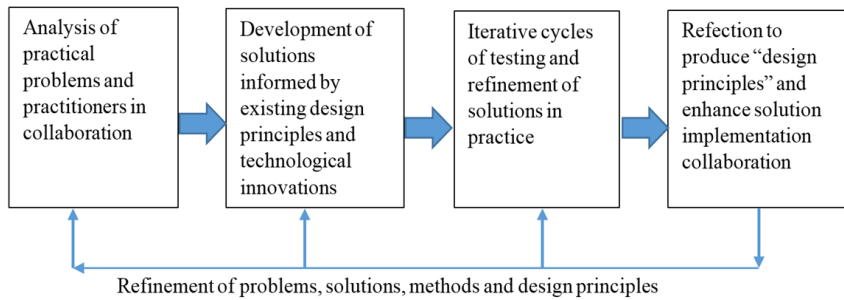


Figure 1. Design research approaches in educational principles (Reeves, 2006)

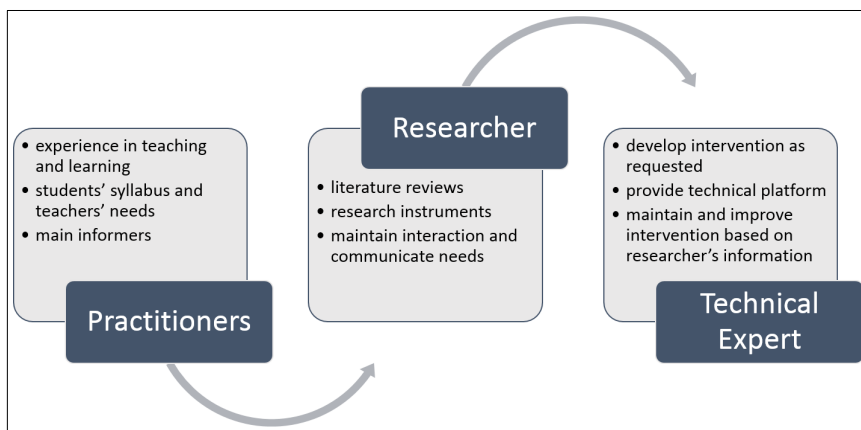


Figure 2. Design research approaches in educational principles

Essay Scorer with Feedback (AESF) prototype.

In this phase, the analysis of practical problems was carried out in eight upper secondary schools, selected purposively where there were Malaysian University English Test (MUET) markers/examiners. By collaborating with these expert markers, the most crucial problem in language teaching and learning, the writing lessons, were scrutinised by means of observation and interviews. Need analysis was done based on three major aspects of content, conduct and context. In terms of content, the writing component is seen as requiring

immediate attention because this component of the language skills is least focused or practised (Weigle, 2007), ironically, the most important medium of communication in the academic arena (Johari, 2004). In terms of conduct, students are normally given an essay question to be attempted, probably after a thorough discussion as practice or as a test without guidance. The essays are eventually collected and marked by the teacher and returned after sometimes. Finally, the context in which this study takes place is the Malaysian upper secondary classroom. In most Malaysian classroom, application of technology is deemed lacking

(Md Yunus, 2007; Yunus et al., 2013). Despite the increasing use of technological innovation in education reported worldwide, the Malaysian schools context are generally lacking in the integration of such advancement (Yunus et al., 2013). They are probably not provided with such facilities or some may have underutilised such provision (Yunus et al., 2013).

In order to suggest a workable solution, eight qualified and experienced MUET examiners, sampled purposively, regarded as subject matter experts were interviewed to identify marking conventions and derive the functionality of AESF based on their perception and expectation of a technological intervention. Analysis of the interviews suggested that AESF had to be reliable, valid in scoring, useful, easy to use, immediate in giving feedback, as well as having easy accessibility (Ng et al., 2015).

Through the findings obtained from literature reviews and practitioners, technological experts in programming and computing were involved to discuss and derived the most possible working and layout of AESF.

Development of Solutions Informed by Existing Design Principles and Technological Innovations

It is necessary to note the development of AESF is based on essays compiled using home grown corpus where essays used for training are essays written by actual students and marked by real expert human grader based on standardised MUET marking scheme, unlike commercially available AES that are trained using first language user that

may not be the same as the local context and the grade provided may not be the same as the MUET marking requirements. In addition, essay topics used to collect gold standard were actual past year MUET questions that were validated. Therefore, AESF resulted from the local context is the solution to assist the actual MUET teaching and learning classrooms.

In this phase, eight schools with 6th form students were approached to collect essays as gold standards. These schools were selected purposively where there were actual experienced examiners trained by the Malaysian Examination Council to ensure validity and reliability in scoring. Apart from being scored by the examiners in school, each essay collected was graded by two other experienced MUET markers hired independently to increase reliability and validity of scoring. A double-blind method was used to overcome biasness and the marks awarded were averaged to calculate the final score obtained by the essays. These scored essays were then typed exactly as written by students into the Notepad programme so that they become machine readable and compiled into a corpus to train the AESF scoring accuracy. Since it is modelled using home-grown corpus it is representative of the cultural and localised marking standard of MUET as mentioned before.

AESF consists of essay management system which enables the user to collect and train essay marking based on the essays feed into the system. Subsequently, from the trained topic, teacher can set task and

monitor progress as well as verify scores generated by AESF.

The core component in AESF is the grading module (GM) which can derive a band between one to six, if given an essay. The GM facilitated the state-of-the-art feature engineering to extract significant indicators in the essay. For the study we facilitated 17 features extracted from essay using natural language processing technologies, which are listed in the followings:

- Total word count
- Unique vocabulary count
- Lexical richness
- Number of sentences
- Average word in a sentence
- Number of spelling error
- Spelling error ratio (against total word count)
- Number of grammatical mistakes
- Grammatical mistakes ratio (against total word count)
- Part of speech count ratio (against total word count)
- Number of high-level part of speech: adverb, adjective, adjective superlative, verb gerund or present participles etc
- Ratio of high-level part of speech (against total word count)
- Number of parameter
- Number of punctuation
- Number of Type 1 conjugate
- Number of Type 2 conjugate
- Number of Type 3 conjugate

We named the above 17 features as surface features, they are used to represent

the essays' technical properties. To score an essay, the features are used to construct scoring model with Support Vector Machine (SVM). The model includes different topics of essay of the gold standard with band one to six. To date, we have constructed a database of about 143 gold standard consisting of essays collected from secondary schools in Sarawak. To provide feedback for each essay, all the features listed above 6-17 were also being used to indicate area of improvement of student essay in according to the gold standard. We also relied on Language Tool, mainly helping identifying spelling error and identify syntactic structure of the sentences. Figure 3 shows a screenshot of the AESF during an essay writing session.

In addition, one of the uniqueness of the AESF over others is it facilitated the state-of-the-art computational semantics technology to detect coherency of the essay. The semantic engine, which was derived from Latent Semantic Analysis (LSA) (Landauer, 2007), is used to compute the semantic similarity within opening, body and closing paragraphs. LSA refers to a "theory and method for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text" (Landauer et al, 1998) while NLP is the use of statistical method by means of annotation of language for analysis (Collobert et al., 2011). The idea for developing AESF originated from the potential seen in general AES to score essays automatically hence could be combined with feedback to be utilised for classroom teaching and learning purposes (Warschauer & Grimes, 2008). If the AES system can

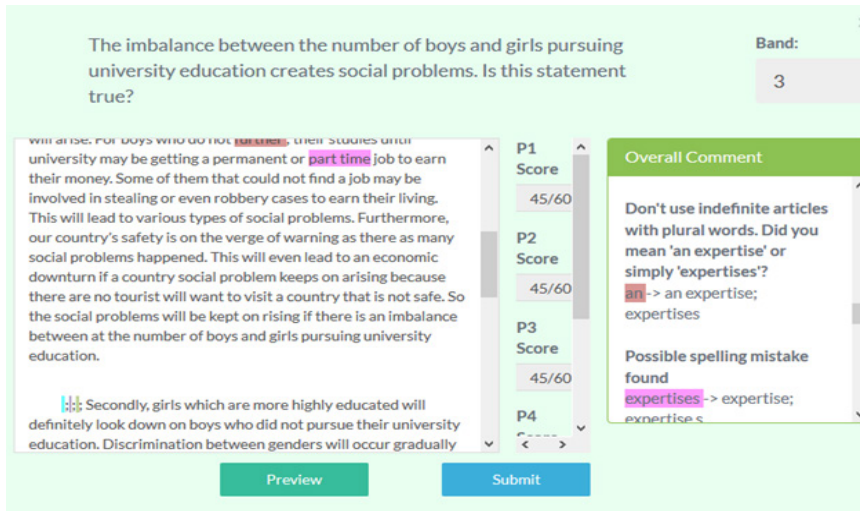


Figure 3. Screen snapshot of the AESF

score essays reliably and validly, it can be used to assist students in the writing process as an indicator on how well their essays are written and to ease teachers' marking process because the essays are being pre-scored and fundamental language errors are being eliminated by the students based on the automated language feedback provided. Thus, AESF is believed to be the solution to the common problem faced in writing and marking essays. The measure is used to detect at paragraph level if the context of the different paragraphs is coherence within the essay and between the gold standard, even if different words are used in the sentences. Lastly, the matrix is not only used for band, but displays it on at the essay editor as a feedback indicator.

For our study, the essay band is predicted through averaging 17 surface features score from the SVM. Our preliminary result has shown that the method could predict the grade correctly (compared to human graders) at accuracy of 75.7%.

Overall, we want to realize an essay management system which is not only able to grade the essay, but also provide constructive feedbacks such as spelling mistakes, incorrect use of punctuation, sentence syntactic structure as well as coherency within an essay.

Iterative Cycles of Testing and Refinement of Solutions in Practise

The first prototype started with a mock design/story board that was presented by the researcher to the technical expert so that the exact need and features could be communicated and implemented. The system has a login page which is shown in Figure 4 and a student page that allows individual students to login and write their essay in either the full essay or the paragraph by paragraph framework based on only one topic.

This prototype was tested by 15 foundation year students at a public

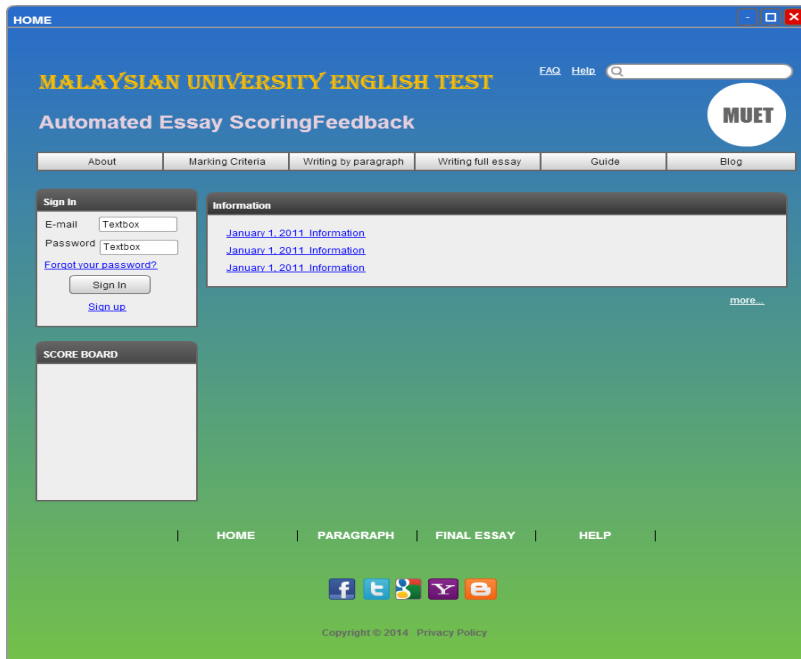


Figure 4. Login page of prototype 1

university who had experienced and sat for MUET before so that they could provide more relevant comments for the next iteration. These students were randomly selected using random number method from the Faculty of Cognitive Sciences and Human Development. Simple elicitation of ideas was conducted verbally because the main reason for this initial testing was for ensuring the stability, authentication as well as a general feel of using AESF. From this first iteration, the issue about authentication and possible login errors were identified. The stability of the system was determined, and the accuracy of the scoring engine was tested.

The outcome of iteration one led to a serious discussion between the researcher and the technical experts to produce the second prototype. This prototype had

added a teacher module to allow teachers to be in control and in charge of their students writing progress and provided teachers with the final decision in approving students' achievement. Besides, it could score more accurately with more detailed annotation and emphasis on discourse markers and with an improved user interface that is straightforward and a welcoming note as shown in Figure 5, minus all the authentication errors identified earlier.

This stage was tested by the same eight expert markers who had contributed in the earlier need analysis stage. Their user experiences were recorded and potential areas for improvement were derived.

This iteration identified some failures when the experts used mobile devices to access the system. The initial design was not

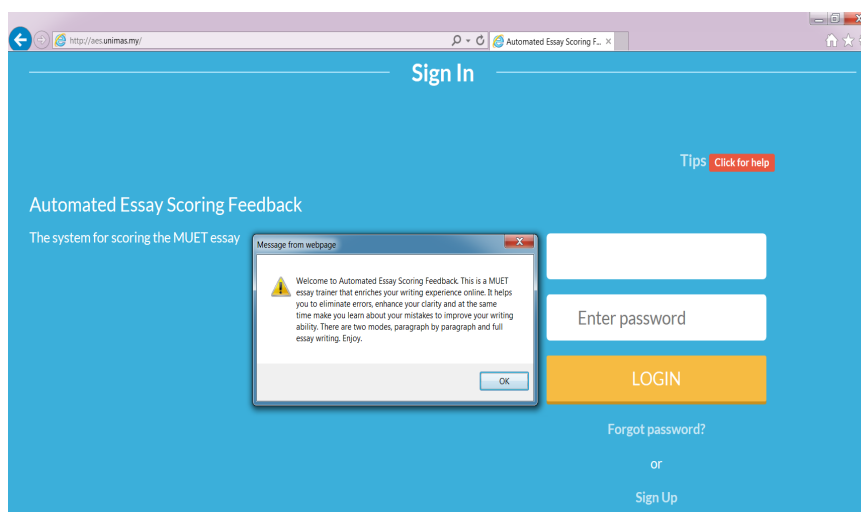


Figure 5. AESF login page

mobile friendly and meant for desktop view. When mobile devices were used, part of the layout went missing and this caused users to feel uneasy. Since most people would use mobile devices like smart phone, phablet and tablet to go online, it is necessary to improve on this area for the next prototype.

In addition, teachers also complained that one topic was not enough, and they would like to have more control like setting multiple tasks and for multiple classes and training more topics to enrich the system.

Reflection to Produce “design principles” and Enhance Solution

In the third iteration, the third prototype witnessed a more complete system with two modules as shown in Figure 6, the teacher module and student module where they could just register themselves by indicating their own status i.e., teacher or student, and start using the system.

The third prototype was used in a pilot study to identify any other possible problems when the system was implemented in actual classroom context with a myriad of users (teachers and students) who might or might not be interested in using the system. This pilot study in the DBR phase was carried out in an upper secondary school with 80 students and four teachers using a quasi-experimental research design where division of classes was set by the school which ranges from 10 to 24 students in each class, students were randomly put in two groups; the controlled and the experimental group. This school was selected because four out of six MUET teachers in this school had experiences in the actual marking of MUET essays, trained by the Malaysian Examinations Council.

A pre-test was conducted to gauge students’ initial level in essay writing for both the controlled and experimental group.

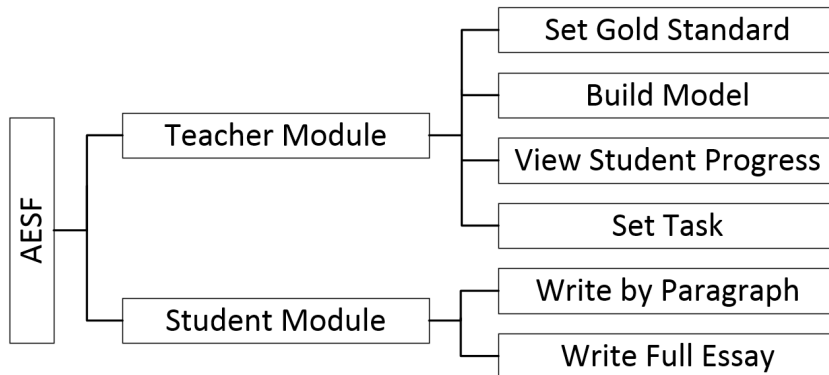


Figure 6. The modules in AESF

The pre-test was an open-ended opinion based essay derived from past year paper to ensure validity and reliability of the test. This was done using the conventional paper and pen method to provide equal grounds for both groups. Then, a common writing lesson was carried out based on the topic in AESF where half of the class (randomly assigned) used AESF and the others used the conventional paper and pen method to attempt the essays. This process was repeated for both essay topics with an interval of two weeks. A post-test, which was the same as the pre-test was conducted to measure students' achievement. The exact essay will allow students to write on similar topic area, but the structuring of the essay would be different, and their language achievement can be assessed.

The pilot test also included a questionnaire combining Instructional Material Motivation Survey (IMMS) developed by Keller (2010) to measure student's motivation level towards using AESF and a Technology Acceptance Model (TAM) survey developed by Davis and Venkatesh (1996) to measure students'

acceptance of AESF. These questionnaires have obtained written approval from the copyrighted owners. The findings of the questionnaire, as shown in Table 1, indicated that the Cronbach alpha coefficient calculated using SPSS showed that the questionnaire was reliable with overall alpha values of 0.94, exceeding 0.7 the criteria seen as appropriate for research purposes (Nunnally, 1978).

This pilot test had ensured that the questionnaire and the general working of AESF were well accepted by teachers and students while on the other hand, some practical problems surfaced. With mass usage, the server could not support the large amount of data entered and this led to server down or loss of link while students were using the system or when students tried to logon to the page. This caused frustration in users when they could not achieve what they wanted. In addition, this phase also detected other possible drawbacks such as students pretending to be teachers and went into the teacher's module to mess up setting of task and other possible functions in the teacher's module.

Table 1
Cronbach alpha coefficients for motivation and acceptance scale

Scale	Number of Items	Cronbach Alpha ($n = 41$)
Attention	11	0.73
Relevance	8	0.77
Confidence	9	0.63
Satisfaction	6	0.74
Total Overall Motivation	34	0.91
Perceived usefulness	6	0.87
Perceived ease of use	6	0.89
Total Overall acceptance	12	0.92
Overall Questionnaire	46	0.94



Figure 7. Teacher module

In line with the findings from the pilot study, it was indicated that AESF had to go back to phase two of the DBR to redesign and restructure the working and functionality of AESF to ensure smoother administration in context.

The fourth prototype had an improved server capacity and classified user permission so that ‘super user’ or the administrator can monitor all progress made by teachers and students. While teachers could build grading database, train new essays, set multiple tasks, view students’ progress, and remark students’ essays as shown in Figure 7.

Students on the other hand could write essay in the paragraph mode or full essay mode (refer Figure 8) where they could keep getting the system to provide marks and comments about their essays over and over again before they submitted their essays online to their teachers.

Teachers registered themselves using a special link given to them while students registered themselves online using their valid email addresses following a simple instruction as shown in Figure 9. This overcomes problems identified in the testing of prototype 3.



Figure 8. Student module

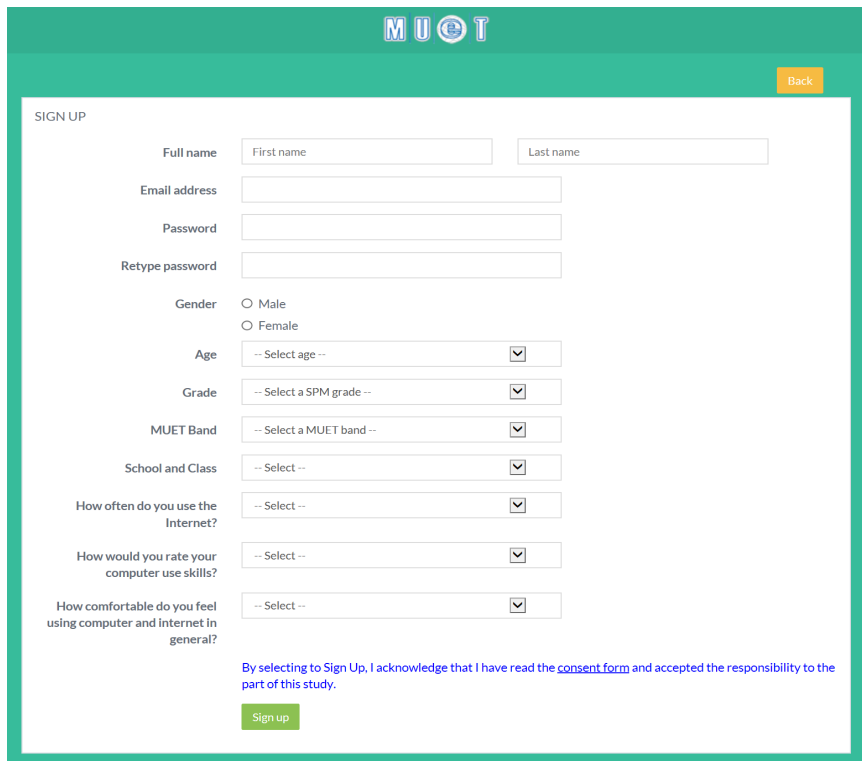


Figure 9. Student registration page

The fourth prototype was tested by 40 computer science students from a public higher education institute to check the performance and stability of the server when the 40 students logged on, used and submitted essays at the same time. This

group of students was selected because computer science students were more aware of the possible problems and ways to crack or hack the system in order to minimise problem of adventurous students in the actual testing ground. This phase

observed minimal challenge, a handful emails received to seek clarification ranging from how register to how to use the system. This iteration had indicated that the system can go through another round of phase 4 DBR in actual teaching and learning context if more guidance were provided.

The fifth prototype included a video walkabout to assist students on how to proceed with written instruction. They can click on the 'Tips button' on the login page if they required additional instructions. Moreover, this video guide as shown in Figure 10 can serve as an introduction to students before they register and logon to use the system.

This prototype was finally introduced to five upper secondary schools involving 24 MUET teachers and a total of 400 students. These were the same five out of eight schools where gold standards were collected but with a different batch of students. Human factor was identified as a vital obstacle in conducting research in real life

context because not all teachers and students were willing to take up the extra effort and time to try out the prototype. Many indicated their willingness to take part but eventually pulled out citing various excuses. Hence, this cycle only secured the commitment and participation of 18 teachers and a total of 300 students where 150 students used the system and 150 in the controlled group with the quasi experimental research design method applied. The other steps taken were the same as testing of the third prototype during the pilot testing.

Moreover, the initial planned time was fixed to be 2 months where each essay from pre-test, topic 1, topic 2 to post test, would have an interval of 2 weeks. However, all schools did not complete the tasks in time due to various unforeseen circumstances like national level exam, school level exams, sports day, celebrations, and teachers or students on emergency sick leave.

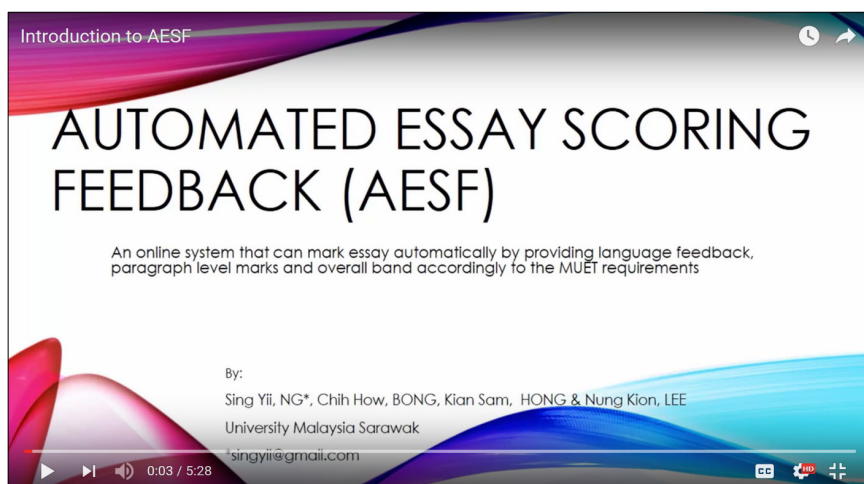


Figure 10. ASEF video guide

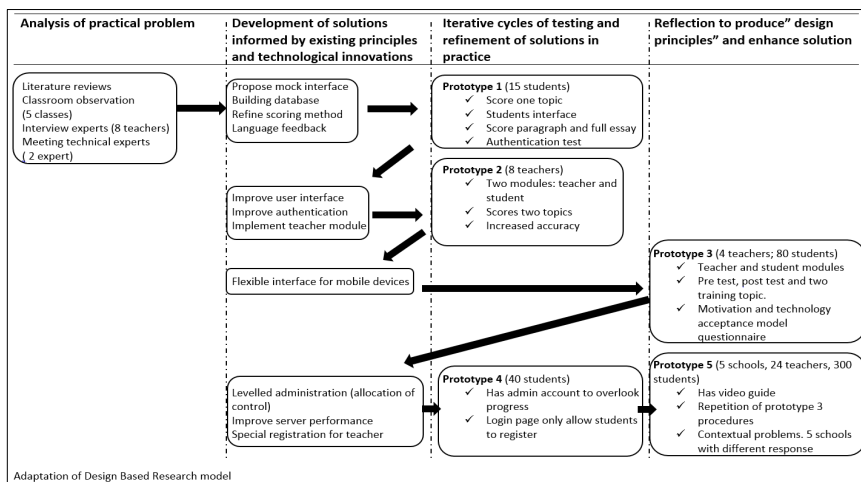


Figure 11. The design and development of AESF based on DBR model

RESULTS AND DISCUSSIONS

With the DBR approach used in suggesting, designing, developing and evaluating AESF as a technological intervention, it has allowed various iterations of refinement from scratch to the final workable AESF in classroom context. An effort made possible by the close cooperation between practitioners, researcher and technical experts to solve, not completely, but assist the most difficult problem in language teaching. The steps and iterations taken in realising AESF based closely on DBR is illustrated in Figure 11.

It can be derived that the process involved in producing the first to currently the fifth prototype is an uphill effort and is still subject to further exploration. DBR allows continuous improvement and introduction of new method, innovation and intervention to take place as recommended by users and experts. It is not as rigid as other research approaches that follow

strictly outlined hypothesis or framework. However, due to time constraint and limited resources, the achievement at the fifth prototype is considered sufficient to show the applicability of such an intervention in the writing classroom.

A quick review from Figure 12 indicated that all teachers accepted AESF intervention in their classroom though some faced problem implementing it, namely no internet access, no time to learn new system, students too busy with assignments and chasing syllabus. A teacher commented that if this system was introduced to them before the semester year started, then it would be easier for them to make way for it.

On the other hand, deriving from the TAM survey, student who used the system for essay writing perceived AESF as useful and easy to use. From Figure 13, most of the students scored the system positively, 3 and above suggesting that AESF could help them in their essay writing and operating it reasonably easy.

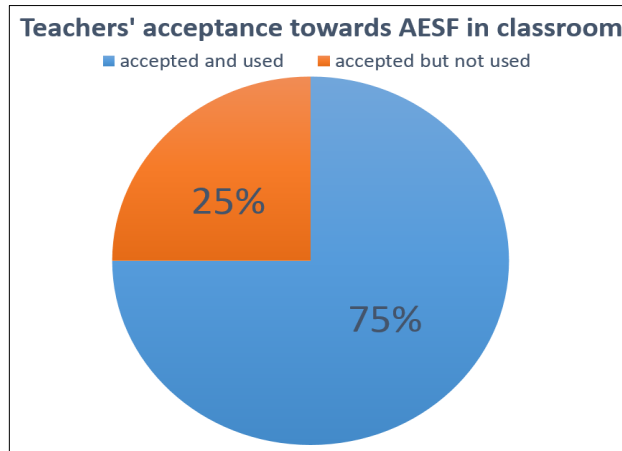


Figure 12. Teachers' acceptance towards AESF

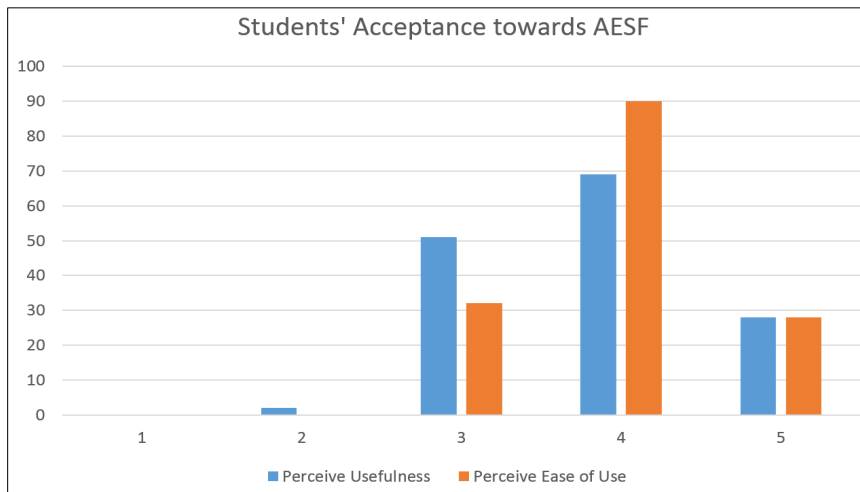


Figure 13. Students' acceptance towards AESF

The system is seen as accepted and is beneficial for both teachers and students despite some additional constructive criticisms that are useful for future iterations. One teacher particularly, confessed that she was initially sceptical about AESF but after she went through the system with her students, they were amazed by what the system can provide.

“...honestly, I was sceptical but after my students use it, I find it amazingly good. Well done.” (Teacher 1)

This teacher even made use of the system for her students to check their other English based subjects, for the system to comment and correct their language command. This kind of remarks definite assured the DBR researcher to further

enhance and improve the system to benefit real classroom.

On the negative note, the system is still limited to only two topics which is insufficient to be used extensively. To have more topics will demand too much time for a teacher to collect, score, digitise and train the system. It is suggested that teachers who are in the same area can cooperate to build on the database to score a wider repertoire of topic. Some teacher also requested that the system provide word count and have option for reading model essays or link to search engine, so students can look up for more resources to craft their writings. These are grass root request that aims at solving real world problems.

CONCLUSION

In conclusion, DBR is an approach that allows every change in the development and evaluation of an intervention to be documented and scrutinised in accordance to the actual learning context where realistic measures can be taken to make it workable as it intended to. The process is laborious and requires flexibility in searching for the right method by intertwining the qualitative and quantitative method in data collection and analysis to seek solution. It is a difficult and time-consuming process, but the outcome is rewarding and fulfilling. AEF may not be the perfect system because it involves too much time to get enough gold standards to train a topic and it depends fully on the internet connection and server performance to be workable, but this effort gives hope to language teachers and

students to tackle the writing component more confidently. AESF allows teacher to train new topic for marking, set task, check progress and certify score generated while for students, this system provides students the platform to practise writing and improve their skills independently. At present the system can grade and provide on-going feedback to users based on two well-trained topics while on untrained topic AESF can still provide grammatical feedback. It is planned to expand the marking repertoire to more topics by getting more students and markers to help expand the corpus for the benefits of all. Our study shows that automated essay scoring using artificial intelligence techniques is a practical and feasible method to score MUET essays.

ACKNOWLEDGEMENT

We would like to thank the teachers and students who had participated in the evaluation of the system and provided valuable feedbacks.

REFERENCES

- Ahmad, M. I., & ul Hussain, S. R. (2012, December). Feedback on students' essays assignments: An appraisal. *Language in India*, 12(12), 471-484.
- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education? *Educational Researcher*, 41(1), 16-25. <https://doi.org/10.3102/0013189X11428813>
- Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, 13(1), 1. https://doi.org/10.1207/s15327809jls1301_1
- Bell, P. (2004). On the theoretical breadth of design-based research in education. *Educational*

- Psychologist*. https://doi.org/10.1207/s15326985ep3904_6
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, *12*, 2493-2537.
- Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: Three experiments. *International Journal of Human-Computer Studies*, *45*(1), 19-45. <https://doi.org/10.1006/ijhc.1996.0040>
- Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing*, *26*(1), 1-15. <https://doi.org/10.1007/s11145-012-9395-2>
- Johari, S. K. (2004). *Second language writing: An investigation into the writing skills of Malaysian primary school pupils*. Kota Kinabalu, Malaysia: Centre for the Promotion of Knowledge and Language Learning, Universiti Malaysia Sabah.
- Joseph, D. (2004). The practice of design-based research: Uncovering the interplay between design, research, and the real-world context. *Educational Psychologist*, *39*(4), 235-242. https://doi.org/10.1207/s15326985ep3904_5
- Keller, J. M. (2010). Motivational design research and development. In *Motivational Design for Learning and Performance* (pp. 297-323). USA: Springer US.
- Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research*, *42*(2), 173-196.
- Kennedy-Clark, S. (2013). Research by design: Design-based research and the higher degree research student. *Journal of Learning Design*, *6*(2).
- Landauer, T. K. (Ed.). (2007). *Handbook of latent semantic analysis*. Mahwah, USA: Lawrence Erlbaum Associates.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259-284. <https://doi.org/10.1080/01638539809545028>
- Lipnevich, A. A., & Smith, J. K. (2009). "I really need feedback to learn:" Students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability*, *21*(4), 347-367. <https://doi.org/10.1007/s11092-009-9082-2>
- Md Yunus, M. (2007). Malaysian ESL teachers' use of ICT in their classrooms: Expectations and realities. *ReCALL*, *19*(01), 79-95. <https://doi.org/10.1017/S0958344007000614>
- Mitchell, R. (2013). *Second language learning theories* (3rd ed.). New York, USA: Routledge.
- Ng, S. Y., Hong, K. S., Lee, N. K., & Bong, C. H. (2015). Teachers' expectation and acceptance of a home-grown networked Automated Essay Scorer with Feedback (AESF) for the Malaysian University English Test (MUET) Classroom. *MELTA 2015*. Kuala Lumpur, Malaysia.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). USA: McGraw-Hill.
- Reeves, T. C. (2006). Design research from a technology perspective. *Educational Design Research*, *1*(3), 52-66.
- Reeves, T. C., & McKenney, S. (2015). Design-based research. In *The SAGE Encyclopedia of Educational Technology* (pp. 189-191). California, USA: SAGE Publications, Inc.
- The Design-Based Research collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5-8. <https://doi.org/10.3102/0013189X032001005>

- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies*, 3(1), 22-36. <https://doi.org/10.1080/15544800701771580>
- Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, 16(3), 194-209. <https://doi.org/10.1016/j.jslw.2007.07.004>
- Yunus, M. M., Nordin, N., Salehi, H., Embi, M. A., & Mahamod, Z. (2013). Managing problems and planning activities involving ICT tools in teaching ESL reading and writing. *Asian Social Science*, 9(10), 222-230. <https://doi.org/10.5539/ass.v9n10p222>