COGNITIVE COMPETENCY ASSESSMENT FOR SHORT FREE TEXT ANSWER VIA HYBRID APPROACHES: *INITIAL FINDINGS*

Anis Fadzilah Binti Mohd Nor@Abdullah[1], Nurul Amirah Jamaluddin[1], Adidah Lajis[1], Haidawati Md Nasir[1], Normaziah Abdul Aziz[2]

[1] Computer Engineering Section, Universiti Kuala Lumpur Malaysian Institute of Information Technology, Kuala Lumpur, Malaysia,
[2] Computer Science Department, International Islamic UniversityMalaysia,
For Correspondence; naa@iium.edu.my anisfadzilah@gmail.com, nurulamirahjamaluddin@gmail.com, adidahl@unikl.edu.my, haidawati@unikl.edu.my, naa@iium.edu.my

ABSTRACT: *The emergence of outcome based education and the increase of number of student in higher education have posed a need of assisted competency based assessment. Most of the assisted assessment available are focusing on performance-based approach. A new hybrid approach has been developed via a combination of Natural Language Processing, Neural Net and Information Theory and has produced an encouraging result.*

1. INTRODUCTION

The increased number of student in Higher Education may also increase the time spent by the lecturer in marking the assessment [1]. According to statistical report from Ministry of Higher Education Malaysia (2010), the population of student has been raised yearly. Assessment is an important process in any educational organization. Assessment is defined as the process of making determination about one individual attitudes, skills and knowledge through various methods such as interview, observation, projects, exam with prior predefined assessment criteria [2].

The emergence of outcome based education (OBE) and competency based assessment (CBE) in higher institution required the proper mixture of learning objective and competencies to be achieved by the student. In this type of assessment, at the end of the learning, student must be able to achieve certain outcome and the competencies set will say how we can certain that the student know it. The evidence is produced whether the student is competent to a certain standard and according to Moris, this kind of assessment approach is different from performance-based assessment [3]. The competency evidence can be get from the, report evaluation, test result, licenses or certificates [4].

Competencies are the knowledge, skills, abilities and behaviors that help and required to the successful performance of the job. Bloom Taxonomy has been applied to categorize competency pieces namely; cognitive, psychomotor and affective. In cognitive domain, Bloom has categorized these into 2 main levels that is low level consists of remembering, understanding, and applying and higher level consists of analyzing, evaluating and creating. Higher level cognitive has been assessed via short free text. As according to Carter et al. [5], 81% of exam is a closed book examination where 57% on application test and 37% on evaluation test. However, grading open ended questions by hand can be time consuming.

There are numbers of assessment systems have been developed which focus to reduce time of marking especially essay and short answers There are also much disagreement on the techniques apply in assisted assessment. According to Wiemer-Hastings et. al. [6] Latent Semantic Analysis (LSA) is focusing on content and mostly used to assess humanities essay and the performance is poor in the domain of casual domains such as research methods [7]. Callear et al. [8] also argued that due to the importance of word order in short free text, the application of LSA is not suitable. In fact, most of the assisted assessment are focusing on performance instead of competency [9]. The increased adoption of CBE and the increase of number of students in HE have pose the need to produce competency assisted assessment. This paper is proposing a technique which can be employed to assess the higher level of cognitive competency focusing on short free text answer.

2. OUTCOME BASE EDUCATION AND BLOOM TAXONOMY

Outcomes based education (OBE) is a process that involves the curriculum restructuring. In OBE, rather than accumulating the course credits, the assessment and reporting practices in education have to reflect the achievement of high order learning [10]. It is defined as a "…comprehensive approach to organizing and operating an education system that is focused in and defined by the successful demonstrations of learning sought from each student" [11]. In Malaysia, it has been in practiced since 1950s and in being implemented at all levels of education tertiary education in 2008 [12]. OBE be implemented in various modalities and the implementation is covering all three learning domains that are effective, cognitive and psychomotor. Cognitive outcomes include demonstrable acquisition of specific knowledge and skills. Affective educational outcomes, defined as learning outcomes that focus on "individual disposition, willingness, preferences, enjoyments …..." [13] can be reintegrated as a critical focus during this restructuring. Evidence of the outcome is required to fulfil the shortage of the soft skill of an employee in the workplace [14-15].

The Bloom's Taxonomy of Educational Objectives cognitive domain has influenced many educationists long years ago [16]. It has been proved to be useful for analyses of cognitive demand from the constructing curricula to the assessment of student performance. Table I, the Blooms' six cognitive skills, the questions cues, and also the skills demonstrated are shown. The questions cues are adapted from [17-18].

Table 1: **Cognitive Level of the Revised Taxonomy [19]**

Cognitive Level	Cognitive Task	Description
Remember	Recognizing Recalling	Retrieving relevant knowledge from long-term memory.
Understand	Interpreting Exemplifying Classifying Summarizing Inferring Comparing Explaining	Determining the meaning of instructional messages, including oral, written, and graphic communication.
Apply	Executing Implementing	Carrying out or using a procedure in a given situation
Analyze	Differentiating Organizing Attributing	Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose.
Evaluate	Checking Critiquing	Making judgments based on criteria and standards.
Create	Generating Planning Producing	Putting elements together to form a novel, coherent whole or make an original product.

3. REVIEWS OF TECHNIQUES IN ASSISTED ASSESSMENT

There are various techniques have been invented in order to produce better assisted assessment for essay and short free text marking. For example, LSA, Probabilistic Latent Semantic Analysis (PLSA), Natural Language Processing (NLP), Vector Space Matrix (VSM) and Pattern Matching and many others. However, most of these assisted assessments were unable to assess competency-based of the student. Table 2, summarizes the existing assisted assessment and the techniques employed and also the assessment focus.

4. NLP

NLP is focus on the development and improvement in the process of learning. It is effectively applied in the education for promoting the language learning and enhancing the academic performance of the students. NLP able to improve learning via the assistance in the process of using computer and internet. There are numbers of assisted assessment that employ NLP in their assessment method [20-23].

Table 2: Summarised Techniques and Assessment Focus for Assisted Assessment for Short Free Text Answer (C-Competency, NM-Not mentioned)

Researchers	Subjects	Level / Age	Sentences	Marks	Techniques	Focus
[20]	Science, Computer Science	Undergraduates and Schools	3-150 words	0-6 point scales	NLP and Information Theory	C
[30-33]	GCSE Biology Examination	14-16 years old	Up to 5 lines	0-2	IE & Computational Linguistic	NM
[34]	Artificial Intelligence	Undergraduate students	Up to 6 lines	NM	Clustering	C
[35-36]	Reading Comprehension	7th and 8th Graders	Up to 100 words	0-2	NLP Rule-based algorithm	NM
[37]	1999 Science National Test Paper A and B	11 years old	Up to short explanatory sentence	0-2	NLP and IE	NM
[38-39]	Introductory course in Computer Literacy	Undergraduate Students	Short answer	NM	LSA and NLU	NM
[40]	Assessment of summaries based on reading comprehension	Undergraduate students	75-100 long	NM	Computational Linguistic	NM
[41]	Summary writing for reading comprehension	6th to 9th graders	4 sentences	0-4	NLP	C
[42-44]	Introductory data structure course	Undergraduate students	Up to 2 lines	0-5	Text Similarity & Corpus based similarity	NM
[45]	87 questions on Object Oriented Programming	Undergraduate Students	Up to 2 lines		NLP and QAML	NM
[46]	High School physics	Undergraduate Students	Up to 2 lines	4 point scale	Statistical Classifier and Rule-based	NM

Convolutional Neural Network

Convolution is an important operation in signal and image processing. Convolutional neural networks (CNNs) are neural networks that make use of the internal structure of data such as the 2D structure of image data through convolution layers [24], where each computation unit responds to a small region of input data. CNN has been applied on text for entity search, tagging, sentence modelling, and others and has been mentioned in [25]. The essence of CNN is to convert small region of data to feature vectors to be used in the computation of student known information.

Information Theory

Knowledge has been viewed as enactment of knowing [26] where its forms a context use of information. And information is an input as in this research is the answer script of the learner and expert. The amount of this information, as in this research will be indicated as known information may be measured as with Shannon's model as logarithm [27] of the inverse of the probability of the state of nature found at the output of the process of learning. Natural language itself shows a balance between order and randomness [28]. Excess entropy has been used to serves as a general purpose measure of a system's structure, regularity, or memory and it provide provides a quantitative measure of structure that may be applied to any one-dimensional symbolic string as mentioned in [29]. Excess entropy also has been used to quantify the amount of known information of the learners [20]. Table 2 summarized the computer assisted assessment that focus on short free text answer and the focus of assessment; performance or competency.

5. PROPOSED HYBRID TECHNIQUES

NCI (NLP, CNN-seq and Information theory) is the hybrid approach consists of a combination of NLP, seq-CNN and Information Theory. The NLP is used to train the data prior undergo the process of CNN to get the pattern and the Information Theory is to quantify the amount of information contain in the text. Figure 1 shows the overview of the proposed techniques.

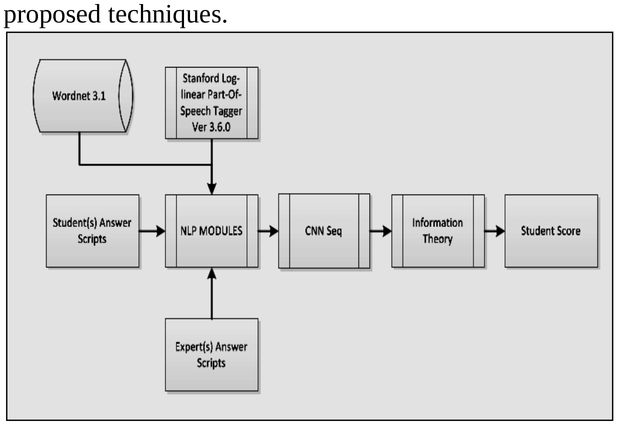

Fig. 1: Overview of NCI

NLP Modules

As shown in Figure 1, both learner and expert answer script will undergo NLP modules. There are 4 processes in this module, they are:

a) Part of Speech Tagging
All text will be tagged by using Stanford Log-linear Part-Of-Speech Tagger version 3.6.0.

b) Term Extraction
After that only noun, verb, adverb and adjective are extracted without ignoring the word order. However, common conjunctive adverbs will not be extract because it is used only to join two clauses together.

c) Term Stemming
Stemming is the process of derived words to their word stem, base or root form. WordNet 3.1 will be used in this process.

d) Term Synonym
The synonym will be identified by using WordNet 3.1.

Seq-CNN

Suppose that the Student Answer Script, $\boldsymbol{S}= (w_1, w_2,\ldots,w_n)$ with Expert Script as a Dictionary $\boldsymbol{D}$. A word is treat as a pixel, treat $\boldsymbol{S}$ as if it is an image of $|\boldsymbol{S}|$ x 1 pixels with $|\boldsymbol{D}|$ channels and to represent each word as a $|\boldsymbol{D}|$-dimensional one-hot vector. In this research, since we are focusing on short free text, thus the region size is 2 and stride is 1. Figure 2 shows the example.

The extracted Expert Answer: *smooth surface further object travel*

The extracted Student Answer: *surface smooth distance travel car far*

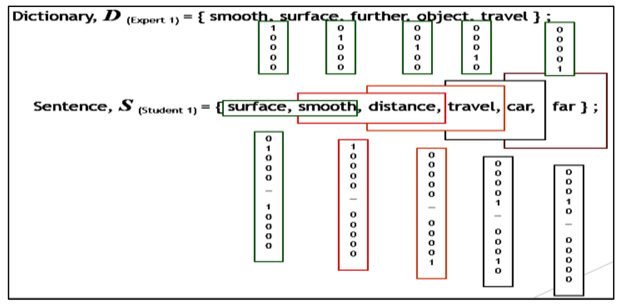

Fig. 2: seq-CNN for Extracted Word

Information Theory

Excess entropy will be employed to quantify the amount of Known-information (K).

$$K = \sum_{i=1}^{m} H(x_i) - H(x_1,....,x_m) \qquad (1)$$

Where: $H(x_i,....,x_m) = -\sum p(x_1,....,x_m) \log p(x_1,....,x_m)$ (2)

Summed over all combinations of values of [48].

For example, the vector of the Expert is shown below in Figure 3 and the amount of Known-information of an Expert is 2.32 bits.

Vector	1 0 0 0 0 — 0 1 0 0 0	0 1 0 0 0 — 0 0 1 0 0	0 0 1 0 0 — 0 0 0 1 0	0 0 0 1 0 — 0 0 0 0 1
Probability, p	0.25	0.25	0.25	0.25

Fig. 3: Example of Vector of an Expert

Score Computation

If there are more than 1 expert answer available, the $\boldsymbol{K_S}$ will be compared against all the $\boldsymbol{K_E}$ and the nearest $\boldsymbol{K_E}$ will be used as a referent expert model and if there is a case where the amount of $\boldsymbol{K_S}$ is more than the amount of $\boldsymbol{K_E}$, the amount of $\boldsymbol{K_S}$ will be assigned to equivalent to $\boldsymbol{K_E}$. The formula of computation of the score is as below [20];

$$Score = \frac{K_S}{K_E} \times mark \qquad (3)$$

6. RESULTS AND DISCUSSION

Table 3 shows sample data used to test NCI. There are two expert scripts and 4 student answer scripts has been used in the testing. Table 4 shows the result of NCI. As shown, K_{S1} is having amount of K of 2.32 for both expert script. However, K_{S1} is more than K_{E1} and equal to K_{E2}. Therefore, the score will be based on K_{E2}. And for K_{S2} is having K as 1.58 bits for both expert script comparisons. Therefore, the score will be based on the lowest differences between K_E and K_S, that is K_{E1} and the score is 0.80. And for K_{S3}, the K differences is lowest with K_{E2} and therefore, the score is computed against K_{E2} that is 1. And same with K_{S4} where the lowest differences are with K_{E2} and the score is computed against K_{E2}.

Table 3: Sample Data.

Category	Answer Script
Expert 1	The smoother the surface, the further the object
Expert 2	The greater the friction of a surface, the shorter the object travel
Student 1	If surface is smooth, the distance travelled by car will be farer
Student 2	A smooth surface has less friction
Student 3	The smoother the surface, it has less friction
Student 4	The more friction, there are the less distance travelled by the car

Table 4: NCI Initial Result

Category	K_E	K_{S1}	Score	K_{S2}	Score	K_{S3}	Score	K_{S4}	Score
Expert 1	2.00	**2.32**	1	1.58	**0.80**	1.58		1.37	
Expert 2	2.32	**2.32**	1	1.58		2.00	**1.00**	1.86	**0.90**

7. CONCLUSION

All the result is check against human-rater and the result is encouraging. In order to validate the techniques, more data will be collected which focus on the three highest Bloom cognitive level form tertiary program from various domain such as Computer Science, Engineering, Business and Management and Information Technology.

8. ACKNOWLEDGEMENTS

This research is supported by the Fundamental Research Grants Scheme (FRGS/1/2015/ICT02/UNIKL/02/2) financed by Ministry of Higher Education Malaysia.

9. REFERENCES

[1] Stephens D., Bull, J. E. and Wade, W. Computer-assisted assessment: suggested guideline for an institutional strategy. Online Learning. Part V: Online Assessment. London: Sage (2011).

[2] Pepper, D. Thematic Working Group 'Assessment of Key Competences': Literature review, Glossary and examples. Brussels: European Commission Directorate - General for Education and Culture (2012).

[3] Moris, A. H. A new Paradigm in Assessment: Competency Base. Presented in Education Conference 2013, April 24 (2013).

[4] Sitthisak, O., Gilbert, L. and Davis, H. C.,. Towards a competency model for adaptive assessment to support lifelong learning. TENCompetence Workshop on Service Oriented Approaches and Lifelong Competence Development Infrastructures, Manchester, UK (2007).

[5] Carter, J., AlaMutka, K., Fuller, U., Dick, M., English, J., Fone, W. & Sheard, J. How shall we assess this?. In Proceeding ITiCSEWGR '03 Working group reports from ITiCSE on Innovation and Technology in computer science education, pp. 107 – 123 (2003).

[6] Wiemer-Hastings, A. C. Graesser, D. Harter, and Tutoring Research Group. . The foundation and architecture of autotutor. In Proc. 4th Int. Conf. on Intelligent Tutoring Systems, Berlin. Springer-Verlag (1998).

[7] Malatesta, P. Wiemer-Hastings,and J.Robertson. Beyond the short answer question with research methods tutor . In Proceedings of the Intelligent Tutoring Systems Conference (2002).

[8] Callear, D., JerramsSmith, J., Soh, V. CAA of short nonMCQ answers. In: Proceedings of the 5th International CAA Conference, Loughborough, UK (2001).

[9] Adidah & Normaziah. Competency Assessment of Short Free Text Answer, Proceeding ICE2T '14 Proceedings of International Conference on Engineering Technologies and Entrepreneurship, August (2014).

[10] Tucker B. Literature Review: Outcomes - focused Education in Universities. Learning Support Network, Curtin University of Technology (2004).

[11] Spady W. G. Outcome-based education: Critical issues and answers. ArlingtonVA: American Association of School Administrators (1994).

[12] Mohayidin, Mohd Ghazali . "Implementation of Outcome-Based Education in Universiti Putra Malaysia: A Focus on Students' Learning Outcomes". International Education Studies. 1 (4). Retrieved 23 October, (2014).

[13] Gronlund, N.E. How to write and use instructional objectives. Toronto: Prentice-Hall (2000).

[14] Clark, D. Softskills and E-learning. London: Epic Performance Improvement Limited (2005).

[15] MacLeod, A. The importance of soft skills in the current Canadian labour market. Sectoral and Occupational Studies Division of Human Resources Development Canada, April (2000).

[16] Krathwohl, D. R., Bloom, B. S. & Masia, B. B. Taxonomy of Educational Objectives. [Handbook 2: Affective Domain]. London: Longman (1964).

[17] Bishop, P. Report on Assessment for a purpose workshop. MSOR Connections, 2(3), 54 (2002).

[18] Paterson, J. S. (2002, July) Linking online assessment in mathematics to cognitive skills, Paper presented at the 6th International Computer Assisted Assessment Conference organized by Loughborough University

[19] Krathwohl, D.R. A revision of Bloom's taxonomy: an overview – Benjamin S. Bloom, University of Chicago (Online)() Avalaible at

http://www.findarticles.com/p/articles/mi_m0NQM/is_4_41/ai_94872707/print (2003).

[20] Adidah & Normaziah. NL Scoring and Bloom Competency Test: An Experimental Result, · Proceeding ICUIMC '12 Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication (2012).

[21] Siddiqi, Raheel, Christopher J Harrison, and Rosheena Siddiqi. "Improving Teaching And Learning Through Automated Short-Answer Marking". IEEE Trans. Learning Technol. 3.3 (2010): 237-249. Web.

[22] Nitin Madnani, Jill Burstein, John Sabatini, and Tenaha O'Reilly.. Automated scoring of a summary writing task designed to measure reading comprehension. In Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications (BEA), pages 163–168, Atlanta, Georgia, June. Association for Computational Linguistics (2013).

[23] Sukkarieh, J. Z., and Blackmore, J. c-rater: Automatic Content Scoring for Short Constructed Responses. Proceedings of the 22nd International Conference for the Florida Artificial Intelligence Research Sciety, Florida, USA (2009).

[24] Yann LeCun, Le on Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In Proceedings of the IEEE , pages 2278–2324 (1986).

[25] Johnson, Rie and Zhang, Tong. Effective use of word order for text categorization with convolutional neural networks. In NAACL HLT (2015).

[26] Cook, S.D.N. & Seely Brown, J. "Bridging epistemologies: the generative dance between organizational knowledge and organizational knowing", Organization Science, 10.(4), 381-400 (1999).

[27] Hartley, R. V. L. Transmission of Information. Bell System Technical Journal, 7, 535 – 563 (1928).

[28] Crutchfield, J. P. The Calculi of Emergence: Computation, Dynamics, and Chaos, Physica D, 75, 11-54 (1994).

[29] Adidah Lajis & Normaziah Ab Aziz. Assessing Learner's Understanding via Node-Link Analysis: An Information Theory Approach, In Proceedings International Symposium on Information Technology, KLCC, August (2008).

[30] Pulman, S.G., Sukkarieh, J.Z.: Automatic short answer marking. In: Proceedings of the Second Workshop on Building Educational Applications Using NLP, EdAppsNLP 05, pp. 9–16. Association for Computational Linguistics, Stroudsburg, PA, USA (2005)

[31] Sukkarieh, J.Z., Pulman, S.G. Information extraction and machine learning: auto-marking short free text responses to science questions. In: Proceedings of the 2005 Conference on Artificial Intelligence in Education: Supporting Learning Through Intelligent and Socially Informed Technology, pp. 629–637. IOS Press (2004).

[32] Sukkarieh, J.Z., Pulman, S.G., Raikes, N. Auto-marking: using computational linguistics to score short, free text responses. In: International Association of Educational Assessment, Philadephia (2004).

[33] Sukkarieh, J.Z., Pulman, S.G., Raikes, N. Auto-marking 2: an update on the ucles-oxford university research into using computational linguistics to score short, free text responses. In: International Association of Educational Assessment, Philadephia (2004).

[34] Sargeant J., Wood M.M., Anderson S. M. A human-computer collaborative approach to the marking of free text answers. Eighth International CAA Conference, Loughborough University, Loughborough, UK, 2004 pp.361-370 (2004).

[35] Sargeant J., Wood M.M., Anderson S. M.: A human-computer collaborative approach to the marking of free text answers. Eighth International CAA Conference, Loughborough University, Loughborough, UK, 2004 pp.361-370.

[36] (Sukkarieh, J.Z. & Blackmore, J. Automatic Content Scoring for short constructed responses. In FLAIRS Conference, AAAI Press (2009).

[37] Mitchell, T., Russell, T., Broomhead, P. & Aldridge N. Towards robust computerized marking free-text responses. Paper presented at the 6th International Computer Aided Assessment Conference, Loughborough University (2002).

[38] Graesser, A. C., Wiemer-Hastings, P. , Wiemer-Hastings, K., Harter, D. Tutoring Research Group, and N. Person. Using latent semantic analysis to evaluate the contributions of students in autotutor. Interactive Learning Environments, 8(2):129–147 (2000).

[39] Wiemer-Hastings, P., Wiemer-Hastings, K. and Graesser. A.. Improving an intelligent tutor's comprehension of students with Latent Semantic Analysis. Artificial Intelligence in Education , pages 535–542 (1999).

[40] Mintz, T. H., Wang, F. H., & Li, J. Word categorization from distributional information: Frames confer more than the sum of their (Bigram) parts. Cognitive psychology, 75, 1-27 (2014).

[41] Madnani, N., Burstein, J., Sabatini, J., Oreilly, T. Automated scoring of a summary writing task designed to measure reading comprehension. In proceeding of the 8th Workshop on Innovative use of NLP for Building Educational Applications, pp. 163-168 (2012).

[42] Fahmy, A. A., Gomaa, W. H. Short Answer grading using string similarity and corpus-based similarity. Int. Journal Adv Computer Science Application (2012).

[43] Mohler, M., Bunescu, R.C., Mihalcea, R. Learning to grade short answer questions using semantic similarity measures and dependency graph

alignments. In: The Association for Computer Linguistics, ACL, pp. 752–762 (2011).

[44] Mohler, M., Mihalcea, R. Text-to-text semantic similarity for automatic short answer grading. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, pp. 567–575. Association for Computational Linguistics (2009).

[45] Siddiqi, R., Harrison, C.J., Siddiqi, R. Improving teaching and learning through automated short-answer marking. IEEE Trans. Learn. Technol. 3 (3), 237–249 (2010).

[46] Dzikovska, M., Bell, P., Isard, A., Moore, J.D. Evaluating language understanding accuracy with respect to objective outcomes in a dialogue system. In: EACL, pp. 471–481. The Association for Computer Linguistics (2012).

[47] Allen, E.B. Information Theory and Measurement. PhD Thesis. Florida Atlantic University Boca Raton (1995).