Competency Assessment of Short Free Text Answers

Adidah Lajis [#1], Normaziah Abdul Aziz [*2]

Malaysian Institute of Information Technology, Universiti Kuala Lumpur

Kuala Lumpur, Malaysia

[1] adidahl@unikl.edu.my

[*] *Kulliyah of Information and Communication Technology, International Islamic University Malaysia*

Gombak, Selangor, Malaysia

[2] naa@iium.edu.my

***Abstract*— The increased adoption of competency-based education has posed the need of an automated competency assessment. Most of the existing assisted assessment does not cater for competency assessment. The high percentage on the use of short free text answer as competency assessment shows that the need of the competency assisted assessment is urgent. This paper studies on the need and also review on existing assisted assessments focusing on short free text answer. A Node Link (NL) Scoring technique is proposed as an alternative automated solution to assess learners' competency in short free text answers.**

Keyword: competency assessment; short free text answers; Node Link Scoring technique

I. INTRODUCTION

Assessment is an important part of learning. It is the process of making inferences about an individuals knowledge, skills, attitudes or other constructs using information from one or more methods such as tests, observations, interviews, projects or portfolios with reference to pre-defined criteria (learning expectations, measurement of learning outcomes) [1]. However, the increased numbers of students in Higher Education and the corresponding increase in time spent by staff on assessment has encouraged interest into how technology can assist in this area. Malaysia alone, the population of student in Higher Education has been increased gradually every year [2] For example, there is an increase of the intake of diploma student in public Higher Education Institute (HEI) form year 2009 for about 82,208 students and 94,026 students in year 2010 and also tremendous increase for the same level of study in private HEI that is from 198,760 students to 220,299 students.

Assessment has always been an aspect of the use of information and communication technology in education. Thelwall (2000) considered computer based assessment as an educational tool in higher education [3]. Bull et al. (2004) and Mills et al. (2002) claimed that computer assisted assessment offers pedagogic and productivity benefit has raised the prospect of reductions in tutors' assessment burden [4][5]. This advantages also has been agreed by Robinson et al. in their paper titled Mathematics Lecturers' Practice and Perception of Computer Aided Assessment (2012) [6].

However, the usage of outcome based education in higher institution requires a well-blended of competencies and learning objectives to be achieved by the student. Assessment based on learning outcome has been broadly defined as what the learner should be able to achieve at the end of learning period and competencies will say how we can be certain they know it. Thus the assessment is not alone but the assessment of the competency also must be done.

Yet, most current competence assessment methods are still exclusively based on paper-and-pencil formats [7]. This assessment medium puts both time and creativity constraints on the assessment process. With the rise of computer assisted education several attempts have been made to transfer existing competence assessment methods to the computer. By using new technology, assessments can be designed to be more authentic and challenging than paper-and-pencil-based assessments allowing for an active demonstration of knowledge, in contrast to talking or writing about it [8]. Apart from the fact that authentic and challenging assessments can provide a deeper insight into a learner's competence, and may be able to measure higher order cognitive skills that cannot be easily tested with a standardized paper-and-pencil instrument [7].

Adidah and Normaziah (2012) have shown that Bloom's competency test can be assessed by using NL Scoring technique [9]. The assessment method used in the study is on short free text answers with a length of 3 to 150 words. The competencies assessed were knowledge, understanding, analysis and evaluation. The experiment results showed that NL Scoring only can be used to assess knowledge, understanding and evaluation. The test data were collected from four different domains namely natural science, computer science, medical and engineering domain at school and tertiary level. This paper intends to further investigate the present landscape of competency assisted assessment through short free text answers and compare to the proposed NL Scoring technique.

A. Assessment Methods

Mansell (2002) stated that the type of knowledge and skills to be assessed can affect the choice of assessment methods needed [10]. And according to Bloom (1956), the length of the answer does not correlate with the levels of Bloom's taxonomy [11].

Many academics examinations make heavy use of questions that require students to write one or two sentences. For example, questions often ask candidates to state, to suggest, to describe, or to explain. Short answers questions are highly valued and included in most examinations. They are also extensively used by educators in assessing learners' understanding of particular content of knowledge. Carter et al. [33] indicated that a web-survey on cognitive level assessment style have shown that a heavy used of closed book examination is 81% with 68% of the total assessment to test remembering, 91% to test understanding, 57% to test application and 37% to test evaluation. This shows that the short free text answer has been used to access the higher cognitive level.

B. *Outcome-based Education and Competency-based Education*

Assessment based on learning outcome has been broadly defined as what the learner should be able to achieve at the end of the learning period. The competencies assist us to be certain that the learner gained the learning outcome. OBE can be implemented in various modalities, including face-to-face, online and hybrid models. Jennings [12] explains the process of outcomes assessment as "specifying the goals and objectives of a program and ways in which the attainment of those goals can be measured."

Whilst, competency-based education (CBE) is a narrower concept, a subset or instance of OBE, where the outcomes are more closely tied to job skills or employment needs, and the methods are typically self-paced. It is producing evidence to make a judgement [decision] about whether the person is competent in relation to a particular standard and the competency-based assessment is not the same as performance based assessment [13]. According to Sitthisak et al. (2007), the competency evidence substantiates the existence, sufficiency, of level of the competency and might include test result, report evaluation, certificates or licenses [14]. Competencies encompass the cluster of skills, knowledge, abilities and behaviors required for success across all professional jobs and this enables an individual to perform task to the standard required for successful job performance. Bloom levels are a way of categorizing competencies pieces such as knowledge, skills, abilities and behavior. Bloom has categorized these into levels and identified behavior for each level along with methods to test or achieve the level.

C. *Educational Taxonomy and Assessment*

Taxonomies of educational objectives can similarly be used to provide a shared language for describing learning outcomes and performance in assessments. They divide educational objectives into three domains, cognitive, affective and psychomotor. These taxonomies have been based on a range of educational theories and research. The most widely cited by educationist and related researchers is the original Bloom's taxonomy [11].

Bloom's (1956) Taxonomy of Educational Objectives relating to the cognitive domain has influenced many educationists over the years – more so than the companion volumes relating to the affective and psychomotor domains respectively [15][16]. Each of these taxonomies is hierarchical, with any higher level subsuming all objectives beneath them in the hierarchy (although the hierarchy may not be clear-cut, as Harrow acknowledges). Whilst the taxonomy relating to the cognitive domain has proved useful for analyses of cognitive demand, whether at the stage of constructing curricula or of assessing students' performance, it has to be used with reference to the epistemological level of the subject material. This taxonomy has been taken as the starting point for analyzing the student's learning competence. As shown in Table I, the six cognitive skills as suggested by Bloom, its questions cues, and skills demonstrated. The questions cues are according to Bishop [17] and Paterson [18] suggestion.

TABLE I

Competence	Question Cues	Skills Demonstrated
Knowledge	Name, define, state, select, show and draw	Recall of information
Understanding	Explain, describe, estimate, classify, rewrite and convert	Grasp the text meaning
Application	Solve, use, demonstrate, apply, show and illustrate	Practical use of material
Analysis	Compare, analyze, differentiate, select, deduce and solve	Notice patterns and hidden data
Synthesis	Draw conclusion based on evidence, formulate, modify, combine, rearrange and generate	Digest information
Evaluation	Recommend and grade, justify and interpret	Judge value for purpose

II. REVIEW OF ASSISTED ASSESSMENT FOR SHORT FREE TEXT ANSWER

There are various techniques applied in the assessment of students' answer text. Different techniques may be reliable to asses a specific area of assessment such as Latent Semantic Analysis has it focus on the content and is mostly used for the assessment of humanities essay [19]. LSA has been applied to essay grading, and high agreement levels obtained. These techniques are more suited to marking essays than short-answer questions, since they focus on metrics which broadly correlate with writing style, augmented with aggregate measures of vocabulary usage [20].

Unlike the holistic assessment of content and style aspects for essays, the interest in short free text answer assessment is solely focused on important content aspects. Various techniques have been used from statistical approaches to applying methods of artificial intelligence, particularly natural language processing. Natural Language Processing (NLP) uses algorithm to identify semantic relations among human language words [21]. Recent advancement in Natural Language Processing has resulted positive impact in the way free text being evaluated today. For example, information extraction has allow some extent of semantic meaning of natural language text to be captured [22][23]. The following will briefly describe the techniques used for the assessment of short free text answer.

A. Natural Language Processing (NLP)

NLP is a sub-field of Artificial Intelligence and Linguistics. Low-level NLP tasks include [34]:

- *Sentence boundary detection*
- *Tokenization*
- *Part-of-speech*
- *Morphology*
- *Shallow parsing (chunking)*
- *Problem-specific segmentation*

Higher-level tasks build on low-level tasks and are usually problem-specific [34]. They include:

- *Spelling/grammatical error identification and recovery*
- *Named entity recognition (NER)*

Auto-marking, C-Rater and NL Scoring apply the above mentioned techniques. However, Auto-marking is combining this approach with pattern matching technique in the system. They combine by including statistical method and involving a deep text parsing with semantic analysis to gather more information to effectively assess the learner's answer. NL Scoring also combine the above techniques with information theory to justify the amount of knowledge produced by the learner.

B. Information Extraction (IE)

It is a sub-discipline of NLP. Information extraction (IE) techniques pull out pertinent information from a partially syntactically analysed text by applying a set of domain-specific patterns typically built from training data. The classic IE tasks include [35]:

- NER
- *Co-reference Resolution (CO)*
- *Relation Extraction (RE)*
- *Event Extraction (EE)*

According to Grishman the process of IE has two major parts [24]. First, the system extracts individual facts from the text of a document through local text analysis. Second, it integrates these facts, producing larger facts or new facts (through inference). As a final step after the facts are integrated, the pertinent facts are translated into the required output format.

Most IE systems identify candidate concepts by applying NLP techniques and simple pattern matching [25]. Preferred candidates are noun phrases or noun-noun collocations [26]. Shallow NLP techniques are also preferred by IE systems to handle large amount of data. Automated Text Marker and Automark are underpinned of these techniques. However, Automark also applied NLP techniques to ignore some mistakes in spelling, typing, syntax or semantics that should not be taken into account.

The main drawback of IE system is low portability, due to language dependent linguistic resources and to domain-specific knowledge (ontology) [27].

C. Clustering

Clustering refers to the grouping of items according to some measure of similarity. In this techniques, a group of essay of similar words pattern to form a cluster with the same score. In 2005, Pappuswamy et al. applied a multitier clustering method to cluster a word a particular concept that defined through the semantic of natural language description [28]. Figure 1 describes the techniques they used to cluster the word. The upper level (cluster and sub-cluster) describe the topic of discussion and the lower level describe the specific principle or misconception. The + sign in each node means the presence of the particular words in a concept description.

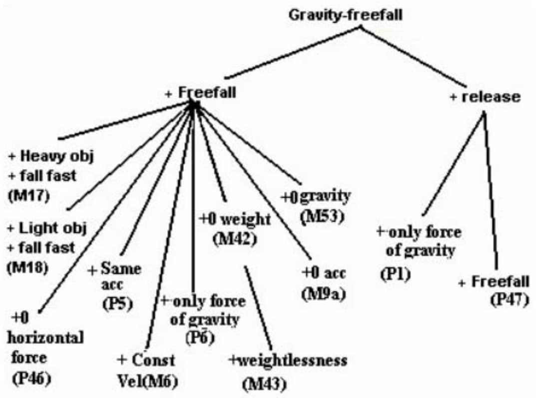

Fig. 1. Tree diagram on features related to the cluster 'Gravity-Freefall' [28]

Clustering can be useful for clarifying and sharpening a vague query, by showing users the dominant themes of the returned results [29]. According to Hearst [30], clustering also works well for disambiguating ambiguous queries; particularly acronyms. Unfortunately, because clustering algorithm are imperfect, they do not neatly group all occurrences of each acronym into one cluster, nor do they allow users to issue follow-up queries that only return documents form the intended sense. The disadvantages of clustering include lack of predictability, conflation of many dimensions simultaneously, the difficulty of labelling the groups and the counter intuitiveness of cluster sub-hierarchies. ABC [31] underpinned of this techniques. However, vector-based clustering approach applied by ABC is not taking into account the word order [32].

D. Comparison of Semantic Network

Network structure analysis is applied to kinds of semantic network analysis in natural language processing, ontology of language, and lexical pattern analysis [36]. Semantic Analysis Grader (SAGrader) has introduced the semantic network of knowledge to recognize pattern in the student answer. SAGrader can be used to assess both short and essay answer script.

E. Hybrid approach

It is also possible to take advantage of the good features of several other techniques to improve a system. For example, Automark has employed both NLP and IE in the system. And E-examiner, C-Rater, Auto-marking, and NL Scoring also have employed more than one technique.

Table II simplify the existing assisted assessment for short free text answer and their focus on competency assessment.

TABLE II

System/ Researcher	Techniques	Length of word	Competency Assessed
Automated Text Marker (Callear et al., 2001)	IE	Not mentioned	Not mentioned
Automark (Mitchell et al., 2002)	NLP IE	Not mentioned	Not mentioned
C-Rater (Leacock & Chodorow, 2003)	NLP Rule-based algorithm	15 to 20 words	Not mentioned
Auto-marking (Pulman & Sukkarieh, 2005)	IE Computational Linguistic	Up to 50 words	Not mentioned
ABC (Sargeant et al., 2004)	Clustering	Less than 100 words	All levels
E-examiner (Gütl,2007)	NLP VSM	Not mentioned	Not mentioned
Automarking (Cutrone & Chang, 2010)	NLP	Not mentioned	Not mentioned
NL Scoring (Adidah & Normaziah, 2012)	NLP and Information Theory	3-150 words	Knowledge, Understanding , Evaluation.
String-based Algorithm (Gomaa & Fahmy, 2012)	String Similarity and Corpus based similarity	Not mentioned	Not mentioned
Cluster based (Brooks et al, 2014)	Clustering	Up to one length	Not mentioned
Okoye et al (2013)	NLP	Not mentioned	Not mentioned
SAGrader	Comparison of Semantic Network	Not mentioned	Not mentioned

As shown above, most of the assisted assessments apply hybrid approach to address the assessment problem of free text answers and improve the assessment performance.

The used of NLP is necessary to extract the students' answer text and later the model can be updated based on small number of language properties and categories [37]. The information theory technique is employed because it is domain independent and quantity independent. It is necessary to propose a technique that can be used in many domains. The proposed technique also takes into account the word order in a sentence [42].

According to [38], IE need to have in the range of 100-500 examples of student answers for each planned test question to assist the creation of IE patterns and it is work best for a specific domain. Text similarity also contribute to some are domain limited and cannot be applied in general [39, 40] and some methods represent a sentence as high-dimensional vector which leads to the sparse data problem and computational inefficiency [41].

III. Conclusion

Automated assessment systems have a variety of different tools and techniques. However, most of the assisted assessment does not assess the learner's competency. The increased adoption of competency-based education has poses the challenge to produce a competency assisted assessment as to replace the conventional competency assessment method that exclusively rest upon paper-based and oral format. Survey shows that at least 37 percent of the short text assessment is used to test evaluation competency and 91 percent of the short test assessment is used to test understanding. This point out the importance of having an automated assisted assessment which focuses on the short free text answer.

References

[1] Pepper, D (2012). KeyCoNet 2012 Literature Review. Assessment for Key competences. Available at http://keyconet.eun.org/c/document_library/get_file?uuid=41099a9b-a557-490c-9955-fb958530e8bd&groupId=11028, 24 February 2014

[2] MOHE (2010). Perangkaan Pengajian Tinggi Malaysia 2010, http://www.mohe.gov.my/web_statistik/perangkaan_2010.pdf, assessed 22 February 2014

[3] Thelwall, M. (2000) 'Computer-based assessment: a versatile educational tool.' Computers & Education 34: 37-49.

[4] Bull, J. & McKenna, C. (2004). Blueprint for Computer-assisted Assessment. London: Routledge Falmer.

[5] Mills, C., Potenza, M., Fremer, J., & Ward, C. (2002). Computer-based Testing building the foundation for future assessment. New Jersey: Laurence Erlbaum Associates

[6] Robinson, C. L., Hernandez-Martinez, P. & Stephen Broughton, S. (2012). Mathematics Lecturers' Practice and Perception of Computer Assisted Assessment. Mapping University Mathematics (pp 105-117). Retrieved from: http://mathstore.ac.uk/headocs/MU-MAP_0.pdf, 22 February 2014

[7] Chung, G. K. W. K. & Baker, E. L. (2003) An Exploratory Study to Examine the Feasibility of Measuring Problem-Solving Processes Using a Click-Through Interface, The Journal of Technology, Learning, and Assessment, 2, https://ejournals.bc.edu/ojs/index.php/jtla/article/download/1662/1504, 23 February 2014

[8] Hunt, N., Hughes, J. & Rowe, G. (2002) Formative Automated Computer Testing (FACT), British Journal of Educational Technology, 33, pp. 525-535.

[9] Adidah, L. & Normaziah, A. A. (2012). NL Scoring and Bloom Competency Test: An Experimental Result, Proceeding of The 6th International Conference on Ubiquitous Information Management and Communication, ICUIMC '12, Kuala Lumpur, Malaysia, February 20-22, 2012. ACM 2012 ISBN 978-1-4503-1172-4

[10] Mansell, J. C. (2002). General issues about assessment of competence. Assessment and evaluation in Higher Education Journal, 19 (1), 3-16.

[11] Bloom, B. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook I, Longman, New York.

[12] Jennings, Edward T. Jr. (1989 Sept/Oct.). Accountability, Program Quality, Outcome Assessment and Graduate Education for Public Affairs and Administration, Graduate Education for Public Affairs and Administration, pp. 438-446.

[13] Moris, A. H. (2013). A new Paradigm in Assessment: Competency Base. Presented in Education Conference 2013, April 2-4, 2013.

[14] Sitthisak, O. Gilbert, L. and. Soonklang T. (2007), Integrating Competence Models with KST for Assessment, CAA 2007

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

[16] Harrow, A.J. (1972). Taxonomy of the Psychomotor Domain: A Guide for Developing Behavioral Objectives. London: Longman.

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

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

[19] Wiemer-Hastings, P., Wiemer-Hastings, K., & Graesser, A. (1999, July). How is Latent Semantic Analysis? Paper presented at the 16th International Joint Congress on Artificial Intelligence, Stockholm, Sweden.

[20] Jordan, Sally and Mitchell, Tom (2009). e-Assessment for learning? The potential of short-answer freetext questions with tailored feedback. British Journal of Educational Technology, 40(2) pp. 371–385.

[21] Girju, R., and Badulescut, A., and Moldovan, D. (2006). Automatic Discovery of Part-whole Relations. Computational Linguistics. 32(1). 83 – 135

[22] Zhu, Q., and Cheng, X. (2008). The Opportunities and Challenges of Information Extraction. 2008 International Symposium on Intelligent Information Technology Application Workshops (IITAW 2008), 597-600.

[23] Wang, H., Yuan, L., and Shao, H. (2008). Text Information Extraction Based on OWL Ontologies. Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery FSKD, 4, 217-222.

[24] Grishman, R. (1997). Information Extraction: Techniques and Challenges Information Extraction (International Summer School SCIE-97), ed. Maria Teresa Pazienza, Springer-Verlag.

[25] Mark, J. (2002, July). A Simple Pattern-Matching Algorithm For Recovering Empty Nodes And Their Antecedents, Paper presented at the 40th Annual Meeting of the Association for Computational Linguistics, ACL 02, Philadelphia, PA, USA

[26] Heid, U.(2000) A linguistic bootstrapping approach to the extraction of term candidates from German text. In Terminology, 161-180

[27] B. Andersson et al. (2002). NLDB 2002 (Eds.), LNCS 2553, pp. 64–75, Springer-Verlag Berlin Heidelberg

[28] Pappuswamy, U., Bhembe, D., Jordan, P. W., & Van-Lehn, K. (2005, May). A multi-tier NL-knowledge clustering for classifying students' essays. Paper presented at the 18th International FLAIRS Conference, Florida USA

[29] Kaki, M. (2005, April). Findex Search result categories help users when document rankings fail. Paper presented at ACM Special Interest Group on Computer-Human Interaction (SIGCHI), Portland

[30] Hearst, M. A. (2006). Clustering versus faceted categories for information exploration. Communication of the ACM, 46(4), 59-61

[31] Sargeant J., M.M. Wood & S.M. Anderson (2004) A human-computer collaborative approach to the marking of free text answers. 8th International Conference on Computer Aided Assessment, Loughborough University, Loughborough, UK, pp.361-370

[32] Wood, M. M., Jone, C., Sargeant, J., Reed, P. Lightweight clustering techniques for short text answer in HCC CAA.

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

[34] Prakash M Nadkarni, Lucila Ohno-Machado, Wendy W Chapman. (2011). Natural language processing: an introduction. J Am Med Inform Assoc. 2011 Sep-Oct; 18(5): 544–551

[35] Piskorski, J. & Yangarber, R. (2013). Information Extraction: Past, Present and Future. In Thierry Poibeau, Horacio Saggion, Jakub Piskorski, and Roman Yangarber, editors. Multisource, Multilingual Information Extraction and Summarization. Theory and Applications of Natural Language Processing, page 23-49. Springer Berlin Heidelberg.

[36] Yeh, H.-C., Yang, Y.-F., & Wong, W.-K. (2010). Interaction Chain Patterns of Online Text Construction with Lexical Cohesion. Educational Technology & Society, 13 (1), 55–68.

[37] Meurers, D. (2012) Natural Language Processing and Language Learning. Encyclopedia of Applied Linguistics, edited by Carol A. Chapelle. Wiley Blackwell

[38] Soderland, S. (1999). Learning Information Extraction Rules for Semi-Structured and Free Text, *Machine Learning,* 34, 233–272

[39] Feng, J., Zhou, Y., & Martin, T. (2008, June). Sentence Similarity based on Relevance. Paper presented at the Information Processing and Management of Uncertainty in Knowledge-based System, IPMU '08, Torremolinos, Malaga

[40] Islam, A. and Inkpen, D. (2008). Semantic text similarity using corpus-based word similarity and string similarity. *ACM Transaction of Knowledge Discovery*, Data. 2, 2, Article 10

[41] Landauer, T. K., Laham, D., Rehder, B., & Schreiner, M. E. (1997, August). How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans, Paper presented at the 19th annual meeting of the Cognitive Science Scociety, Mawhwah, NJ.

[42] Callear, D., Jerrams-Smith, J., & Soh, V. (2001). CAA of short non-MCQ answer. Paper presented at the 5th International CAA Conference organized by Loughborough University