7. CONCLUSION AND FUTURE WORK

There are many existing systems that evaluate student work. However, most of these systems are either limited to evaluating MCQ type answers or to evaluating answers that can be compared against a single teacher-provided answer or a set of possible answers.

There are few systems that evaluate multi-step mathematics answers, but they rely on one of the two approaches: 1. clustering a large number of answers and getting the teacher to manually award marks to one member of the cluster to propagate the marks to the other cluster members, 2. breaking down the question into multiple sub-questions so that the student is forced to provide single step answers to all the sub-questions. These two approaches create two problems: the need to have a large number of answers and losing the purpose of allowing multi-step answers, respectively.

The system developed by us solves these two problems by allowing the student to provide a multi-step answer and giving the grade as soon as the student submits the answer without having to wait for a large number of students to submit the answers. According to the best of our knowledge, this system is the first to grade multi-step open-ended answers according to a marking scheme to award full/ partial credit. The other most important feature is the requirement of zero teacher-involvement during answer grading.

There is research work done in the area of error identification of multi-step answers, but this research is limited to pedagogy, where no automation has been considered. Our system is capable of automatically identifying the type of error made by the student in partially correct answers. The system does this by replicating errors that are common in these types of questions, and comparing the results with the wrong step entered by the student.

As future work, a performance improvement of the error identification module is required. As mentioned in section 5.4.1, the system should be supported with some intelligence in order to have the required thinking ability. Creating an intelligent agent

that is capable of integrating multiple buggy procedures is a possible approach for performance improvement. Also a mechanism can be developed to extract the similar student answers that do not receive a specific feedback. Then we can consult a teacher and get their feedback on what the student has done. That way we can add more buggy procedures to the system.

Currently the G.C.E. O/L marking scheme is strictly followed for grading the student answers. As future work, a practice-mode can be enabled that allows follow-through marking. The student assessment results can be used in all questions to create a student model that gives information of the areas that the student needs to pay more attention.

Currently when the student makes a mistake in the answer, the system identifies the error. This should be presented to the student in the form of feedback. The simplest way to do this is to tell the student that she made a mistake and then to give the correct answer. But the most effective way is to first tell the student the type of the mistake that she made, thus allowing the student to figure out her own mistake. One possible approach will be having multiple feedback levels as follows for partially correct answers:

Level 1: Tell the student that there is a mistake. Ask the student to find the mistake and correct on her own.

Level 2: If the student fails to identify the error, show the step where the mistake has been made. Ask the student to correct on her own.

Level 3: If the student still fails to identify the error, give a sample correct answer to the student. Direct the student to similar questions which will test the area the student lacks knowledge and/ or provide the student with study materials on the particular topic.

REFERENCES

[1] Bennett, R. E., & Bejar, I. I. (1998). Validity and automad scoring: It's not only the scoring. *Educational Measurement: Issues and Practice*, *17*(4), 9-17.

[2] Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, *25*(2-3), 259-284.

[3] Zipitria, I., Elorriaga, J. A., Arruarte, A., & de Ilarraza, A. D. (2004). From human to automatic summary evaluation. In *7th International Conference on Intelligent Tutoring Systems* (pp. 432-442). Springer Berlin Heidelberg.

[4] Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and instruction*, *22*(3), 333-362.

[5] Pérez, D., Alfonseca, E., & Rodríguez, P. (2004). Application of the BLEU Method for Evaluating Free-text Answers in an E-learning Environment. In *Language Resources and Evaluation Conference*. 1351-1354

[6] Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*. 311-318. Association for Computational Linguistics.

[7] Mohler, M., & Mihalcea, R. (2009). 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* . 567-575. Association for Computational Linguistics.

[8] Pado, U., & Kiefer, C. (2015). Short answer grading: When sorting helps and when it doesn't. In *Proceedings of the 4th workshop on NLP for Computer Assisted Language Learning at NODALIDA 2015* 114, 42-50. Linköping University Electronic Press.

[9] Pulman, S. G., & Sukkarieh, J. Z. (2005). Automatic short answer marking. In *Proceedings of the second workshop on Building Educational Applications Using NLP*. 9-16. Association for Computational Linguistics.

[10] Ala-Mutka, K. M. (2005). A survey of automated assessment approaches for programming assignments. *Computer science education*, *15*(2), 83-102.

[11] Reek, K. A. (1989). The TRY system-or-how to avoid testing student programs. In *ACM SIGCSE Bulletin*. 21(1), 112-116. ACM.

[12] Higgins, C., Symeonidis, P., & Tsintsifas, A. (2002). Diagram-based CBA using DATsys and CourseMaster. In *Proceedings. International Conference on Computers in Education*. 167-172. IEEE.

[13] Jackson, D., & Usher, M. (1997). Grading student programs using ASSYST. In *ACM SIGCSE Bulletin*. 29(1), 335-339. ACM.

[14] Cheang, B., Kurnia, A., Lim, A., & Oon, W. C. (2003). On automated grading of programming assignments in an academic institution. *Computers & Education*, *41*(2), 121-131.

[15] Ellsworth, C. C., Fenwick Jr, J. B., & Kurtz, B. L. (2004). The quiver system. In *ACM SIGCSE Bulletin*. 36(1), 205-209. ACM.

[16] Bettini, L., Crescenzi, P., Innocenti, G., & Loreti, M. (2004). An environment for self-assessing Java programming skills in first programming courses. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies.* 161-165. IEEE.

[17] Arnow, D., & Barshay, O. (1999). WebToTeach: an interactive focused programming exercise system. In *29th Frontiers in Education Conference.* 1, 12A9-39. IEEE.

[18] Truong, N., Bancroft, P., & Roe, P. (2003, February). A web based environment for learning to program. In *Proceedings of the 26th Australasian computer science conference,* 16, 255-264. Australian Computer Society, Inc.

[19] Annual Performance Report(2013). (n.d.). Retrieved May 20, 2015, from http://www.moe.gov.lk/english/images/publications/Annual_performance_Report2013/Annual_performance_report_e.pdf

[20] Hansen, H., & Ruuska, M. (2003). Assessing time-efficiency in a course on data structures and algorithms. In *Proceedings of the Third Finnish/Baltic Sea Conference on Computer Science Education.* 93-100.

[21] Chen, P. M. (2004). An automated feedback system for computer organization projects. *IEEE Transactions on Education*, *47*(2), 232-240.

[22] Livne, N. L., Livne, O. E., & Wight, C. A. (2007). Can automated scoring surpass hand grading of students' constructed responses and error patterns in mathematics. *MERLOT Journal of Online Learning and Teaching*, *3*(3), 295-306.

[23] Sangwin, C. J., & Naismith, L. (2008). Implementing Computer Algebra Enabled Questions for the Assessment and Learning of Mathematics. *International Journal for Technology in Mathematics Education*, *15*(1).

[24] Sangwin, C. J., & Grove, M. (2006). STACK: addressing the needs of the neglected learners. In *Proceedings of the 1st Web Advanced Learning Conference and Exhibition* (pp. 81-96).

[25] Dougiamas, M., & Taylor, P. (2003). Moodle: Using learning communities to create an open source course management system in *11th World Conference on Educational Media and Technology*.

[26] Lan, A. S., Vats, D., Waters, A. E., & Baraniuk, R. G. (2015). Mathematical language processing: Automatic grading and feedback for open response mathematical

questions. In *Proceedings of the Second ACM Conference on Learning@ Scale*. 167-176. ACM.

[27] Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science (Washington)*, *228*(4698), 456-462.

[28] Walonoski, J. A., & Heffernan, N. T. (2006). Detection and analysis of off-task gaming behavior in intelligent tutoring systems. In *International Conference on Intelligent Tutoring Systems*. 382-391. Springer Berlin Heidelberg.

[29] Koedinger, K. R., Brunskill, E., Baker, R. S., McLaughlin, E. A., & Stamper, J. (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine*, *34*(3), 27-41.

[30] Melis, E., Andres, E., Budenbender, J., Frischauf, A., Goduadze, G., Libbrecht, P., Pollet, M., & Ullrich, C. (2001). ActiveMath: A generic and adaptive web-based learning environment. *International Journal of Artificial Intelligence in Education (IJAIED)*, *12*, 385-407.

[31] Ambekar, D. (2015). *Evaluation of essays using incremental training for Maximizing Human-Machine agreement* (Doctoral dissertation), Indian Institute of Technology, Bombay.

[32] "edX", *edX*, 2016. [Online]. Available: https://www.edx.org/. [Accessed: 30- Dec- 2015]

[33] Pellegrino, J. W., & Goldman, S. R. (1987). Information processing and elementary mathematics. *Journal of Learning Disabilities*, *20*(1), 23-32.

[34] Norman, D. A. (1981). Categorization of action slips. *Psychological review*, *88*(1), 1-15.

[35] Gonzalez, C. S., Guerra, D., Sanabria, H., Moreno, L., Noda, M. A., & Bruno, A. (2010). Automatic system for the detection and analysis of errors to support the personalized feedback. *Expert Systems with Applications*, *37*(1), 140-148.

[36] Peng, Aihui, and Zengru Luo. (2009) A framework for examining mathematics teacher knowledge as used in error analysis. *For the learning of mathematics* 29(3) 22-25.

[37] Sarwadi, H. R. H., & Shahrill, M. (2014). Understanding students' mathematical errors and misconceptions: The case of year 11 repeating students. *Mathematics Education Trends and Research*, 1-10.

[38] Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive science*, *2*(2), 155-192.

[39] Dhlamini, Z. B., & Kibirige, I. (2014). Grade 9 Learners' Errors And Misconceptions In Addition Of Fractions. *Mediterranean Journal of Social Sciences*, *5*(8), 236.

[40] Makonye, J. P., & Matuku, O. (2016). Exploring Learner Errors in Solving Quadratic Equations. *International Journal of Science Education.* 12(1), 7-15

[41] Wolfram, S. (1991). *Mathematica: a system for doing mathematics by computer*. Addison Wesley Longman Publishing Co., Inc.

[42] Kutzler, B. (1996). *Improving Mathematics Teaching with DERIVE: a guide for teachers*. Chartwell-Bratt.

[43] Joyner, D., Čertík, O., Meurer, A., & Granger, B. E. (2012). Open source computer algebra systems: SymPy. *ACM Communications in Computer Algebra*, *45*(3/4), 225-234.

[44] "HP 50g CAS Enabled RPN Programmable Engineering Calculator : Educalc.net", HP-50, 2016. [Online]. Available: https://www.educalc.net/1861088.page. [Accessed: 30- Dec- 2015].

[45] Fife, J. H. (2011). Automated scoring of CBAL mathematics tasks with m-rater. *Research Memorandum. 11-12. Princeton, NJ: Educational Testing Service*.

[46] Driver, D. (2008). Pedagogical Use of a CAS. In *Proceedings of 13th Asian Technology Conference in Mathematics: Enhancing understanding and constructing knowledge in mathematics with technology*.

[47] Li, G., Li, L., Su, W., & Zhao, Y. (2006). Design and implementation of MAML. In *Proceedings of 11th Asian Technology Conference in Mathematics*.

[48] Klai, Saliha, Theodore Kolokolnikov, and Norbert Van den Bergh.(2000) Using Maple and the web to grade mathematics tests. In *Proceedings of International Workshop on Advanced Learning Technologies*. IEEE.

[49] Gould, J. D. (1975). Some psychological evidence on how people debug computer programs. *International Journal of Man-Machine Studies*, *7*(2), 151-182.

[50] Boehme, R., Fairweather, P., Farooq, U., Lam, D., & Singley, K. (2003). *U.S. Patent Application No. 10/722,926*.

[51] Heid, M. K., Blume, G. W., Hollebrands, K., & Piez, C. (2002). Computer algebra systems in mathematics instruction: Implications from research. *The Mathematics Teacher*, *95*(8), 586-591.

[52] "TI-Nspire", *TI-Nspire* Available: https://education.ti.com/en/us/products/calculators/graphing-calculators/ti-nspire-cas-with-touchpad/tabs/overview. [Accessed: 30- Dec- 2015]

[53] "ClassPad II", *ClassPad II*, 2016. [Online]. Available: https://edu.casio.com/products/cg/cp2/. [Accessed: 30- Dec- 2015]

Appendix A: Sample Marking Scheme for Linear Equation Solving Question

```
<rubric>
 <type>LinearEquations</type>
 <sectionid>1</sectionid>
 <question id="q01" totalMarks="10">
   <sub_question id="q01-01" totalMarks="3">
      <data required="true" marks="1">
        <concept>
          <name>LCD</name>
        </concept>
      </data>
      <data required="true" isFinal="false" marks="1">
        <concept>
          <name>CM</name>
        </concept>
      </data>
      <data required="true" isFinal="false" marks="1">
        <concept>
          <name>FinalAns</name>
        </concept>
      </data>
   </sub_question>
 </question>
</rubric>
```

Appendix B: Sample Marking Scheme for Quadratic Equation Solving Question

```
<rubric>
 <type>QuadraticEquations</type>
 <sectionid>1</sectionid>
 <question id="q01" totalMarks="10">
   <sub_question id="q01-01" totalMarks="5" positiveOnly="0" decimalPlaces="2">
      <data required="true" marks="1">
        <concept>
          <name>PS</name>
        </concept>
      </data>
      <data required="true" marks="1">
        <concept>
          <name>SQRT</name>
        </concept>
      </data>
       <data required="true" marks="1">
        <concept>
          <name>SUBST</name>
        </concept>
      </data>
      <data required="true" marks="2">
        <concept>
          <name>FinalAns</name>
        </concept>
      </data>
   </sub_question>
 </question>
</rubric>
```