2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #32406

Design of CAIR Assessment-monitoring Display

Dr. Bahar Memarian, University of Toronto

Bahar Memarian is a researcher, educator, and analyst with research interests in Human Factors Engineering and Engineering Education. She received her Ph.D. (2021) in Industrial Engineering and the Collaborative Specialization in Engineering Education at the University of Toronto, Canada. Before that, she completed her MASc. (2015) and BASc. (2012) in Electrical Engineering from the University of Toronto.

Dr. Susan McCahan, University of Toronto

Susan McCahan is a Professor in the Department of Mechanical and Industrial Engineering at the University of Toronto. She currently holds the positions of Vice-Provost, Innovations in Undergraduate Education and Vice-Provost, Academic Programs. She received her B.S. (Mechanical Engineering) from Cornell University, and M.S. and Ph.D. (Mechanical Engineering) from Rensselaer Polytechnic Institute. She is a Fellow of the American Association for the Advancement of Science in recognition of contributions to engineering education has been the recipient of several major teaching and teaching leadership awards including the 3M National Teaching Fellowship and the Medal of Distinction in Engineering Education from Engineers Canada.

Design of CAIR Assessment Monitoring Display

Abstract

This work in progress (WIP) paper describes the conceptual design of an information visualization display. Earlier the authors proposed a 6-item coding scheme for the evaluation of engineering problem-solving skills. The coding scheme is called CAIR which stands for Constructive Alignment Integrated Rating system. CAIR can offer insight into the abstractions and depth of errors present in the solution to any closed-ended engineering problem. However, on its own, CAIR cannot communicate the grading styles of teaching assistants or problem-solving abilities of students in aggregate for a test, course, courses, and so on. The goal of this work is to visualize and make meaning of CAIR-related assessment data. Our display design is inspired by concepts from the domain of human factors engineering. A low-fidelity conceptual design and walk-through of the display are provided and key scenarios and tasks the instructor can achieve via using the display are explored. The display can inform the instructor on both the quality of the marking done by the assessor(s) and common problem-solving errors committed by the students across a problem, test, and so on.

Introduction

Meeting the pedagogical goals of Constructive Alignment, Formative and outcomes-based Assessment are deemed significant for learning [1]–[3]. Constructive Alignment promotes a social negotiation and mapping between the goals set and evaluations made for an activity such as engineering problem-solving [1], [4], [5]. In this view, assessment can be conceptualized as a cycle where each group of instructors, assessors, and students needs to interact and successfully fulfill the expectations surrounding their role to facilitate student learning and complete an assessment loop.

Instruments in the form of marking and feedback schemas that facilitate the outcomes-based assessment of student competencies exist [6]–[17]. There also exists data mining algorithms that use internal and external assessments along with student demographic and psychometric factors to predict student performance [18]–[21]. Most of the assessment systems (i.e., instruments or techniques), however, do not outline how the instructor should make use of the data collected and synthesize measured skills and meta-competencies (i.e., engineering problem solving) that transcend across domains. Besides, existing systems tend to present tabular data and descriptive summaries (e.g. median), or histograms only, rather than illustrate criterion-based evaluation data such that they are entirely visual [22], [23].

Earlier, the authors proposed an instrument to enable data collection around the assessment of engineering problem-solving skills per student solution to any given closed-ended engineering problem. For a complete description of the instrument and its use, see [24]. The instrument is called CAIR which stands for Constructive Alignment Integrated Rating system. The CAIR instrument was inspired by the concept of Work Domain Analysis and its tool called the Abstraction Hierarchy from the field of human factors engineering [25]. The Work Domain Analysis advises designing a system based on the goals and governing constraints of the environment in which the system operates. In our case, the system of interest is a formative assessment tool. The goal of the instrument is to attribute solution-specific feedback with generic engineering problem-solving competencies as expressed by accreditation agencies such as ABET and CEAB [26], [27]. The Abstraction Hierarchy further outlines an organizational structure for piecing the information requirements of a system. Using these concepts and findings from engineering problem-solving and assessment literature, we derived a model for error tagging and feedback for any given closed-ended engineering problem-solving activity.

When using CAIR, each error-tag has two key attributes, namely has 1) an error abstraction type classification of Goal, Theory, or Calculation and 2) an error depth or decomposition type classification of Surface or Deep. Having this information at hand is useful because the literature reveals there are some patterns in engineering problem-solving expertise and errors committed. For example, underperforming student solutions (i.e., receive a failing mark) tend to have a higher number of errors than achieving student solutions (i.e., receive a passing mark). Also, underperforming students exhibit a higher proportion of goal and theory errors as well as deep errors as compared to achieving students in their problem-solving work [28]–[30].

Situating the CAIR instrument within the assessment cycle alone underscores the prominent role the instructor plays from start to finish. More specifically, the instructor not only influences the assessment cycle at the beginning during the creation of problems and assessment guides but also after the completion of marking. The assessment cycle has the potential to both provide students with formative feedback on their work and provide the instructor with insights into student learning. This work in progress paper shares the conceptual design of a display that takes as input the data collected from the digital version of CAIR. A version of this design was first proposed by [31].

Display Design Process

We first identified the goals of the primary user (instructor): 1) to track the quality of evaluations made by the assessors and 2) to track the types of problem-solving errors demonstrated on student solutions. Second, the data that is produced by assessors and students during the assessment process was outlined. Examples include the total mark of a problem, types of errors tagged by the assessors, performance grade of student solutions, etc. Using this data, the goals were then broken down and detailed into a set of user stories, as shown in Figure 1. The advantage of user stories was that they could be turned into formulaic statements that are understandable by computer logic (if-then statements and quantitative relationships). Interface (i.e., semantic mapping) and universal design (i.e., consistency) principles were then utilized to reformulate these statements into a visual, which we refer to as the building block visual [25], [32]. This building block visual effectively communicates all the functional statements (i.e., 1.1 to 1.3, 2.1 shown in Figure 1) through its structural organization. Our building block visual can transcend across multiple aggregates of the data.

As the instructor I want to:

1. Evaluate the quality of the grading process carried out by the assessors…
   - 1.1 I want to monitor the counts of CAIR error tags on graded solutions to check that failing solutions have more Goal and Theory error tags combined than the passing student solutions. This is to check that the grade being given and the feedback the student receives is in alignment.
   - 1.2 I want to monitor the counts of CAIR error tags on graded solutions to check that the Deep error count of failing student solutions is greater than passing student solutions. Again, this is to check that the grade being provided aligns with the feedback the student receives.
   - 1.3 I want to monitor the time spent by the assessors evaluating student solutions and check that it is in the expected range or consistent among assessors overall and across passing and failing performance categories. This is to make sure assessors are spending their assigned teaching assistantship time or check if I have assigned an appropriate time to them.

2. Stay informed about the quality of the problem-solving skills being exhibited by my students.
   - 2.1 I want to know what types of errors students are making overall, and what types of errors did students who have passing solutions make? And what types of errors did students who have failing solutions make? This is to inform my instructional approach and help me provide overall feedback to the class.

Figure 1. User stories are broken down into functional statement(s), taken from [24].

Figure 2 illustrates the building block visual as it may appear on the proposed display. Here an example data distribution is shown along with the legend. It is important to note that the visual shown in Figure 2 is only a preliminary conceptual design. Our future work requires testing and iteration to arrive at a polished design. Due to the nature of obtained formulaic statements, we gathered that the instructor mostly needs to compare the distribution and proportions of a variable (e.g., deep errors, surface errors, or time spent) across two groups (e.g., receive a passing versus failing mark). Following visual and universal design guidelines, we use a stacked bar structure repeatedly around three sides of a hypothetical data aggregate box. Further, we decomposed a hypothetical data aggregate box into two performance groups and used the stacked bar structure around the three sides of the “failing” and “passing” boxes. We then used three shades of distinct colors (e.g., grey, blue, and green here but a more accessible texture color space could be used in the final design) to differentiate between our three main variables. The title at the top of a building block visual would reveal the data aggregate under analysis (e.g., individual problems on a test, or all the problems one a test together, etc.). The number of student solutions within the failing, passing, and whole categories is listed underneath each box. The left side of each box shows the distribution of Goal, Theory, and Calculation error tags for solutions with at least one error marked by the TA. The top side of each box presents the distribution of Deep and Surface errors for the same set of solutions. The right side of each box shows the distribution of time spent by the assessors on marking each of the solutions. The time stacked bar accounts for the expected time range set a priori by the instructor. The time stacked bar indicates if assessors spent less (under time) or more (over time) than expected in their evaluations. Using the stacked bar graphical element repeatedly and the isomorphic (repeated) structure of the building block as a whole and for the failing and passing solution data sets follow the consistency principle from universal design.

The user stories are addressed through this visualization as follows (taken from [24]):

1.1 The instructor compares the left side of the passing box to the left side of the failing box to check to see that student solutions that receive a failing mark have a higher proportion of the darker grey shade than passing solutions (i.e., failing solutions have more Goal and Theory errors than passing solutions).
1.2 The instructor compares the top of the passing box to the top of the failing box to check to see that student solutions that receive a failing mark have a higher proportion of darker blue shade than passing (i.e., failing solutions have more Deep errors than passing solutions).
1.3 The instructor examines the right side of all of the boxes to check to see that for each category there is a minimal degree of the lighter shade of green present (i.e., assessors spend at least the minimum expected time on each solution). The instructor might also be concerned with how much dark green is represented in the display because this would indicate that the assessors spent more time than expected.
2.1 The instructor compares the left and top sides of each box to deduce the CAIR-specific tags present.

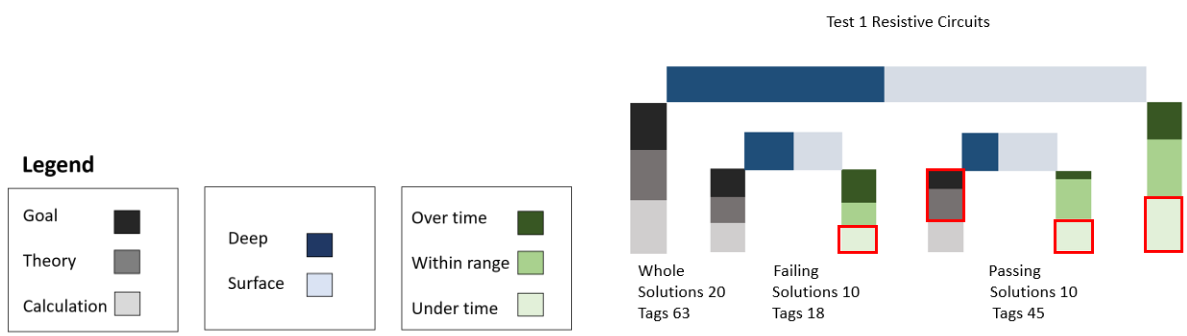

Figure 2. Legend and overview of building block visual, taken from [24].

Conclusion and Future Work

This work in progress paper shares the conceptual design of a display that uses data from CAIR, an instrument previously proposed by the authors for marking and feedback on closed-ended engineering problem-solving activities. The display aims to open a discussion around the potentials of using and processing outcomes-based assessment data. The display can help the instructors to first examine the quality of assessor evaluations and once necessary adjustments are made then use the data to monitor the state of problem-solving errors in student solutions at various levels: the problem level, test level, course level, and potentially across multiple courses. An assessment approach built on the Constructive Alignment principle that cultivates a marking system like CAIR has the capacity for monitoring various aggregations of data within the assessment cycle. Working with data aggregations that are constructively aligned can help the instructors to achieve outcomes-based assessments at the course and curriculum levels. Future work requires testing and analysis of the instructor's uptake with the CAIR display. Future work should also consider the development of tools and resources that can intuitively guide the faculty with the development of user stories when wanting to evaluate the quality of their course assessments.

References

[1] J. Biggs, "Enhancing teaching through constructive alignment," *High. Educ.*, vol. 32, no. 3, pp. 347–364, 1996, doi: 10.1007/BF00138871.

[2] R. Stiggins, "From formative assessment to assessment FOR learning: A path to success in standards-based schools," *Phi Delta Kappan*, vol. 87, no. 4, pp. 324–328, 2005, doi: 10.1177/003172170508700414.

[3] P. Black and D. Wiliam, "'In praise of educational research': Formative assessment," *Br. Educ. Res. J.*, vol. 29, no. 5, pp. 623–637, 2003, doi: 10.1080/0141192032000133721.

[4] J. Biggs, "Assessment and Classroom Learning: A Role for Summative Assessment?," *Assess. Educ. Princ. Policy Pract.*, vol. 5, no. 1, pp. 103–110, 1998, doi: 10.1080/0969595980050106.

[5] J. Biggs, *Teaching for quality learning at university: What the student does.* McGraw-Hill Education (UK)., 2011.

[6] AAC&U, "Problem Solving VALUE Rubric," 2009. https://www.aacu.org/value/rubrics/problem-solving (accessed Jan. 01, 2017).

[7] S. J. Grigg, "A process analysis of engineering problem solving and assessment of problem solving skills.," Dissertation. Clemson University, Clemson, SC., 2012.

[8] C. D. McAllister, X. Jiang, and F. Aghazadeh, "Analysis of engineering discipline grade trends," *Assess. Eval. High. Educ.*, vol. 33, no. 2, pp. 167–178, 2008.

[9] S. M. Barnes, K. D., & Buring, "The effect of various grading scales on student grade point averages," *Am. J. Pharm. Educ.*, vol. 76, no. 3, 2012.

[10] S. M. Humphry and S. a. Heldsinger, "Common Structural Design Features of Rubrics May Represent a Threat to Validity," *Educ. Res.*, vol. 43, no. 5, pp. 253–263, 2014, doi: 10.3102/0013189X14542154.

[11] A. Jonsson and G. Svingby, "The use of scoring rubrics: Reliability, validity and educational consequences," *Educ. Res. Rev.*, vol. 2, no. 2, pp. 130–144, 2007, doi: 10.1016/j.edurev.2007.05.002.

[12] M. M. Hull, E. Kuo, A. Gupta, and A. Elby, "Problem-solving rubrics revisited: Attending to the blending of informal conceptual and formal mathematical reasoning," *Phys. Rev. Spec. Top. Educ. Res.*, vol. 9, no. 1, p. 010105, 2013, doi: 10.1103/physrevstper.9.010105.

[13] M. Besterfield-Sacre, J. Gerchak, M. R. Lyons, L. J. Shuman, and H. Wolfe, "Scoring concept maps: An integrated rubric for assessing engineering education," *J. Eng. Educ.*, vol. 93, no. 2, pp. 105–115, 2004, doi: 10.1002/j.2168-9830.2004.tb00795.x.

[14] B. M. Moskal and J. A. Leydens, "Scoring rubric development: Validity and reliability," *Pract. assessment, Res. Eval.*, vol. 7, no. 1, p. 10, 2000.

[15] S. M. Brookhart *et al.*, "A century of grading research: Meaning and value in the most common educational measure," *Rev. Educ. Res.*, vol. 86, no. 4, pp. 803–848, 2016, doi: 10.3102/0034654316672069.

[16] C. Marshman, E., Sayer, R., Henderson, C., & Singh, "Contrasting grading approaches in introductory physics and quantum mechanics: The case of graduate teaching assistants," *Phys. Rev. Phys. Educ. Res.*, vol. 13, no. 1, p. 010120, 2017, doi:

10.1103/PhysRevPhysEducRes.13.010120.
[17] A. Carberry, M. Siniawski, S. A. Atwood, and H. A. Diefes-Dux, "Best practices for using standards-based grading in engineering courses," 2016, doi: 10.18260/p.26379.
[18] S. K. Mohamad and Z. Tasir, "Educational data mining: A review," *Procedia-Social Behav. Sci.*, vol. 97, pp. 320–324, 2013, doi: 10.1016/j.sbspro.2013.10.240.
[19] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1432–1462, 2014, doi: 10.1016/j.eswa.2013.08.042.
[20] P. Roy, D., Bermel, K. Douglas, H. Diefes-Dux, M. Richey, K. Madhavan, and S. Shah, "Synthesis of clustering techniques in educational data mining," 2017, doi: 10.18260/1-2--28897.
[21] A. M. Shahiri and W. Husain, "A review on predicting student's performance using data mining techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015, doi: 10.1016/j.procs.2015.12.157.
[22] S. S. Lin, E. Z. F. Liu, and S. M. Yuan, "Web-based peer assessment: feedback for students with various thinking-styles," *J. Comput. Assist. Learn.*, vol. 17, no. 4, pp. 420–432, 2001.
[23] N. R. Aljohani and H. C. Davis, "Learning analytics and formative assessment to provide immediate detailed feedback using a student centered mobile dashboard," in *IEEE*, 2013, pp. 262–267.
[24] B. Memarian, "Design for Assessment of Engineering Problem-solving Skills", Ph.D. dissertation, Dept. Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, 2021.
[25] K. J. Vicente, *Cognitive Work Analysis - Toward Safe, Productive, and Healthy Computer-Based Work*. CRC Press, 1999.
[26] Engineers Canada, "Canadian Engineering Accreditation Board Accreditation Criteria and Procedures," 2019. [Online]. Available: https://engineerscanada.ca/sites/default/files/accreditation/Accreditation-Criteria-Procedures-2019.pdf.
[27] ABET, "Criteria for accrediting engineering programs," 2016. http://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2016-2017/.
[28] M. T. Chi, R. Glaser, and E. Rees, "Expertise in problem solving," in *R. Sternberg (Ed.), Advances in the psychology of human intelligence*, NJ: Erlbaum, 1982.
[29] E. B. Coleman and M. Shore, "Problem-solving processes of high and average performers in physics," *J. Educ. Gift.*, vol. 14, no. 4, pp. 366–379, 1991, doi: 10.1177/016235329101400403.
[30] D. R. Woods *et al.*, "Developing problem solving skills: The McMaster problem solving program," *J. Eng. Educ.*, vol. 86, no. 2, pp. 75–91, 1997, doi: 10.1002/j.2168-9830.1997.tb00270.x.
[31] B. Memarian and S. McCahan, "Outcomes-Based Assessment Instrument for Engineering Problem-Solving Skills," 2020.
[32] D. Vernon, C. Reising, and P. M. Sanderson, "Ecological interface design for Pasteurizer II: A process description of semantic mapping," *Hum. Factors*, vol. 44, no. 2, pp. 222–247, 2002.