Expanding a National Network for Automated Analysis of Constructed Response Assessments to Reveal Student Thinking in STEM

Mark Urban-Lurain, Michigan State University, Overall Project PI[1]
Andrea Bierema, Michigan State University, Post-doctoral Researcher
Kevin Haudek, Michigan State University, co-PI
Anne-Marie Hoskinson, Michigan State University, Post-doctoral Researcher
Jennifer Kaplan, University of Georgia, PI
Jennifer Knight, University of Colorado Boulder, PI
Paula Lemons, University of Georgia, PI
Carl Lira, Michigan State University, co-PI
Jill McCourt, University of Georgia, Post-doctoral Researcher
John Merrill, Michigan State University, PI
Rosa Moscarella, Michigan State University, Post-doctoral Researcher
Ross Nehm, SUNY-Stony Brook, PI
Karen Pelletreau, University of Maine, Post-doctoral Researcher
Luanna Prevost, University of South Florida, PI
Michelle Smith, University of Maine, PI
Matthew Steele, Michigan State University, Post-doctoral Researcher
Mary Anne Sydlik, Western Michigan University, PI – Project Evaluation

Need: Faculty who wish to respond to, and build upon, students' existing understandings of key STEM concepts must first know what and how students think about the concepts. While multiple-choice assessments are easy to administer, they cannot measure students' abilities to organize individual bits of knowledge into a coherent and functional explanatory structure. Writing is an authentic task that can reveal student thinking, but is time-consuming to evaluate and therefore difficult to implement in large classes typical of many introductory STEM courses. The Automated Analysis of Constructed Response (AACR, pronounced "acer") project combines educational research-based methods with computerized linguistic analysis to quickly evaluate student writing, generating useful and timely feedback for faculty to inform their instruction. www.msu.edu/~aacr

Goals: We are a large, multi-institutional collaboration (TUES 3 and WIDER funding) with several connected goals: 1) create a national web portal to access AACR conceptual assessments and analysis; 2) use resulting reports to focus community collaborations between STEM education researchers and instructors; 3) transport AACR innovations through ongoing faculty professional development; 4) expand the range of STEM disciplines in which we pursue this research from our biology into chemistry, chemical engineering, physics/astronomy, and statistics; 5) engage in ongoing project evaluation for continuous quality improvement; and 6) lay the foundation for sustainability.

---

[1] Co-authors are listed in alphabetic order after the first author

Approach: We use a variety of computerized lexical analysis and machine learning tools to create statistical models that predict expert rating of student writing with inter-rater reliability as good as expert-to-expert IRR (>0.8). These models are used to generate reports for faculty that detail both scientific and alternative conceptions in their students' responses. Local Faculty Learning Communities (FLCs) meet to discuss the reports and create instructional interventions to improve student outcomes. We are developing a web portal that will allow any faculty to obtain questions and upload their students' responses for analysis.

Outcomes: Our research has led to new insights into students' struggles with key concepts, such as the Central Dogma of Biology, with FLC faculty collectively creating new instructional materials to address these challenges. Research on the FLC members shows faculty are moving from asking "How many students got the right answer?" to reflecting on student thinking and modifying instruction to address common learning challenges. We continue to explore a variety of lexical analysis and classification techniques to speed up and improve the development of questions and analytic resources. We are working on the web portal to completely automate report generation and make these analyses widely available to participating FLCs.

Broader Impacts: We created a set of FLCs across multiple institutions that engage STEM faculty teaching foundational courses to administer AACR questions, reflect on the results and implement revised instruction. In this current year we have added additional faculty to each FLC.   We are expanding questions development beyond biology into chemistry, statistics, thermodynamics, and physics/astronomy. We have presented at conferences, published several papers, and created two web sites to disseminate our results.

Introduction

Developing robust measures of student thinking about core scientific ideas is a challenge that may be too complex to accomplish via multiple-choice assessments such as concept inventories (CIs). [1]. Moreover, multiple-choice CIs introduce significant validity threats as they are constrained to "either-or" forced-choice ("misconception" vs. scientific key concept) item preference, and do not typically allow the detection of students who harbor "mixed models" of correct and incorrect conceptions [1-3]. They are also constrained in that they can only evaluate students' ideas about individual concepts, so they may not be up to the task of evaluating how students connect ideas - important to development of disciplinary expertise [4]. Constructed response (CR) assessments that capture students' explanatory models are needed to mitigate the constraints and reveal students' mixed models. CR assessments, for which students have to use their own language to demonstrate knowledge, are widely viewed as providing greater insight into student thinking than closed form (e.g., multiple-choice)

assessments [5]. Until now, financial and time constraints made CR assessments challenging to execute in large-enrollment courses than closed form assessments.
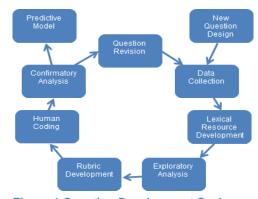
In the Automated Analysis of Constructed Response (AACR) project, a collaborative project among 7 institutions, we employ cutting-edge lexical analysis, machine learning, and statistical technology to develop conceptual constructed response assessments and build computer models that predict how experts would score student responses. Faculty administer the questions to students as online homework then we provide them with reports that summarize the ideas present in their students' writing, showing the distributions of both correct and alternative conceptions.

Methodological overview of the AACR approach

Our approach to developing, validating and implementing AACR assessments is captured by the Question Development Cycle (QDC) shown in Figure 1. In general, we use linguistic feature-based methods [6] to extract words and phrases from students' writing and then use those linguistic features as variables in statistical models that predict human raters' scores of the students' writing.

Figure 1 Question Development Cycle

In the first stage of the QDC, we *Design New Questions* to measure student thinking about important disciplinary constructs that are important and/or challenging for students to learn. *Data Collection* is done by administering the questions via online course management systems. *Lexical Resource Development* is done using lexical analysis software to extract key terms and scientific concepts from the students' writing. These terms and concepts are used as variables for *Exploratory Analysis* which aid in *Rubric Development*. We use the rubrics for *Human Coding* of student responses. During *Confirmatory Analysis,* the *Lexical Resources* are used as dependent variables in statistical classification techniques to predict expert human coding of student responses based on scoring models trained using human scored data. The entire process is iterative, with feedback from the various stages informing the refinement of other components. The final product of the QDC is a *Predictive Model* that can be used to completely automate the scoring of a new set of student responses, predicting how experts would score the responses. These models are used to generate reports for faculty that summarize the ideas present in their students' responses.

If you build it, they will (not necessarily) come

3

Making AACR assessments available is a necessary, but insufficient, condition for faculty adoption. Because STEM faculty at universities rarely have formal training in teaching or learning theory they need explicit support for their conceptual change to facilitate reformed teaching. Critical features of professional development (PD) programs that successfully promote change among faculty [7] include: 1) An extended period of professional development; 2) performance evaluation and feedback; and 3) a focus on changing faculty conceptions about teaching and learning.

Key to the AACR project are disciplinary *Faculty Learning Communities* (FLCs) to support faculty who are interested in new methods of assessment and willing to use AACR questions and reports to inform their teaching. We are building sustainability by:
  1) Creating local FLCs for PD to support the use of AACR assessments at each AACR collaborating institution.
  2) Connecting the local FLCs in a cross-institutional virtual community of instructors who use AACR questions and share materials, where support emerges from the community itself.
  3) Laying the foundation for expanding the FLC network in the future, providing a roadmap for structuring support of transformed teaching and learning.

The structure of the entire project is shown in Figure 2. The FLCs at participating institutions (Michigan State University, University of Georgia, University of South Florida, University of Maine, University of Colorado Boulder, and SUNY Stony Brook) are represented by the six larger circles arrayed about the FLC "hub." Each FLC consists of the local PI and participating faculty at that institution. Participating faculty (represented by the small red circles in each FLC circle in Figure 2) in the FLCs use the AACR assessments in their courses and receive feedback reports to inform their teaching. FLCs are supported locally at each institution by the local PI (small blue circles in Figure 2), administrator(s) (small green circles in Figure 2) and regular meetings. FLC meetings include discussions of AACR questions and reports and more general teaching and learning topics. FLCs are connected in a larger cross-institutional community of practice via face-to-face meetings and virtual meetings focused on developing curricula related to AACR questions. This larger community is supported by a website and email lists that allow resource sharing and discussion relating to teaching practice and AACR assessments.
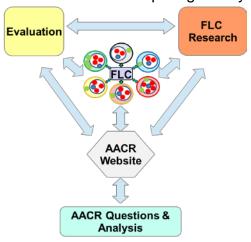

Figure 2: Logical structure of AACR: FLCS, Virtual Community of Practice Website, FLC Research and Evaluation

The community of all local FLCs is the central component in the overall project structure. FLCs communicate about the AACR Questions and Analysis (shown in the light green rectangle) to exchange data and receive feedback reports. FLCs also interact directly with the FLC research hub (shown in orange). Research on faculty adoption and institutional change are investigated by studying the interactions and FLC activity. Finally, project evaluators (shown in yellow) directly interact with the FLCs and the research hub to provide feedback on the project progress.

Overview of key outcomes

To date, we have retained our original 19 FLC faculty members and have added an additional 12 faculty across the project. In academic year 2014-2015, we held 33 local FLC meetings; faculty administered 71 AACR questions generating a total of 24,948 student responses. We generated 123 reports for faculty about their students' responses. These assessments and corresponding reports and local FLC meetings have shown promise in impacting student learning. For example, we have a set of "stop codon" questions that have been used extensively by faculty for pre/post-testing; the FLC participants have developed an instructional activity to address learning challenges identified in these AACR questions. The learning gains in the spring 2015 semester were double what they were during the previous semester. Performance on common exam questions was between 18-38% higher when compared to the fall.

Our research on FLC participants addresses two broad questions: 1) What is the best way to support the teaching professional development of the faculty who use AACR questions and reports? 2) How does AACR participation impact teaching practices as well as attitudes and thinking about teaching? Using Expectancy Value Theory (EVT) as a lens to address research question 1, we have found that faculty participants' expectancies for FLCs are high due to the ease of participation. Faculty participants also attach high value to the FLCs, because the FLCs provide knowledge about how to use AACR, an enjoyable forum for exchanging ideas about teaching with peers, and practical guidance about questions and concerns about their teaching.

During the 2014-15 academic year, we used the Classroom Observation Protocol for Undergraduate STEM (COPUS) to document instructional practices of FLC faculty members on multiple days [8]. The COPUS data were used to characterize FLC instructional styles based on a model developed by Lund et. al. [9] describing courses structured primarily as Lecture, Socratic, Peer Instruction, and Collaborative Learning. The AACR FLC faculty represented all four of these instructional styles, with many also occurring as hybrids. Understanding this diversity in teaching practices allows us to monitor the use of, and the supports needed for, the implementation of AACR questions, and to determine whether faculty teaching practices change.

We are developing a web portal that will allow faculty to obtain AACR questions, administer them to their students, and upload the responses to receive reports automatically. See the Evograder web site as an example of our portal plans [10]. We are also interested in engaging more engineering thermodynamics instructors to develop constructed response questions regarding energy balance concepts.

For additional information see the papers *A Community of Enhanced Assessment Facilitates Reformed Teaching* for information about the FLCs, *An Iterative Approach To Developing, Refining And Validating Machine-Scored Constructed Response Assessments* for information about the QDC, *Building Next-Generation STEM Assessments using Machine Learning Methodologies* for information about machine learning models for analyzing text, and the poster *Modeling Student Thinking in STEM: Insights from the Automated Analysis of Constructed Response (AACR) Project* for information about what the AACR analysis reveals about student thinking.

1. Nehm, R.H. and I.S. Schonfeld, *Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview.* Journal of Research in Science Teaching, 2008. **45**(10): p. 1131-1160.
2. Nehm, R.H. and L. Reilly, *Biology majors' knowledge and misconceptions of natural selection.* BioScience, 2007. **57**(3): p. 263 - 272.
3. Haudek, K.C., et al., *What are they thinking? Automated analysis of student writing about acid/base chemistry in introductory biology.* CBE - Life Sciences Education, 2012. **11**(3): p. 283-293.
4. Smith, J.I. and K. Tanner, *The problem of revealing how students think: Concept inventories and beyond.* CBE Life Sci Educ, 2010. **9**(1): p. 1-5.
5. Birenbaum, M. and K.K. Tatsouka, *Open-ended versus multiple-choice response formats - It does make a difference for diagnostic purposes.* Applied Psychological Measurement, 1987. **11**: p. 329-341.
6. Deane, P., *Strategies for evidence identification through linguistic assessment of textual responses*, in *Automated scoring of complex tasks in computer-based testing*, D.M. Williamson, I.I. Bejar, and R.J. Mislevy, Editors. 2006, Lawrence Erlbaum Associates: Mahwah, N. J. p. 313-372.
7. Henderson, C., A. Beach, and N. Finkelstein, *Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature.* Journal of Research in Science Teaching, 2011. **48**(8): p. 952-984.
8. Smith, M.K., et al., *The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices.* CBE-Life Sciences Education, 2013. **12**(4): p. 618-627.
9. Lund, T.J., et al., *The best of both worlds: Building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice.* CBE-Life Sciences Education, 2015. **14**(2).
10. Moharreri, K., M. Ha, and R.H. Nehm, *EvoGrader: an online formative assessment tool for automatically evaluating written evolutionary explanations.* Evolution: Education and Outreach, 2014. **7**(15): p. 14.

Author Biographies: co-authors are listed in alphabetic order after the first author

Mark Urban-Lurain, AACR Project PI. www.msu.edu/~urban Dr. Urban-Lurain is an Associate Professor and Acting Director of the Center for Engineering Education Research in the College of Engineering at Michigan State University. His research interests are in theories of cognition, how these theories inform the design of instruction, how we might best design instructional technology within those frameworks, and how the research and development of instructional technologies can inform our theories of cognition. He is also interested in preparing future STEM faculty for teaching, incorporating instructional technology as part of instructional design, and STEM education improvement and reform. Much of his research has focused on incorporating technology in the context of instructional design and using technology to provide assessments for formative feedback in the improvement of instruction.

Andrea Bierema, Michigan State University. Dr. Bierema is a research associate in the Center for Engineering Education Research, College of Engineering, Michigan State University. Her training is in both biological and science education research. Research interests include undergraduate student conceptions and sense-making of fundamental biological concepts, the portrayal of fundamental biological concepts in curriculum, and student group discourse.

Kevin Haudek, Michigan State University. Dr. Haudek is an instructor in the Department of Biochemistry and Molecular Biology. His broad research interests are strategies to increase student writing and feedback in undergraduate STEM courses and in student application of chemistry knowledge in biological contexts, such as how students use fundamental chemistry concepts to explain biological processes. He also has research interests in how assessment item construction and features influence student writing and explanations.

Anne-Marie Hoskinson, Michigan State University AACR Research Associate. Dr. Hoskinson is a research associate in the Center for Engineering Education Research, College of Engineering, Michigan State University. She was trained in mathematical modeling in biology and earned her Ph.D. in conservation biology. Her research focuses on how undergraduate students develop expertise in biological concepts, especially the role of scientific practices in the development of biological expertise.

Jennifer Kaplan, UGA PI. Dr. Kaplan is an Associate Professor in the Department of Statistics at the University of Georgia. Her research interests are undergraduate student learning in statistics, the pedagogical and content knowledge needs of instructors of statistics, particularly of GTAs, and the types of professional development that will lead to gains in knowledge for both instructors and students.

Jennifer Knight, University of Colorado Boulder PI. Dr. Knight is an Associate Professor in the Department of Molecular Cellular and Developmental Biology (MCDB). She has a Ph.D. in Neuroscience, and previously worked as a developmental biologist and geneticist. Her research now focuses on developing and using active learning materials and concept assessments, and studying the factors that influence students' in-class discussion. Dr. Knight coordinated the MCDB Science Education Initiative for 7 years, and is actively involved in CU's Center for STEM Learning, as well as other national organizations devoted to science education research.

Paula Lemons, UGA PI and Co-PI. http://sites.bmb.uga.edu/lemonslab/ Dr. Lemons is an Associate Professor in the Department of Biochemistry and Molecular Biology at the University of Georgia. Her research interests are in faculty development, with a focus on the process by which faculty change their teaching beliefs and practices while engaged in activities like faculty learning communities. She also studies problem solving among biology undergraduates, focusing on students' application of threshold concepts in biochemistry to problems involving visual representations.

Carl T. Lira, Associate Professor of Chemical Engineering at Michigan State University, integrates computer technology into classroom learning through assignments, interactive classroom activities, clickers. His participation on the project concerns the development of constructed response questions in engineering, initially focusing on energy concepts.

Jill McCourt, University of Georgia. Dr. McCourt is a Postdoctoral Associate in the Department of Biochemistry and Molecular Biology. She is an active participant in the AACR project. Her primary research interests focus on how faculty change both their beliefs about teaching and learning and their teaching practices in response to participation in teaching-related professional development.

John E. Merrill, Michigan State University, PI and Co-PI. Dr. Merrill is Associate Professor of Microbiology and Molecular Genetics (College of Osteopathic Medicine) and Director of the Biological Sciences Program (College of Natural Science). Primary research interests include assessment of student learning in foundational undergraduate biology courses. Previous work on concept inventory type assessment instruments led to an interest in finding better ways to explore student thinking about important biological concepts.

Rosa A. Moscarella, Michigan State University AACR Research Associate. Dr. Moscarella is a Research Associate at the Center for Engineering Education Research in the College of Engineering at Michigan State University. Formally trained as a biologist, she is interested in understanding students' learning obstacles in biology and more specifically in genetics. Her research focuses on three main aspects: 1) developing assessments and diagnostic tools that better reveal students' thinking in biology, 2) understanding the basis of students' learning difficulties and misconceptions in biology, and 3) designing a learning progression for college genetics.

Ross Nehm, Stony Brook University PI. Dr. Nehm is Associate Professor of Ecology & Evolution and Associate Director of the Ph.D. Program in Science Education. He studies student thinking about biological concepts such as natural selection and evolution. Additional work has examined novice and expert reasoning strategies, psychometric evaluation of education instruments, science teacher belief revision and professional development, conceptual structuring of scientific understanding, and the comparative efficacy of educational innovations. Currently, several projects are focusing on developing and evaluating machine-learning models for automated assessment of complex scientific practices, such as biological explanations.

Karen N. Pelletreau, University of Maine AACR Research Associate. Dr. Pelletreau is a research associate in the School of Biology and Ecology at the University of Maine. Her research interests include exploring new ways to promote and facilitate faculty collaboration and discussion of new teaching practices and how to effectively support faculty in developing and implementing active based teaching instruments for large enrollment courses.

Luanna Prevost, University of South Florida, PI. Dr. Prevost is an Assistant Professor in the Department of Integrative Biology at the University of South Florida. She is interested in exploring undergraduate student thinking in biology. Her research employs written assessment, automated analysis tools, and game design to explore student understanding of biology. She is also interested in how these approaches can be used to foster active learning environments in undergraduate biology classrooms.

Michelle Smith, UMaine PI, http://umaine.edu/center/directory/faculty-page/michelle-smith/ Dr. Smith is an Assistant Professor in the School of Biology and Ecology at the University of Maine and holds the C. Ann Merrifield Professorship in Life Sciences Education. Her research laboratory engages undergraduate

and graduate students, postdocs, K-12 teachers, and university faculty in research on teaching and learning. Together they focus on: 1) developing tools to understand student conceptual difficulties and conduct classroom observations, 2) studying what aspects of peer discussion make it an effective learning tool, and 3) understanding what factors influence faculty members' decisions about teaching.

Matthew M. Steele, Michigan State University AACR Research Associate. Dr. Steele is a Research Associate at Michigan State University's Center for Engineering Education Research. His research is focused on the development of tools and methods to support interactive and student-centered pedagogies in online and blended classroom environments. Specifically, he is interested in creating tools to allow rapid individualized feedback and foster collaborative discussion in large enrollment physics and astronomy courses.

Mary Anne Sydlik is the Director of the Science and Math Program Improvement (SAMPI) Center, an outreach division of Western Michigan University's Mallinson Institute for Science Education. SAMPI specializes in evaluation, research, and technical assistance for higher education institutions and K-12 schools. She is the external evaluator for AACR III. Dr. Sydlik has been the lead external evaluator for a number of STEM and NSF-funded projects. Her interests are in adding to efforts to improve the educational experiences and outcomes of undergraduate STEM students.