



# Open Student Modeling Research and its Connections to Educational Assessment

Diego Zapata-Rivera<sup>1</sup> 

Published online: 6 October 2020

© International Artificial Intelligence in Education Society 2020

## Abstract

Research in the area of Open Student Models (OSMs) has shown that external representations of the student model can be used to facilitate educational processes such as student reflection, knowledge awareness, learning, collaboration, negotiation, and student model diagnosis. OSMs can be integrated into existing learning systems or become a framework for the creation of learning systems. This paper discusses how early work with Jim Greer in the area of open student modeling has inspired and continues to inspire a line of research on innovative assessments and the design and evaluation of score report systems that are used to share assessment/student modeling information with various educational stakeholders to support learning. Several projects are discussed as well as their connections to my work with Jim while working at the ARIES laboratory.

**Keywords** Open student models · Score reporting systems · Educational assessment · Indirectly visible Bayesian student models · Evidence-based interaction · Privacy and data security

## Introduction

Different types of external representations (e.g., graphical and verbal representations) and interaction approaches have been implemented and evaluated in the area of Open Student Models (OSMs). Strategies for interacting with OSMs include free exploration, guided exploration, negotiation with a human or the system (e.g., with a virtual tutor), and collaboration with a human or a virtual peer (Bull and Kay 2007, 2016). My work with Jim Greer in the late 90s, early 00's gave me the opportunity to explore many aspects of this area of research including student modeling representations, inference mechanisms, guidance mechanisms, technical architectures and server modules used

---

✉ Diego Zapata-Rivera  
DZapata@ets.org

<sup>1</sup> Educational Testing Service, 661 Rosedale Road, Princeton, NJ 08540, USA

for implementing distributed student modeling applications, and intelligent tutoring systems that involve the use of OSMs. My dissertation on learning environments based on inspectable student models describes the range of topics explored during my years at the ARIES laboratory working with Jim and other members of the lab (Zapata-Rivera 2003).

This paper describes how work in the area of innovative assessments and score reporting systems have been influenced by my early work with Jim. Areas discussed include work on innovative approaches to interacting with student models: the indirectly visible Bayesian student model, and the evidence-based interaction with open student models; work on the design and evaluation of score reporting systems and issues related to privacy and data security.

I am grateful for having had the opportunity to learn from and share with Jim. His personal and professional advice was always valuable. My time at the ARIES laboratory was an exciting, engaging, and productive time. Jim and all the members of the lab created a diverse, challenging and stimulating environment that encouraged us to explore many interesting research topics.

### **Indirectly Visible Bayesian Student Models**

Designing modes of interacting with external representations of the student model requires thinking about who the user is. Individual differences may play a role on how students react to the information maintained by the system. One of the concerns is how students with low self-esteem may react to student model representations that show very low knowledge/skill levels. Work on different guidance mechanisms to interacting with OSMs can be useful in these situations (Dimitrova 2003; Dimitrova and Brna 2016; Zapata-Rivera and Greer 2002). In order to address this concern, an indirect approach to interacting with OSMs was devised and implemented in an assessment-based learning environment for English language learners called English ABLE (Zapata-Rivera 2007).

Several approaches to interacting with OSMs have been explored in the context of English language learning. For example, systems such as Mr. Collins (Bull and Pain 1995) implement different forms of negotiation of the contents of the student model with the system. The Notice OLM (Shahrouh and Bull 2008) allows students to compare the learner model and the system model side-by-side and decide what to do next (e.g., focus their learning on weak areas or answer additional questions to demonstrate their knowledge). Finally, the ‘See Yourself Write’ system (Bull 1997) and the ‘Do it Yourself’ modeling system (diyM; Bull 1998) allow students to negotiate the contents of the model with a teacher.

In this indirect approach the student is placed in a learning by teaching situation. That is, the student is asked to “teach” a pedagogical agent (Jorge or Carmen) by helping fix grammatical errors. This work was inspired by work on pedagogical agents (Chan and Baskin 1990; Graesser et al. 2001; Johnson et al. 2000, Johnson and Lester 2016) for learning including work on teachable agents (Biswas et al. 2001). However, in this work the knowledge levels of the pedagogical agent are also the student’s knowledge levels, which are made available so the student can see if the pedagogical agent is making progress (or not). The student also received feedback from a second pedagogical agent (Dr. Grammar). Jorge (or Carmen based on student selection)

exhibited some emotional states and associated utterances based on performance levels (Zapata-Rivera et al. 2007a, 2007b). The goal was to help attenuate possible negative reactions while keeping the student focused on helping the pedagogical agent learn. English ABLE included an internal Bayesian student model that was used to implement adaptive sequencing of activities and adaptive feedback mechanisms. Figure 1 shows a screenshot of English ABLE.

English ABLE was one of the first adaptive English learning systems that made use of a database of retired TOEFL items in a learning context. Available Item Response Theory (IRT; Lord 1980) parameters for these items (2PL: difficulty and discrimination) were calibrated with data from test-takers who shared similar characteristics as the intended learners and used to generate the conditional probability tables of the internal Bayesian student model (Chrysafiadi and Virvou 2013; Pandarova et al. 2019; Slavuj et al. 2017). More information about the Bayesian student model implemented in this system can be found in Zapata-Rivera (2007) and Zapata-Rivera et al. (2007b).

The evaluation of the system included 149 English language learners from English as a second language centers at universities and a high school in the US (Zapata-Rivera et al. 2007b). Students were randomly assigned to one of three conditions based on increasing levels of sophistication of the system (Condition 1 – control: test preparation with immediate verification feedback and a fixed sequence of items; Condition 2 – English ABLE simple: with a simple student model based on performance on the last three tasks, immediate verification feedback and limited interaction with pedagogical agents; and Condition 3 –

The screenshot displays the English ABLE interface. At the top, the title "English ABLE" is shown in a blue banner, with a "Ver en Español" button on the right. Below the title, a speech bubble from a male avatar says, "I don't understand how to make the verb agree with the rest of the sentence." A progress bar shows "Sentence Number 4", "Total Sentences 4", and "Total Correct 1". A "Review Previous Sentences" button and a "Next Sentence" button are also visible. The main text area contains the sentence: "Pollution ~~created~~ by humans ~~have~~ been a topic of interest by environmentalists for years." The word "created" has a red 'X' above it, and "have" has a green checkmark above it. Below "have" is the word "has" in red. A "Suggested word:" input field is empty. On the left, a "Knowledge Levels" sidebar shows "Details" and "Low" and "High" options. On the right, a "Dr. Grammar" character is shown with an "Ask for a Hint" button. A large text box at the bottom provides feedback and grammar rules for Verb-Noun, Verb-Pronoun, and Verb-Verb agreement.

Fig. 1 A screenshot of English ABLE. The student is helping Jorge (upper left) fix grammar errors while receiving help from Dr. Grammar (bottom right)

English ABLE enhanced: with a visible student model, immediate verification and adaptive instructional feedback, adaptive item sequencing, and pedagogical agents). Results showed that students assigned to conditions that included pedagogical agents and the OSM (English ABLE simple and enhanced) significantly outperformed those assigned to the control condition (test preparation condition with verification feedback and a predefined sequence of items) in terms of learning gains. These results showed that English ABLE had potential for supporting student motivation and learning. A longer study (i.e., in this study students interacted with the system for only one hour) with additional types of tasks (e.g., constructed response tasks) could provide additional insights on the value of the adaptive features of the system.

Results of a usability questionnaire showed that most students indicated that they understood the knowledge levels, thought that they had learned by using the system, liked helping Jorge/Carmen (when available), and found the feedback provided by Dr. Grammar useful (when available) (Zapata-Rivera 2007). Students' comments suggested that they were engaged in the process of helping Jorge/Carmen. Some students' comments showed an emotional connection with the characters. For example, "My Carmen is happy. Her knowledge levels are increasing," or "Poor Carmen, she is not learning a lot from me." Some aspects of the system that could be improved include adding dialogue capabilities to the agents so they could engage in conversations with the students, include more agents (e.g., agents with different ethnicities), and include various types of activities (e.g., interactive constructed response tasks).

This approach has the potential for engaging students in productive learning sessions. Assessment information was used to guide learning and relevant feedback (i.e., immediate verification and adaptive instructional feedback) was provided to students. When I told Jim about this work, he said this was a clever approach to deal with some key issues in the area of OSMs (e.g., dealing with potential negative effects of students exploring the student model and engaging students in an activity that involves providing help while receiving feedback from the system). He was glad to see that many of the concepts that we explored in the past were part of this assessment-based learning system (e.g., Zapata-Rivera and Greer 2002; Zapata-Rivera and Greer 2004a).

Jim and I reconnected at the invited Paul G. Sorenson Distinguished Graduate Lecturer event in 2013, where I presented some of this work as well as work on reporting systems. I remember that we talked about my current work in the area of assessment-based learning environments and innovated reporting systems. His advice was encouraging. He mentioned that formative adaptive learning systems that make use of assessment data to provide relevant, actionable feedback to teachers and students can benefit many learners. This line of research continues to inform our current work on the design and evaluation of adaptive learning systems. This work includes the design and evaluation of conversation-based assessments (Zapata-Rivera et al. 2015) to assess English language skills in naturalistic contexts that require the integration of various language skills (e.g., Forsyth et al. 2018; So et al. 2015), English language and mathematical skills (e.g., Lopez et al. 2017;), and science inquiry skills (Zapata-Rivera et al. 2014). Also, work on games that make use of agents to help students practice English language pragmatics (Yang and Zapata-Rivera 2010).

## EVIDENCE-BASED INTERACTION WITH OSMs

Inspired by previous work on interacting with Bayesian student models (Zapata-Rivera and Greer 2004a), we created an open student modeling approach for assessment-based learning systems called evidence-based interaction with open student models (EI-OSM) (Zapata-Rivera et al. 2007a). In this approach, a graphical interface based on Toulmin's argument structures (Toulmin 1958) was designed to allow teachers and students to engage in a collaborative, continuous assessment process, in which students and teachers could react to the current state of the student model (e.g., assessment claims or skills/knowledge levels) by providing alternative explanations and supporting evidence.

Evidence-based Interaction with OSMs (EI-OSM) is an example of an approach that makes available the state of the student model and some of the internal mechanisms used to infer and support assessment claims to teachers and students. A similar approach of using Toulmin's argument structures to share internal inferring mechanisms with students is described in Van Labeke et al. (2007). Bull (2016) reviews different approaches to negotiating the contents of the student model. In addition to Toulmin's based arguments patterns, approaches based on dialogue moves have been explored in the past (e.g., Dimitrova 2003; Bull and Pain 1995; Ginon et al. 2016; Kerly and Bull 2008). Given the amount of student data from different data sources used by adaptive learning environments and the need for explaining inferences made by these systems, it is expected that types of negotiated approaches will continue to be explored in the future.

In EI-OSM, the strength of each argument structure is calculated based on supporting evidence parameters such as its credibility, relevance, and quality. Types of supporting evidence and instructional materials are added to the system by the teacher. Students can use the learning materials and update the model based on additional pieces of evidence.

Figure 2 shows how ViSMod, a tool for interacting with Bayesian student models (Zapata-Rivera and Greer 2004a), can be used by a student to select a node with a low knowledge level ("Calculate slope from points") to be explored using EI-OSM. Figure 3 shows the node ("Calculate slope from points") in EI-OSM. An alternative explanation ("My knowledge level of 'Calculate Slope from points' is Medium") and new supporting evidence added by the student. EI-OSM was integrated in a mathematics adaptive system called Math Intervention Module (MIM; Shute 2006). This system provided the context for the evaluation of EI-OSM.

An evaluation of this approach with eight teachers was carried out. Teachers were presented with three evaluation scenarios: (a) exploring assessment claims and supporting evidence maintained by the system; (b) exploring assessment claims and supporting evidence added by a student; (c) and assigning credibility and relevance values and providing adaptive feedback. After that, they answered a usability and feasibility questionnaire and participated in a discussion session. Two interviewers and 1–3 teachers participated in each session.

Some of the main findings include: Teachers found ViSMod useful to identify the student's weaknesses and strengths. EI-OSM was found to be useful to inform next instructional actions for individuals or small groups of students (e.g., assign homework and plan instructional activities). Some of the teachers wanted the student to provide

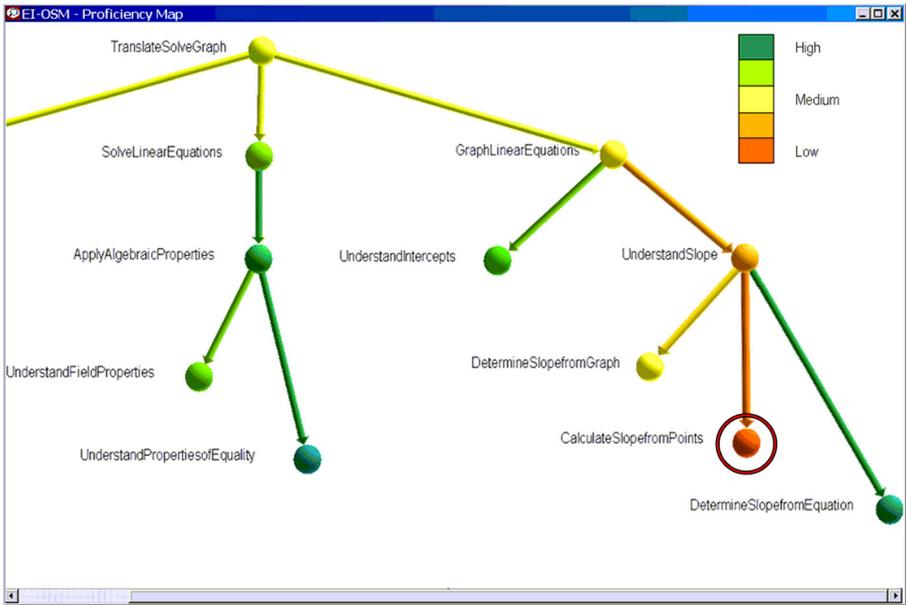


Fig. 2 The student selects “CalculateSlopefromPoints” for further exploration. Color is used to show current knowledge levels

Fig. 3 The student provides an alternative explanation (“My knowledge level of ‘Calculate Slope from points’ is Medium”) and supporting evidence

additional evidence before accepting the alternative explanation. However, once additional evidence was added, they accepted the alternative explanation. Teachers considered that this approach could help students become more active and participants and accountable for their learning. They expressed their interest in having the system to automatically assign credibility and relevance values based on sources of evidence and topic proximity. Teachers preferred the use of levels (e.g., high, medium, low) to numerical values.

Teachers also provided suggestions for making the interaction more effective given the lack of time teachers had to focus on each student. These suggestions include sending automated messages to inform teachers about situations that may require their attention and considering the role of teacher assistants who can help provide guidance to students. This is an important feature, since teachers may have limited time to interact with each student. Also, teachers may have additional information about students and would like to override the system's values in some cases (e.g., in cases they talk to the student and reach an agreement). More information about this study can be found in Zapata-Rivera et al. (2007a).

A related area of current research is the interpretation of complex models (e.g., interpretable machine learning algorithms). Intelligent learning and assessment systems collect a variety of data from students' interactions with the system. When making claims about student knowledge or skills it is important to have the evidence and argument structures necessary to show users how these claims are supported. This type of information can be used to provide users of adaptive applications with explanations about why particular recommendations are being suggested (Zapata-Rivera 2019).

Research in the area of OSMs can contribute to the challenges of improving interpretability to Artificial Intelligence (AI) systems (Conati et al. 2018). Providing appropriate mechanisms for humans to explore how the system makes use of student performance data to implement adaptive components (e.g., adaptive scaffolding and recommendations), can increase trust in AI educational systems and improve accountability on decisions made by AI educational systems. EI-OSM is a possible approach to provide humans with interpretable student model information that builds on the availability of assessment-based arguments in well-designed assessments (Mislevy 2012).

Work with Jim and others on visualizing Bayesian networks (Zapata-Rivera et al. 1999) and interacting with Bayesian student models (Zapata-Rivera and Greer 2004a, 2004b) focused on designing interpretable Bayesian structures, visualization techniques, graphical interfaces, and guidance mechanisms to facilitate teachers and students' interactions with OSMs. With the renewed interest in visualizing complex models, this research becomes relevant (Al-Shanfari et al. 2016; Champion and Elkan 2017).

## OSMs AND SCORE REPORTING SYSTEMS

Research in the areas of OSMs and score reporting systems share similar concerns. For example, a main concern in both areas is how to clearly communicate potentially complex student model/assessment information to different stakeholder groups (e.g., students, parents, administrators), who each comprehend the information from their unique viewpoints and for their unique purposes (Zapata-Rivera and Katz 2014).

Information in score reports usually include individual-, item- and classroom-level performance data based on the results of an assessment (e.g., a summative assessment) or a group of assessments (e.g., interim and formative assessments). They can include recommendations on what to do next (e.g., formative hypotheses, links to educational materials and ancillary materials with additional information about the report (e.g., appropriate uses and interpretation guidelines) (Zapata-Rivera et al. 2012). Active Reports describe a continuous assessment framework where educational stakeholders can interact with score report information and use it to support teacher instruction and student learning (Zapata-Rivera et al. 2005). In fact, EI-OSM builds on the notion of Active Reports (Zapata-Rivera et al. 2007a).

New types of assessments include computer-based simulations and other performance-based tasks that provide both student response and process data. Reports in this context can include information such as strategies used by students to solve problems (Zapata-Rivera et al. 2018). As more data are available from students' interactions with these data-rich assessments, the use of dashboards showing data analytics becomes more popular. Designing and evaluating these dashboards with the intended audience is an important area of research. Research in OSMs and reporting systems can inform this work (e.g., Bodily et al. 2018; Corrin 2018; Ez-Zaouia 2020; Feng et al. 2018; Jivet et al. 2018; Klerkx et al. 2017, Maldonado et al. 2012).

Research in the area of score reporting can offer several insights to the field of OSMs such as the use of development frameworks and standards to ensure clear understanding of assessment results by assessment users (Zapata-Rivera 2018a). Frameworks for designing and evaluating score reports propose an iterative approach that includes activities like gathering assessment information needs, reconciling these needs with assessment information that will be available to the audience, designing prototypes, and evaluating these prototypes with the intended audience at different stages of assessment development and after operational deployment (Hambleton and Zenisky 2013; Zapata-Rivera et al. 2012). The professional standards (AERA, APA and NCME 2014) include information such as the type of information that should be reported to users to support appropriate use of assessment results, and the quality criteria on scores and subscores that are required to support different types of decisions.

The design of score reports takes into account design principles from areas such as cognitive science, information visualization, and human computer interaction (Hegarty 2018). Score reporting systems can be designed and evaluated considering the characteristics of the intended users (i.e., needs, knowledge, and attitudes). Zapata-Rivera and Katz (2014) shows how audience analysis was applied to the design of score reports for administrators, teachers and students considering their needs, knowledge, and attitudes. For example, in the case of teachers, their *needs* include receiving assessment results that can be used to inform instruction (e.g., What are the students' strengths and weaknesses? How does a student's score compare to other students' scores in the class? How difficult were the tasks for students?). Regarding *knowledge*, teachers know their students, and have access to additional information about the context of learning educational resources, types of instructional activities that would be appropriate in a particular case. They may have limited knowledge about measurement error (Zapata-Rivera et al. 2016b; Zwick et al. 2014). In terms of *attitudes*, teachers may have limited time to explore/process data and may not be aware of appropriate uses for test scores (Zapata-Rivera et al. 2012).

Zapata-Rivera (2019) includes a table showing sample assessment information needs for various types of users (i.e., teachers, students, parents, administrators, policy makers, and researchers). Evaluation of score reporting systems may include small-scale studies such as cognitive labs (Kannan et al. 2018), usability studies (Zapata-Rivera and VanWinkle 2010; Zapata-Rivera et al. 2013), and focus groups (Underwood et al. 2007) as well as large-scale research studies evaluating users' comprehension and preference of the various elements included in the score reports (e.g., Zapata-Rivera et al. 2016b; Zwick et al. 2014). More information about evaluation methods for score reporting systems can be found in Hambleton and Zenisky (2013) and Zapata-Rivera (2018b).

Figures 4 and 5 show two screenshots of a reporting system designed for teachers. Information about the Mathematics performance level of a student is shown graphically. Figure 4 depicts the score, the corresponding performance level, a confidence bar and a footnote for the teacher mentioning that the score is subject to measurement error which should be taken into consideration when making decisions. Links to definitions and additional information are included in the report. Figure 5 describes performance expected at the current and next proficiency levels as well as information about possible next steps. The system also provides information about appropriate and inappropriate uses of this type of information (Zapata-Rivera et al. 2012). This work was done as part of the Cognitively based Assessments *as, for, and of* Learning (CBAL) initiative (Bennett 2010; Bennett and Gitomer 2009).

Figure 6 shows a report system designed for administrators. It includes an interactive question-based interface that was also designed based on the assessment information

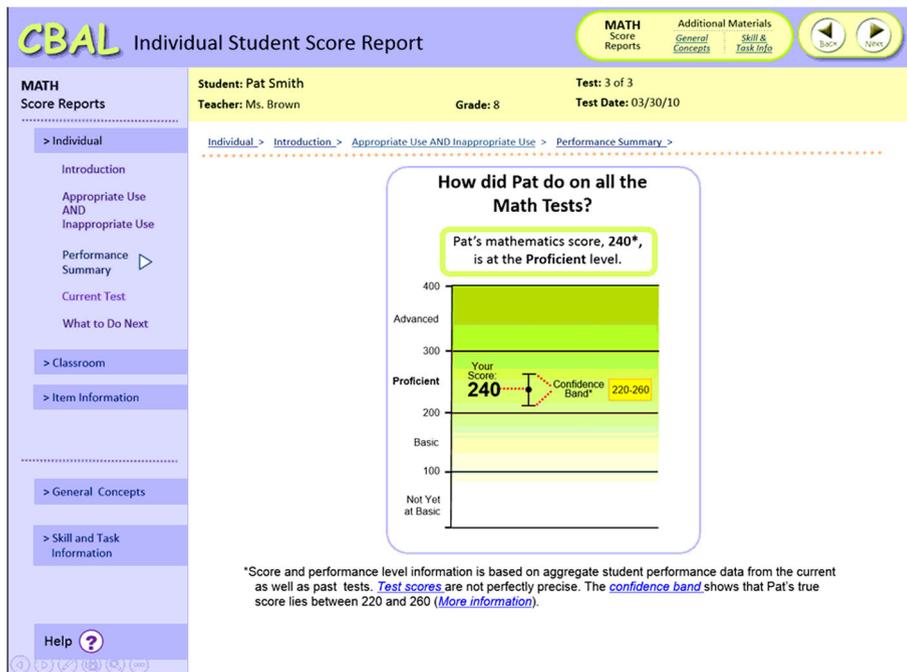


Fig. 4 A screenshot of an individual report for teachers depicting the proficiency level, score and confidence band

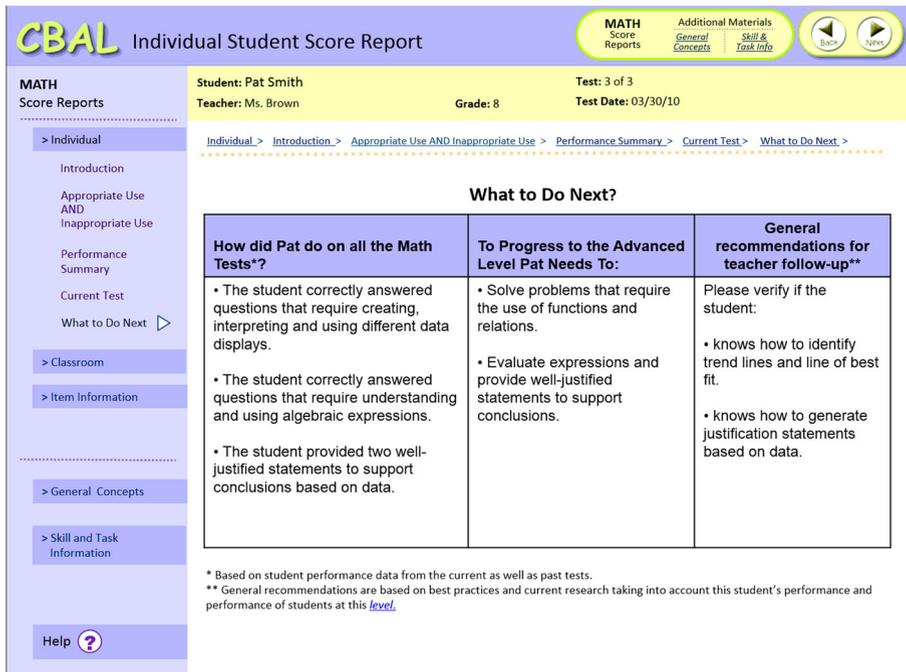


Fig. 5 Information on ‘What to Do Next’ from an individual score report for teachers

needs/responsibilities of administrators, their attitudes, and knowledge. The report shows the percentage and number of 8th grade students above and below proficient in the district level in several areas. The bar is clickable, and the table is automatically updated based on the proficiency level selected. The information is also aggregated by groups of interest in order to facilitate decision making. A textual description of the table including key findings and results is provided to facilitate understanding (Underwood et al. 2010).

Figure 7 shows a screenshot of a recent report system for teachers. This report shows information about a student who interacted with a conversation-based assessment of English language skills (So et al. 2015). This reporting system shows assessments results based on student response and process data. In addition to performance level, scores, and number of answers per categories (e.g. blank, correct, incorrect), this report includes information such as number of words produced, amount of scaffolding, sample responses and misconceptions. This information is intended to provide teachers with additional context to support instructional decisions.

This reporting system was evaluated and refined through several iterations of focus groups with teachers (3 iterations; 12 teachers). Teachers interacted with individual- and classroom-level reports. Teachers valued the availability of process data. They found the displays and features both easy to understand and useful. They provided suggestions such as adding explanations about relevance and use of particular features (e.g., number of words), changing the name of some of the features to facilitate comprehension (e.g., ‘number of words’ instead of ‘generation’), adding comparison

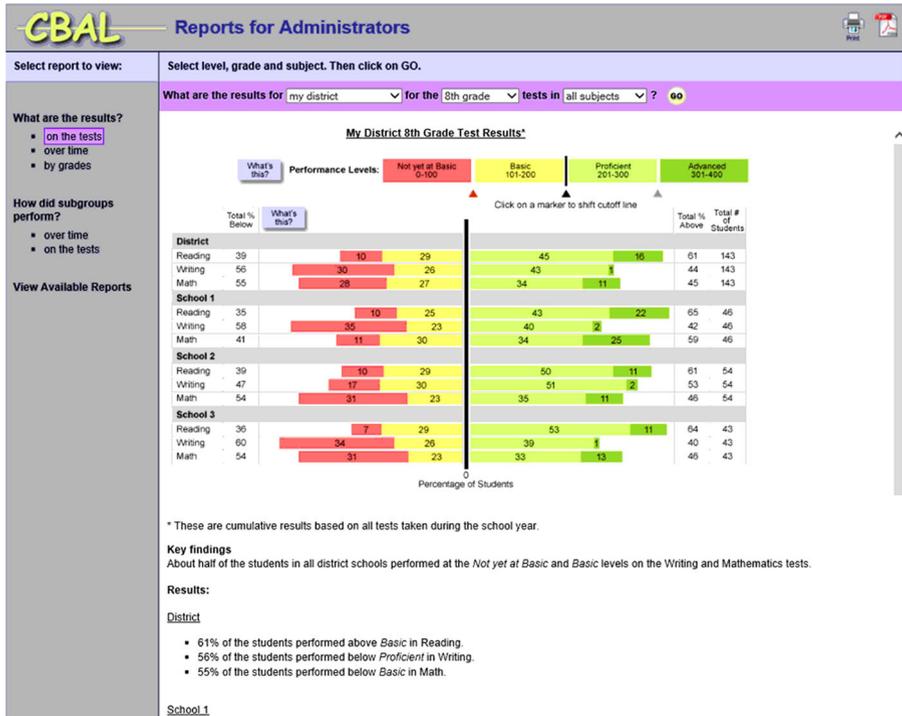


Fig. 6 An interactive question-based score report for administrators

data (e.g., class- and school-level indicators), and adding misconceptions (Peters et al. 2017).

Additional work in this area includes a dialogue-based tutor designed to teach teachers about measurement error provided in score reports (Forsyth et al. 2017). AutoTutor was the dialogue engine used to implement this tutorial (Graesser et al. 2001). This work builds on prior work on evaluating the effectiveness of using a video tutorial on measurement error to facilitate teacher understanding of score report information (Zapata-Rivera et al. 2016b).

My work with Jim on interacting with Bayesian student models has influenced our thinking in these projects (Zapata-Rivera and Greer 2004a, 2004b). For example, our early work on interacting with Bayesian student models included a variety of guidance mechanisms to support teacher and student interaction with OSMs (e.g., following a protocol, interacting with human peers and virtual guiding agents, interacting with the teacher, and exploring the model as a group). These mechanisms were designed to support student reflection, student model accuracy, and student learning. These work can be reflected in current work on designing and evaluating score reports that include the use of graphical and verbal representations to communicate information such as measurement error with teachers and parents as well as the use of innovative designs and guidance mechanisms to support appropriate use of assessment information (e.g., Kannan et al. 2018; Zapata-Rivera et al. 2016b; Zwick et al. 2014).

We continue to make progress on how to clearly communicate assessment results to different audiences. This work has implications for the validity of assessments since the

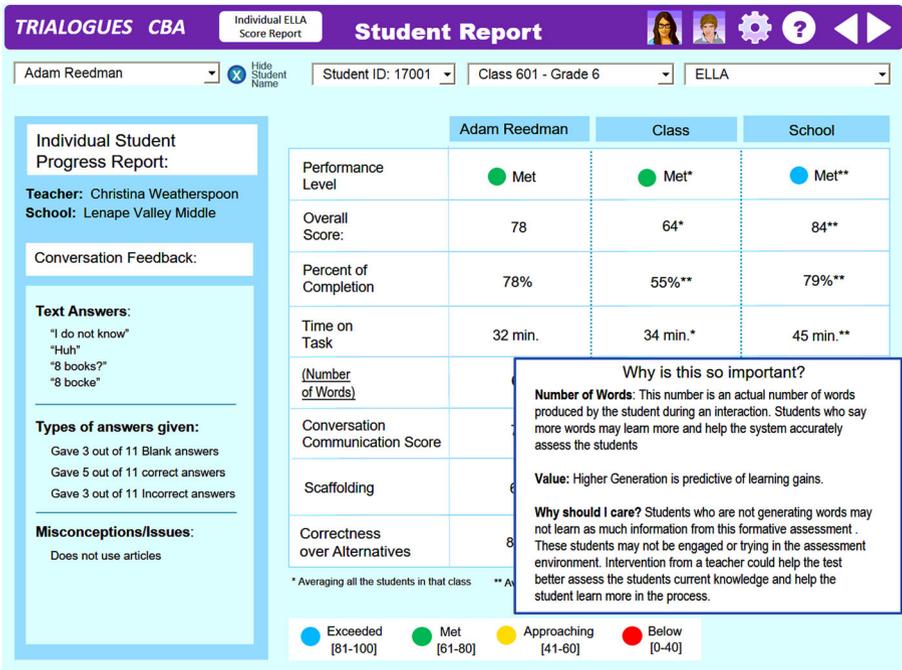


Fig. 7 A screenshot of an individual reporting system for teachers showing performance data of a student interacting with a conversation-based assessment

intended audience should clearly understand assessment results to make sound decisions based on this information (Zapata-Rivera et al. 2020).

### Privacy and Data Security

Privacy and data security concerns have been an important issue in the design and implementation of adaptive systems (Kobsa 2001). Our past work on student modeling servers includes the design and evaluation of a server called SModel (Zapata-Rivera and Greer 2004b). This server offered various services to agents in a distributed platform.

Using a Student Model Viewer (SMV) available in SModel, users could specify types of relationships with particular actors (e.g., persona agents, application agents, teachers, peers, and external agents), establish disclosure agreements with them, and grant them different levels of access to particular variables. In order to facilitate students’ control of student model information, every agent that wanted to access the student model needed to establish a disclosure agreement with the student. These agreements were used to specify privacy levels (e.g., whether a student model variable was private or not) and levels of access (e.g., permissions to read, update, and/or share the content of student model variables). This work is a good example of how systems could be created to facilitate human interaction with OSMs. These systems play an important role on addressing issues such as who owns the data and who has access and control on how personal data are collected and used by modern adaptive systems.

This early work includes some of the elements found in more recent work by Mohd Anwar and Jim in user-controlled identity management that makes it possible for users to control the type of information that will be shared depending on the context of the interaction. This approach allows for privacy customization in terms of purposes, roles, and relationships (Anwar and Greer 2012b). Anwar and Greer (2012a) describe an approach to facilitating trust by allowing the use of multiple identities as a mechanism for secure reputation transfer. These privacy approaches were implemented and evaluated in the I-Help System (Greer et al. 2001).

As new privacy laws are being introduced (e.g., GRDP 2018) and expected to spread to different countries, this line of research becomes even more relevant. Our work on privacy and security issues in the context of OSMs is also relevant to new types of assessments that make use of both process and response data to assess the strategies that students use to solve a problem (Zapata-Rivera et al. 2016a). Knowing how these data are used by the system to augment scores or provide relevant feedback to teachers and students is a current area of research (Ercikan and Pellegrino 2017; Zapata-Rivera et al. 2018). Research in the area of OSMs may provide useful insights to address these challenges.

Privacy and data security concerns are relevant to our current work on Caring Assessments (Zapata-Rivera 2017). Caring assessments consider additional information about the student (e.g., motivation, metacognition, and affect) and the learning context (e.g., learning opportunities) to create situations that students find engaging while maintaining the quality of the assessment. Challenges in this area include having access to information about the student and the educational context, developing tools that teachers, students, and parents can use to inspect and maintain their student model information, and implementing privacy and data security mechanisms so the information can be used for the intended purposes.

## Summary

I am indebted to Jim for his mentorship. As I've discussed, our early work together on OSMs—and his wisdom—guided my research and career to this day. Thanks to Jim, I've been able to expand on that work, researching new types of open student modeling approaches, the design and evaluation of assessment-based learning environments, work on innovative score reporting systems and current research in conversation and caring assessment systems.

Several areas for future research have been identified including: exploring the use of OSMs and score reporting systems in the design of systems (e.g., dashboards) that allow human interaction with complex inference models and large amounts of process and response data, and their role in facilitating the implementation of new privacy and data security laws.

Jim's work continues to inspire my current work in the areas of educational assessment and artificial intelligence in education. As assessments become more interactive and technology-rich, more opportunities for cross fertilization in the areas of artificial intelligence in education and educational assessment will occur.

**Acknowledgements** I would like to thank James Carlson, Priya Kannan, Blair Lehman, Irv Katz, and three anonymous reviewers for their comments on a previous version of the manuscript.

## References

- AERA, APA, & NCME. (2014). *American Educational Research Association, American Psychological Association, & National Council on measurement in education: Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Al-Shanfari, L., Epp, C. D. & Bull, S. (2016). Uncertainty in open learner models: Visualising inconsistencies in the underlying data. In *LAL @ LAK*. 23–30.
- Anwar, M., & Greer, J. (2012a). Facilitating trust in privacy-preserving e-learning environments. *IEEE Transactions on Learning Technologies*, 5(1), 62–73. <https://doi.org/10.1109/TLT.2011.23>.
- Anwar, M., & Greer, J. (2012b). Role- and relationship-based identity Management for Privacy-enhanced E-learning. *International Journal of Artificial Intelligence in Education*, 21(3), 191–213.
- Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (Cbal): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research & Perspective*, 8(2–3), 70–91.
- Bennett, R. E., & Gitomer, D. H. (2009). Transforming K–12 assessment: Integrating accountability testing, formative assessment and professional support. In C. Wyatt-Smith & J. Cumming (Eds.), *Assessment issues of the 21st century* (pp. 43–61). New York, NY: Springer.
- Biswas, G., Schwartz, D., Bransford, J., & the Teachable Agent Group at Vanderbilt (TAG-V). (2001). Technology support for complex problem solving: From SAD environments to AI. In K. D. Forbus & P. J. Feltovich (Eds.), *Smart machines in education: The coming revolution in educational technology* (pp. 71–97). Menlo Park: AAAI/MIT Press.
- Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: A systematic review. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 41–50). <https://doi.org/10.1145/3170358.3170409>.
- Bull, S. (1997). See yourself write: A simple student model to make students think. In A. Jameson, C. Paris, & C. Tasso (Eds.), *User modeling: Proceedings of the sixth international conference* (pp. 315–326). Wien New York: Springer.
- Bull, S. (1998). ‘Do it Yourself’ student models for collaborative student Modelling and peer interaction. In B. P. Goettl, H. M. Half, C. L. Redfield, & V. J. Shute (Eds.), *Intelligent Tutoring Systems* (pp. 176–185). Berlin, Heidelberg: Springer-Verlag.
- Bull, S. (2016). Negotiated learner modelling to maintain today’s learner models. *Research and Practice in Technology Enhanced Learning*, 11(10), 1–29.
- Bull, S., & Kay, J. (2007). Student models that invite the learner in: The SMILI open learner modelling framework. *International Journal of Artificial Intelligence in Education*, 17(2), 89–120.
- Bull, S., & Kay, J. (2016). SMILI©: A framework for interfaces to learning data in open learner models, learning analytics and related fields. *International Journal of Artificial Intelligence in Education*, 26(1), 293–331.
- Bull, S., & Pain, H. (1995). ‘Did I say what I think I said, and do you agree with me?’: Inspecting and questioning the student model. In J. Greer (Ed.), *Proceedings of world conference on artificial intelligence and education* (pp. 501–508). Charlottesville VA: AACE.
- Champion, C., & Elkan, C. (2017). Visualizing the consequences of evidence in Bayesian networks. *Champion, C., & Elkan, C. (2017). Visualizing the consequences of evidence in bayesian networks. arXiv preprint arXiv:1707.00791*.
- Chan, T. W., & Baskin, A. B. (1990). Learning companion systems. In C. Frasson & G. Gauthier (Eds.), *Intelligent Tutoring Systems: At the crossroads of AI and Education* (p. 6-33). Ablex pub..
- Chrysaifiadi, K., & Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications*, 40(11), 4715–4729.
- Conati, C., Porayska-Pomsta, K. & Mavrikis, M. (2018). AI in education needs interpretable machine learning: Lessons from open learner Modelling. *ArXiv, abs/1807.00154*. Corrin, L. (2018). Evaluating students’ interpretation of feedback in interactive dashboards. In D. Zapata-Rivera (Ed.), *Score reporting research and applications*. New York, NY: Routledge. 145–159.

- Corrin, L. (2018). Evaluating students' interpretation of feedback in interactive dashboards. In Zapata-Rivera (Ed.), *Score reporting research and applications* (pp. 145–159). New York, NY: Routledge.
- Dimitrova, V. (2003). StyLE-: Interactive open learner Modelling. *International Journal of Artificial Intelligence in Education*, 13(1), 35–78.
- Dimitrova, V., & Brna, P. (2016). From interactive open learner Modelling to intelligent mentoring: STYLE-OLM and beyond. *International Journal of Artificial Intelligence in Education*, 26(1), 332–349.
- Ercikan, K., & Pellegrino, J. W. (2017). Validation of score meaning using examinee response process for the next generation of assessments. In K. Ercikan & J. W. Pellegrino (Eds.), *Validation of score meaning for the next generation of assessments: The use of response processes* (pp. 1–8). New York, NY: Routledge.
- Ez-Zaouia, M. (2020). Teacher-centered dashboards design process. In 2nd international workshop on eXplainable learning analytics, *Companion Proceedings of the 10th International Conference on Learning Analytics & Knowledge LAK20*. <https://doi.org/10.35542/osf.io/p7cdv>.
- Feng, M., Krumm, A., & Grover, S. (2018). Applying learning analytics to support instruction. In D. Zapata-Rivera (Ed.), *Score reporting research and applications* (pp. 145–159). New York, NY: Routledge.
- Forsyth, C. M., Peters, S., Zapata-Rivera, D., Lentini, J., Graesser, A. C., & Cai, Z. (2017). Interactive score reporting: An AutoTutor-based system for teachers. In R. Baker, E. Andre, X. Hu, T. Rodrigo, & B. du Bouley (Eds.), *In proceedings of the international conference on artificial intelligence in education, LNCS* (pp. 506–509). Switzerland: Springer Verlag.
- Forsyth, C. M., Luce, C., Zapata-Rivera, D., Jackson, G. T., Evanini, K. & So, Y. (2018). Evaluating english language learners' conversations: Man vs. machine. *Computer Assisted Language Learning*, 1-20. <https://doi.org/10.1080/09588221.2018.151712>.
- Ginon, B., Boscolo, C., Johnson, M. D., and Bull, S. (2016). Persuading an open learner model in the context of a university course: An exploratory study. *Proceedings of the 13th International Conference on Intelligent Tutoring Systems*. 307–313.
- Graesser, A. C., Person, N., Harter, D., & TRG. (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, 12, 257–279.
- GRDP (2018). General data protection regulation. Art. 22 GDPR. Automated individual decision-making, including profiling. <https://gdpr-info.eu/art-22-gdpr/>. Accessed 11/30/19 2019.
- Greer, J., McCalla, G., Vassileva, J., Deters, R., Bull, S., & Kettel, L. (2001). Lessons learned in deploying a multi-agent learning support system: The I-help experience. *Proceedings International AI and Education Conference AIED'2001*, San Antonio, IOS press: Amsterdam, 410–421.
- Hambleton, R., & Zenisky, A. (2013). *Reporting test scores in more meaningful ways: A research-based approach to score report design*. *APA handbook of testing and assessment in psychology* (p. 479–494). Washington, DC: American Psychological Association.
- Hegarty, M. (2018). Advances in cognitive science and information visualization. In D. Zapata-Rivera (Ed.), *Score reporting research and applications* (pp. 19–34). New York, NY: Routledge.
- Jivet, I., Scheffel, M., Specht, M., & Drachler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 31–40). <https://doi.org/10.1145/3170358.3170421>.
- Johnson, W. L., & Lester, J. C. (2016). Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial Intelligence in Education*, 26(1), 25–36.
- Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11(1), 47–78.
- Kannan, P., Zapata-Rivera, D., & Leibowitz, E. A. (2018). Interpretation of score reports by diverse subgroups of parents. *Educational Assessment*, 23(3), 173–194. <https://doi.org/10.1080/10627197.2018.1477584>.
- Kerly, A., & Bull, S. (2008). Children's interactions with inspectable and negotiated learner models. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent Tutoring Systems* (p. 132–141).
- Klerkx, J., Verbert, K., & Duval, E. (2017). Learning analytics dashboards. In C. Lang, G. Siemens, A. F. Wise, & D. Gaevic (Eds.), *The handbook of learning analytics, pages 143–150*. *Society for Learning Analytics Research (SoLAR)* (1st ed.). Alberta.
- Kobsa, A. (2001). Tailoring Privacy to Users' Needs (Invited Keynote). In M. Bauer, P. J. Gmytrasiewicz, & J. Vassileva (Eds.), *User Modeling 2001: 8th international conference UM2001* (pp. 303–313). Berlin - Heidelberg: Springer Verlag.
- Lopez, A. A., Luce, C., Zapata-Rivera, D., & Forsyth, C. (2017). Using formative conversation-based assessments to support students' English language development. *IEEE Technical Committee on Learning Technology Bulletin*, 19(1), 6–9.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Mahwah: Lawrence Erlbaum Associates, Inc..

- Maldonado, R. M., Kay, J., Yacef, K., and Schwendimann, B. (2012). An interacEve teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In *Intelligent Tutoring Systems*. 482–492.
- Mislevy, R. J. (2012). *Four metaphors we need to understand assessment*. Commissioned paper for the Gordon Commission on the future of assessment in education. Princeton, NJ: Educational Testing Service. Retrieved April 28, 2020, from [www.ets.org/Media/Research/pdf/mislevy\\_four\\_metaphors\\_understand\\_assessment.pdf](http://www.ets.org/Media/Research/pdf/mislevy_four_metaphors_understand_assessment.pdf)
- Pandarova, I., Schmidt, T., Hartig, J., Boubekki, A., Jones, R. D., & Brefeld, U. (2019). Predicting the difficulty of exercise items for dynamic difficulty adaptation in adaptive language tutoring. *International Journal of Artificial Intelligence in Education*, 2019, 342–367.
- Peters, S., Forsyth, C. M., Lentini, J., & Zapata-Rivera, D. (2017, April). Score reports for conversation-based assessments: Identifying and interpreting evidence. *Presented at the annual American Education Research Association (AERA)*, Conference in San Antonio, TX.
- Shahrour, G., & Bull, S. (2008). Does 'notice' prompt noticing? Raising awareness in language learning with an open learner model. *Proceedings of the 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*. 173–182.
- Shute, V. J. (2006). Tensions, trends, tools, and technologies: Time for an Educational Sea change. Research report. ETS RR-06-16. ETS research report series.
- Slavuj, V., Meštrović, A., & Kovačić, B. (2017). Adaptivity in educational systems for language learning: A review. *Computer Assisted Language Learning*, 30(1–2), 64–90. <https://doi.org/10.1080/09588221.2016.1242502>.
- So, Y., Zapata-Rivera, D., Cho, Y., Luce, C., & Battistini, L. (2015). Using dialogues to measure English language skills. In J. García Laborda, D. G. Sampson, R. K. Hambleton, & E. Guzman (Eds.), *Journal of educational technology and society* (pp. 21–32). Technology Supported Assessment in Formal and Informal Learning: Special Issue.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: University Press.
- Underwood, J. S., Zapata-Rivera, D., & VanWinkle, W. (2007). *Growing pains: Teachers using and learning to use IDMS (research memorandum 08–07)*. Princeton: Educational Testing Service.
- Underwood, J. S., Zapata-Rivera, D., & VanWinkle, W. (2010). *An evidence-centered approach to using assessment data for policymakers (ETS research rep. No. RR-10-03)*. Princeton: ETS.
- Van Labeke, N., Brna, P., & Morales, R. (2007). Opening up the interpretation process in an open learner model. *International Journal of Artificial Intelligence in Education*, 17, 305–338.
- Yang, H. C., & Zapata-Rivera, D. (2010). Interlanguage pragmatics with a pedagogical agent: The request game. *Computer Assisted Language Learning*, 23(5), 395–412.
- Zapata-Rivera, J. D. (2003). *Learning Environments based on Inspectable Student Models*. Ph.D. Thesis. Saskatoon: Department of Computer Science. University of Saskatchewan.
- Zapata-Rivera, D. (2007). Indirectly visible Bayesian student models. In K. B. Laskey, S. M. Mahoney, & J. A. Goldsmith (Eds.), *Proceedings of the 5th UAI Bayesian Modelling applications workshop, CEUR Workshop Proceedings, Vol. 268*. 9 pp.
- Zapata-Rivera, D. (2017). Toward caring assessment systems. In *Adjunct Publication of the 25th conference on user modeling, adaptation and personalization (UMAP '17)*. ACM, New York, NY, USA, 97–100.
- Zapata-Rivera, D. (2018a). Why is score reporting relevant? In D. Zapata-Rivera (Ed.), *Score reporting research and applications* (pp. 1–6). New York: Routledge.
- Zapata-Rivera, D. (Ed.). (2018b). *Score reporting research and applications*. New York, NY: Routledge.
- Zapata-Rivera, D. (2019). Supporting human inspection of adaptive instructional systems. In R. Sottilare & J. Schwarz (Eds.), *Adaptive instructional systems. HCI 2019, Lecture Notes in Computer Science, vol (Vol. 11597, pp. 482–490)*. Cham: Springer.
- Zapata-Rivera, D., & Greer, J. (2002). Exploring various guidance mechanisms to support interaction with Inspectable learner models. In *Proceedings of Intelligent Tutoring Systems ITS 2002*. Pp. 442–452.
- Zapata-Rivera, J. D., & Greer, J. (2004a). Interacting with Bayesian student models. *International Journal of Artificial Intelligence in Education*, 14(2), 127–163.
- Zapata-Rivera, J. D., & Greer, J. (2004b). Inspectable Bayesian student modelling servers in multi-agent tutoring systems. *International Journal of Human Computer Studies*, 61(4), 535–563.
- Zapata-Rivera, D., & Katz, I. R. (2014). Keeping your audience in mind: Applying audience analysis to the design of score reports. *Assessment in Education: Principles, Policy & Practice*, 21, 442–463.
- Zapata-Rivera, D., & VanWinkle, W. (2010). *A research-based approach to designing and evaluating score reports for teachers. ETS research memorandum no.RM-10-01*. Princeton: ETS.
- Zapata-Rivera, J. D., Neufeld, E., & Greer, J. (1999). Visualization of Bayesian belief networks. *IEEE Visualization 1999 Late Breaking Hot Topics Proceedings*. 85–88.

- Zapata-Rivera, D., Underwood, J. S., & Bauer, M. (2005). Advanced reporting Systems in Assessment Environments. In J. Kay, A. Lum, & J.-D. Zapata-Rivera (Eds.), *Proceedings of Workshop on Learner Modelling for Reflection, to Support Learner Control, Metacognition and Improved Communication between Teachers and Learners (pp. 23-32)*, 10th international conference on artificial intelligence in education. Sydney.
- Zapata-Rivera, D., Hansen, E. G., Shute, V. J., Underwood, J. S., & Bauer, M. I. (2007a). Evidence-based approach to interacting with open student models. *International Journal of Artificial Intelligence in Education.*, 17(3), 273–303.
- Zapata-Rivera, D., VanWinkle, W., Shute, V., Underwood, J., & Bauer, M (2007b). English ABLE. In artificial intelligence in education - building technology rich learning contexts that work. Luckin, R., Koedinger, K., & Greer, J. (Eds.) Vol. 158. 323–330.
- Zapata-Rivera, D., VanWinkle, W., & Zwick, R. (2012). Applying score design principles in the Design of Score Reports for CBAL™ teachers. In *ETS research memorandum RM-12-20*. Princeton: ETS.
- Zapata-Rivera, D., Vezzu, M., & K. Biggers (2013). Supporting teacher communication with parents and students using score reports. Paper presented at the annual meeting of the American Educational Research Association (AERA), San Francisco.
- Zapata-Rivera, D., Jackson, T., Liu, L., Bertling, M., Vezzu, M., and Katz, I. R., (2014). Science inquiry skills using Trialogues. *12th International conference on Intelligence Tutoring Systems*. 625–626.
- Zapata-Rivera, D., Jackson, T., & Katz, I. R. (2015). Authoring conversation-based assessment scenarios. In R. A. Sottolare, A. C. Graesser, X. Hu, & K. Brawner (Eds.), *Design Recommendations for Intelligent Tutoring Systems Volume 3: Authoring Tools and Expert Modeling Techniques*. Pp. 169–178. U.S. Army Research Laboratory.
- Zapata-Rivera, D., Liu, L., Chen, L., Hao, J., & von Davier, A. (2016a). Assessing science inquiry skills in immersive, conversation-based systems. In K. Daniel (Ed.), *B* (pp. 237–252). Big Data and Learning Analytics in Higher Education: Springer International Publishing.
- Zapata-Rivera, D., Zwick, R., & Vezzu, M. (2016b). Exploring the Effectiveness of a Measurement Error Tutorial in Helping Teachers Understand Score Report Results. *Educational Assessment*, 21(3), 215–229.
- Zapata-Rivera, D., Kannan, P., Forsyth, C., Peters, S., Bryant, A. D., Guo, E., & Long, R. (2018). Designing and evaluating reporting Systems in the Context of new assessments. In D. Schmorow & C. Fidopiastis (Eds.), *Augmented cognition: Users and contexts. AC 2018, Lecture Notes in Computer Science*, vol (Vol. 10916, pp. 143–153). Cham: Springer.
- Zapata-Rivera, D., Graesser, A., Kay, J., Hu, X., & Osofsky, S. J. (2020). Visualization implications for the validity of ITS. In *In Design Recommendations for Intelligent Tutoring Systems: Volume 8 – Data Visualization*. Orlando: U.S. Army Research Laboratory.
- Zwick, R., Zapata-Rivera, D., & Hegarty, M. (2014). Comparing graphical and verbal representations of measurement error in test score reports. *Educational Assessment.*, 19(2), 116–138.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.