Using GIFT to Develop an Adaptive Distributed Learning Environment Supporting Data Science Competencies

Fadjimata Anaroua[1], Hong Liu[1], and Naomi Malone[2]
[1]Embry-Riddle Aeronautical University; [2]Cybermedia Technologies Inc.

INTRODUCTION

This paper presents our latest research on the efficient support of Data Science (DS) students in higher education, particularly focusing on underserved universities. The need for new graduates and professionals to upskill in DS surpasses the capacity of universities to offer conventional classes, particularly in underserved universities (NASEM, 2018). Our solution provides otherwise unavailable DS courses for all students by implementing the Generalized Intelligent Framework for Tutoring (GIFT, Sottilare et al., 2012) to develop a multi-university adaptive distributed learning (ADL) environment that can share DS courses and facilitate student learning from anywhere at any time. This distributed learning ecosystem using Department of Defense (DOD)-initiated technologies (ADL, 2018) allows students from 11 networked universities to share courses and resources, providing equal access to underserved and better-equipped research universities within the system. Besides GIFT, the ADL environment integrates the learning management system (LMS), Moodle, competency management software such as Competence and Skill System (CaSS, 2021), and Learning Record Stores (LRSs) to collect and analyze data for personalized learning. Our instructional design and course development efficiently align learning objectives, activities, and assessments of DS student competencies based on the Edison DS Competency Framework (Edison, 2017).

In previous research (National Science Foundation Grants 1244967 and 1626602, 2014-2020), we developed and delivered five distributed learning (DL) courses using online and teleconferencing tools for 3 Underrepresented Racial Minority (URM) universities. Our latest National Science Foundation (NSF) grant (NSF Grant 2145214 2022-2025) expands this project to develop and deliver an program of 10 courses in DS for students in 11 universities, including 8 URM universities. The scope of the funded project is shown in Figure 1. The overarching goal of the newly NSF-funded project is to adapt GIFT and other DOD-initiated educational technologies to an academic setting to develop and analyze their efficacy in producing DS graduates who can succeed in the expanding DS workforce.

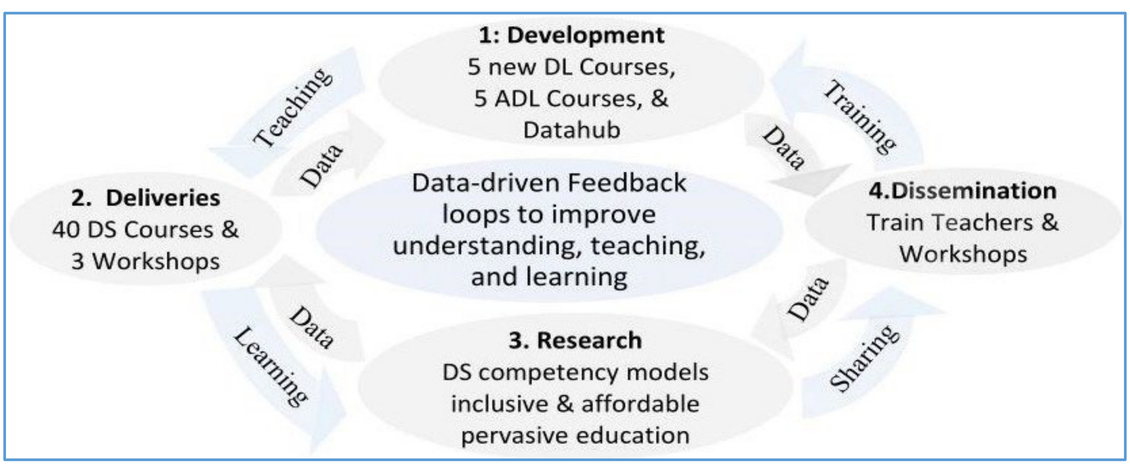

Figure 1. Data-driven formative evaluations are used to inform project progression.

There are several challenges to developing DS courses in a multi-university distributed environment. First, due to increasing class sizes, universities must already accommodate large classes that often mix in-person and remote students. Second, the interdisciplinary nature of DS courses attracts students from diverse academic backgrounds, bringing another challenge to offering courses that are customizable to student backgrounds. ADL courses enabled by GIFT not only can tailor the course to fit the individual student's learning paces and preferences but also increase teaching efficiency. Our current work focuses on automating non-cognitive tasks (Liu et al., 2022). The next phase will automate cognitive aspects of formative learning assessments: grading homework, monitoring team collaboration, and providing feedback. The goal is to automate all repeating routines and reduce instructor tasks.

This paper is organized as follows. Section 2 reviews the literature on team competency assessment in DS applications. Section 3 explains the instructional design and strategies to automate part of the assessment and aggregate lower-level competencies into higher-level ones. Section 4 describes the system integration of Moodle, GIFT, CaSS, and LRS. Section 5 presents three use-cases for students, instructors, and educational researchers, respectively. Section 6 discusses the innovation, impact, challenges, and future work of the project.

LITERATURE REVIEW ON TEAM COMPETENCY ASSESSMENT

The primary objective of our previous and current NSF-sponsored projects is to use DL and ADL technologies to enhance active learning that is competency-based, inclusive, and cost-sensitive. Experiential learning is a popular active learning approach that boosts metacognition and reflection (Kolb & Kolb, 2009). Developed in our previously funded projects, the instructional design of our DL course materials followed the Online Learning Initiative (OLI) framework originated from Carnegie Mellon University to align learning objectives, learning activities, and learning assessments (Brooks 2017; Lovett et al., 2008). The primary tasks for content development were to supplement current non-DL course materials with case studies, applications, and data analytics projects. In addition, these DL courses included a Course-based Undergraduate Research Experience (CURE) component to promote deep experiential learning (Bangera & Brownell, 2014; Liu et al., 2018).

A major task of our current project is to transfer existing DL courses into ADL courses that follow current data interoperability standards and provide personalized and flexible learning leveraging AI, big data, and communication technologies. ADL courses also collect and analyze learner data to provide feedback loops that support learners and educators, improve courses, and advance our understanding of human learning (ADL, 2018; Bienkowski et al., 2012; Sottilare et al., 2018);. In Liu et al. (2020), we presented a preliminary ADL matrix algebra course using GIFT and the Moodle LMS and stored learning activity data in the LRS.

INSTRUCTIONAL DESIGN AND ASSESSMENT STRATEGIES

The 10 courses include five domain-agnostic courses in Introduction to Data Science (IDS), Data Visualization (DV), Mathematical Modeling & Simulation (MMS), Data Mining (DM), and Cloud Computing, as well as five domain-specific courses in Genomics and Bioinformatics (GBI), Advanced Computing Resources in Biology (ACRB), Data-driven Decision Making (DDM), Environmental Data Analytics (EDA), and Data Analytics for Public Health (DAPH). Five courses - MMS, DV, DM, GBI, and ACRB, were DL courses that were developed and offered in hybrid learning modes under the two prior funded projects from 2014-2020. The other five courses have only been offered before as traditional lecture classes in a single university with the corresponding instructor. In year 1 of the project, Embry-Riddle Aeronautical University (ERAU) will develop the first adaptive DL course (MMS). In year 2, ERAU will create two more adaptive DL courses (DV and DM). All 5 courses listed above will be upgraded as adaptive by year 3 and adjusted based on student outcomes over the final two years of the grant.

We aim to teach students effectively by building on their prior knowledge and addressing their learning gaps. Our research method uses competency-based learning assessment and learning analytics**,** defined as "the use of data and analysis to understand and improve learning and its environment". We foster teamwork and timely feedback among students during problem-solving activities and promote student motivation through collaborative learning, both online and in-person. The primary theoretic challenge for our DS project is how to assess team competency in the context of DS applications (Owens & Goldberg, 2022; Salas et al., 2017; Vatral et al., 2022;).

Our competency-based learning strategy is to decompose the top-level competency from the Edison DS Competency Framework into tasks and roles of teamwork and then roll up lower-level competencies into high-level competencies reversely. At the top level, the DV course targets two competencies: 1) Applying visualization tools and DS ethics to communicate effectively for diverse audiences; and 2) identifying relevant data sources, retrieving data, cleansing data, transforming data, and warehousing data. While homework assignments, quizzes, and tests assess the basic skills and knowledge, the overall competencies are trained and evaluated through teamwork in the last five weeks of classes. The two objectives are decomposed into six lower-level weekly learning objectives, as the project process is shown in Figure 2.

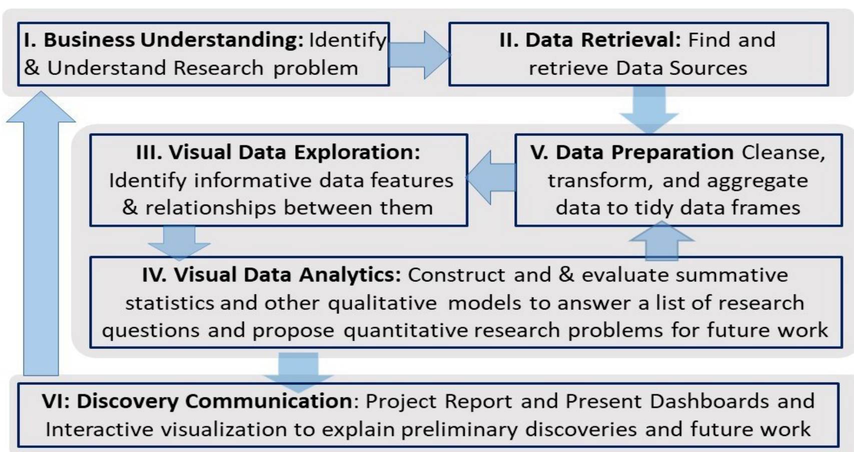

Figure 2. Team Project Process for Data Visualization Course

As Shown in Table 1, the weekly learning objectives are further divided into tasks and roles in the size that a student can complete in a week. Each team, typically 3-5 students, must use a teamwork tool (e.g., Slack) to record the team meetings and task assignments. The instructor tracks the team process and collects data for evaluation using the teamwork tool. Due to time restraints, the instructor often only has time to check records at the end of the term. Therefore, a software application called the BotCaptain is assigned an email and added to each team to perform formative assessments, monitor progress, and provide feedback. BotCaptain collects data using an xAPI (eXperience Application Programming Interface) format and stores it in a noisy LRS.

Table 1. Tasks and Roles for DV Team Project

Weekly Objectives	**Roles and Task Description**	**Deliverables to store in the team-shared folder**
Week 1: Business Understanding	*Domain Researchers*: One to two members with DS application domain knowledge other than DS, Math, or Computer Science (CS). Start literature review and internet search, brainstorm with other members and instructor to identify the research problem and possible data sources. *Data Retrievers*: Two members with Database and programming skills. Identify, extract, and pull together available and pertinent heterogeneous data, including modern data sources such as social media data, open data, and governmental data.	***Submit (1):*** *Team members and weekly meeting times are set up and posted in the shared folder.* ***Submit (2):*** *Sign team agreement in page 7 and upload to shared folder!* *Team members and weekly meeting times* are set up and posted in the shared folder. Who serves as domain researchers? Who serves as data retrievers?
Week 2: Data Preparation	*One domain researcher* may continue to propose business questions that DM may help to get answers. *One Data Retriever* may continue to retrieve more datasets and merge datasets. *Data Wranglers*, 2 (+ depending on the need above) members with strong programming skills. Use Numpy, Pandas, and Microsoft Excel to cleanse and transform data into a tidy data frame.	***Submit (3):*** *One page proposal* about the topics of the project and data sources that your team agrees to work on. Who serves as data wranglers and what are the roles of other members?
Week 3: Data Visual Exploration	*Data Wranglers:* Continue to cleanse and transform data based on the inputs from the explorative analyzers. *Exploratory Analyzers*: Two members. Use Tableau, Pandas, Matplotlib, and Seaborn to identify and rank the informative values of the features to be modeled for visual analytics. Questions are asked to explore the data distributions, relationships, and their implications for business actions.	***Submit (4)****: upload a tidy data frame* (table) that is ready for data mining. Who serves as *Exploratory Analyzers,* and what are the roles of members? What questions are to be explored?

		What features are selected to answer the questions?
Week 4:	*Visual Developers:* Each member needs to develop visual analytics *independently* to answer the questions that the team developed. The team will agree on which graphs will be selected and merged for the project report and final presentation.	***Submit (5):*** Preliminary Visual analytics to explain the answers to the questions of research interest. Have you answered the questions?
Week 5: Discovery communication. Final presentation and report due Date.	*Dashboard developer:* Two members will be selected to develop the dashboard and merge the selected visual analytics and graphs. *Writers*: All members will be responsible for providing technical details of the visual analytics you developed, and the writer will help merge and polish writing. Future work for the data mining course is encouraged. *Presenters: All members* are in-person. The team will give a final presentation.	***Submit (6):*** Email your peer reviewer! ***Submit (7):*** a white paper/report about the findings and ***Submit (8):*** presentation slides, and seven meta questions to be answered. See rubrics below for evaluation. ***Submit (9):*** Executable visualization code in Python and Tableau.

Using CaSS to roll up lower-level competencies into higher-level competencies and match compatible competencies is a work in progress. Since the bottom-level skills and competencies associated with every learning activity are all recorded in xAPI format ("Who *did* what"), we need to configure the taxonomy of the hierarchically organized competencies in CaSS roll-up rules and compatible rules to certify higher-level competencies after a student completes a set of lower-level competencies. This system will allow automated learning assessment to make adaptive content recommendation during learning.

SYSTEM ARCHITECTURE AND TOOL INTEGRATION

Universal data exchange and Learning Tool Interoperability (LTI) for long-term sustainability were our primary concerns in the development of platforms and tools. We selected the GIFT tutoring system because it is an integrated component of the Future Learning Ecosystem (Smith & Ram, 2019). Moodle was selected as a LMS because Moodle is an LTI tool that facilitates the xAPI (https://github.com/adlnet/xAPI-Spec) standard for data exchanges. Moodle and GIFT, can not only directly exchange data as consumers or producers but also share mutually accessible data through the xAPI data format (Hruska1 et al., 2015) and LRSs. In this section, we present the tool configuration, content deployment, and use cases, as well as the components that map the pedagogy and instructional design into the artifacts of the GIFT tutoring system.

We installed Moodle (www.icycle.cloud), GIFT (http://3.12.146.191:8080/dashboard/#login), and a LRS (https://erau.xapi.io/) in the EC2 servers of the Amazon Web Service (AWS) cloud. Veracity Learning, Inc.

donated the LRS and its accompanying data visual analytics tools. Figure 3 shows the system components distributed in cloud services and data exchange through three levels of the LRS. BotCaptain is an in-house developed web bot that helps instructors to collect teamwork data and perform primitive learning assessments (Liu et al., 2020). In our recent implementation, we replaced the in-house developed Natural Language Processing (NLP) component with ChatGPT. The course contents are posted on the Moodle site mostly through linked pages, while the online exercises, formative assessment, and content recommendations are delegated to GIFT (using its default pedagogical model). The learner state transition uses the three default levels - below, meet, or exceed, for the students to change state and move to the next learning activity. The student learning records are stored in a noisy LRS.

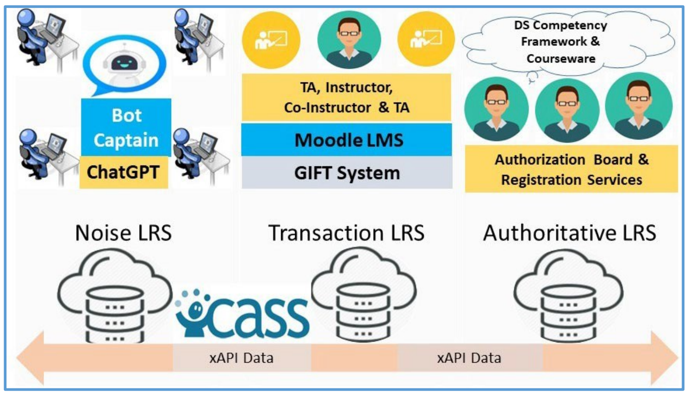

Figure 3. The Conceptual Architecture Design of the Distributed Learning Environment

At the current stage, we are focusing on automating the non-cognitive tasks, such as collecting data or reminding students about a deadline (Liu et al., 2022). Next, we will gradually automate the cognitive aspect of formative learning assessments, such as grading homework, monitoring team collaboration, and providing feedback. The ultimate goal is to automate all repeating routines and reduce instructor tasks to grading test papers and project reports at the end of the project.

USE CASES FOR STUDENTS, INSTRUCTORS, AND RESEARCHERS

This section illustrates three use cases of our ADL environment to support students with self-paced and personalized education, instructors with evidence for iterative course improvement, and educational researchers with the data for learning analytics.

The first example use case shows how GIFT provides learning anywhere at any time. Joe is a junior Biochemistry major and student-athlete at Hampden-Sydney College (HSC). Under recommendation by his teacher, Dr. Smith, for Data Analytics for Public Health (DAPH) at HSC, Joe watches five video lessons from ERAU instructor and project Principal Investigator (PI) Dr. Liu's Data Visualization GIFT courseware from his smartphone while traveling for a game at another institution. GIFT recommends progression from one lesson to another based on its real-time learning assessment. After Joe returns to his room that evening, he completes the assigned online homework. Each activity triggers an xAPI assertion in a noisy LRS for Joe. Several assertions collectively trigger the summative assertion in a transactional LRS, "Joe completed the Visual Data Analytics Module (Novice)." Following up on the DS courses, we form a hybrid team consisting of one pair from Dr. Smith's DAPH class (e.g., including Joe), and the other pair from Dr. Liu's

Data Mining Class. The collective skills of students cover both domain knowledge in DAPH and Data Mining. After Joe completes the workshop, an authoritative xAPI assertion, “Joe demonstrated expertise in visual data analytics of DAPH (Intermediate),” is certified.

The second use case demonstrates how the instructors use visual analytics derived from data in the LRS to iteratively improve the instructional design and course content. As illustrated in Figure 2, this project's emphasis is on formative measures that are likely to improve both students’ and teachers’ performance. For DL courses, an impact study will be conducted that includes measures of student performance and standardized tests. Each individual course will be designed to have pre- and post-tests to measure content competencies among students. Pre-tests will help instructors to develop personalized learning activities. Student preference questions on pre- and post-tests will be used to determine the degree to which learner engagement and motivation efforts were successful. We can use a dashboard LRS (donated by Veracity Learning) to visualize the learning outcomes in a transactional LRS for individual learners, teams, and cohorts to identify the problematic content and ineffective instructional designs.

Multiple delivery modes and heterogeneous student backgrounds make the instructional design of our courses particularly challenging. A one-size-fits-all best model is unrealistic. As a practical incremental approach, the instructors and researchers of this project meet in a symposium each summer to share findings and identify the “best practice so far” for each network course, allow peers to duplicate and observe the outcomes, and determine what is effective and what requires improvement.

The third use case shows how the researchers use the data to conduct learning analytics research and identify repeating routines to be automated. The proposed project related to this research will require us to log our time efficiencies to use Artificial Intelligence (AI) to relieve repetitive routines for teachers. Time cost-effectiveness is crucial for the sustainability and adaptability of proposed DS courses. In the first two years of this project, we will determine which instructor routines are suitable for potential replacement by a virtual Teaching Assistant (TA). Specifically, we will measure the instructor's time in mentoring classes and workshops as well as student learning outcomes and feedback. To accomplish this, we will classify instructors' duties into two categories: domain-agnostic routines and domain-specific instruction. The data for time efficiency will be collected from surveys of key personnel that show the hours spent on creative tasks on lecture preparation, learning assessment, lecturing, mentoring, and tutoring.

CONCLUSIONS AND FUTURE WORK

This paper presents the large-scale practice of transferring DOD's advanced distributed learning technologies from military training into academic settings to promote pervasive DS education. Based on student-centered pedagogy, it was built on the previous work described by Liu et al. (2020). In the instructional design of this paper, we illustrated how a higher-level competency is decomposed into a set of lower-level competencies for delivering learning activities in the size of launchable units LMS compatible with the Sharable Content Object Reference Model (SCORM). Such an instructional design enables us to apply the associated technology, such as CaSS, to automate part of learning assessment and content recommendation. The progress of technology and applications could be faster, but steadfast in improving online courses' quality and reducing human intervention costs. The three use cases of the ADL environment illustrate how it provides students with self-paced and personalized education, instructors with evidence for iterative course improvement, and educational researchers with the data for learning analytics.

This NSF-funded project involves 11 institutions, 23 researchers and instructors, and at least two hundred DS mini-bachelor’s degree awardees. We are developing 5 ADL courses, the first ADL course this year, two more courses in 2024, and the last two in 2025. The courses will be revised annually based on student feedback and learning outcome data. As an ongoing work in the next three years, the instructors must transmit their course contents into the Moodle LMS so that our DS courses can be integrated with GIFT and LRSs. The AWS cloud hosting solution has the long-term benefit of using a small business model to sustain the project after the NSF fund terminates.

ACKNOWLEDGEMENT

This research was partially sponsored by the National Science Foundation (NSF) of the USA under the IUSE Grant NSF Grant 2145214 2022-2025. The authors would like to thank Veracity Learning Inc. for donating the LRS software system and the colleagues in GIFT research and development for their free online technical support.

REFERENCES

ADL (2018). DOD's Advanced Distributed Learning Initiative: Total Learning Architecture: 2018, Reference Implementation Specifications and Standards, W900KK-17-D-0004, https://adlnet.gov/projects/tla .

Bangera G & Brownell S.E. (2014). Course-based undergraduate research experiences can make scientific research more inclusive. *CBE-LSE 13*(4), 602-606.

Bienkowski, M., Feng, M., & Means, B. (2012). Center for Technology in Learning, SRI International, U.S Education Department, Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, https://files.eric.ed.gov/fulltext/ED611199.pdf .

Brooks, J. (2017) Learning Design: The Simon Approach, LearnLab Summer School, Open Learning Initiative Track, July 10 - 14, 2017, https://docs.google.com/presentation/d/1kiwK2L0WYYW0WJMJapc_tuNV_NrsFwhegJY6Y9YXM8c/edit#slide=id.p4

CaSS (2021), Competencies and Skill System by EduWorks Inc., https://www.cassproject.org/

Edison (2017). Edison Data Science Framework: Part 1. Data Science Competence Framework (CF-DS), https://creativecommons.org/licenses/by/4.0/.

IBM Analytics, (2020). The Data Science Skills Competency Model, A blueprint for the growing data scientist profession, https://www.ibm.com/downloads/cas/7109RLQM#:~:text=A%20data%20scientist%20must%20demonstrate,must%20be%20good%20problem%20solvers.

Kolb, A. Y., & Kolb, D. A. (2009). The learning way: Meta-cognitive aspects of experiential learning. *Simulation & gaming, 40*(3), 297-327.

Liu, H., Bernard, T., & Acharya, K. (2020). Using GIFT to Develop Adaptive Remedial Courses for Graduate Degree Programs in Data Science, Proceedings of the 8th Annual GIFT Users Symposium, 61-68.

Liu, H., Bernard, T., Cankaya, E., & Hall, A. (2022). Task-agnostic team competence assessment and metacognitive feedback for transparent project-based learning in data science. *International Journal of Smart Technology and Learning, 3*(2), 138-162.

Liu, H., & Malone, N. (2021). Building a Competency-Based Data Science Certificate Program Using CASS – A Prototype, Presentation accepted by iFest 2021 "Learning and Thriving in the New Normal" Virtual and In-person Conference organized by National Training & Simulation Association, Washington DC, 8/31 - 9/2, 2021.

Liu, H, Warner, T., Ikle, M, and Mitten S. (2020) Harness Big Data by iCycle - Intelligent Computer-supported hybrid collaborative learning environment. *International Journal of Smart Technology and Learning, 2*(1), 31-47.

Lovett, M., Meyer, O., & Thille, C. (2008). The Open Learning Initiative: Measuring the Effectiveness of the OLI Statistics Course in Accelerating Student Learning. *The Journal of Interactive Media in Education*, 14, 1-16.

NASEM 2018, National Academies of Sciences, Engineering, and Medicine 2018. Data Science for Undergraduates: Opportunities and Options, Washington, DC: The National Academies Press. https://doi.org/10.17226/25104.

NIST Big Data PWG (2015-1) NIST Big Data Public Working Group , Definitions and Taxonomies Subgroup, NIST Big Data Interoperability Framework: Volume 1, Big Data Definition, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf.

Owens, K., & Goldberg, B. (2022). Competency-Based Experiential-Expertise. In Sinatra, A., Graesser A., Hu, X., Goldenberg, B., Hampton, A., Johnson, J., (eds.) Design Recommendations for Intelligent Tutoring Systems, Volume 9, Competency-Based Scenario Design.

Salas, E.,. Reyes, D.L, & Woods, A.L. (2017). The Assessment of Team Performance: Observations and Needs, In A.A. von Davier et al. (eds.), Innovative Assessment of Collaboration, Methodology of Educational Measurement and Assessment, Springer International Publishing Switzerland.

Smith, B., & Ram, P. (2019). Interoperability. In Walcutt, J.J. & Schatz, S. (eds), Modernizing Learning: Building the Learning Ecosystem, pp. 107 – 128.

Sottilare, R.A., Brawner, K., Goldberg, B., & Holden, H. (2012). The Generalized Intelligent Framework for Tutoring (GIFT), https://gifttutoring.org/projects/gift/documents.

Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2018). Designing adaptive instruction for teams: A meta-analysis. *International Journal of AI in Education, 28*(2), 225-264.

Vatral, C., Biswas, G., Mohammed, N., & Goldberg, B.S. (2022). Automated Assessment of Team Performance Using Multimodal Bayesian Learning Analytics, 2022 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).

ABOUT THE AUTHORS

Fadjimata Anaroua *is a Ph.D. student and research assistant at Embry-Riddle Aeronautical University. She was awarded a master's degree in software engineering at Embry-Riddle Aeronautical University, a bachelor's degree in Aeronautical Science, and a bachelor's degree in industrial engineering.*

Dr. Hong Liu *was awarded a Ph.D. in Mathematics and MS in Computer Science at the University of Arkansas, Fayetteville in 2000. He serves as a professor in Mathematics and Computing at Embry-Riddle Aeronautical University. He served as PI and Co-PI of 14 sponsored projects and published numerous articles in mathematics, data science, and STEM education. His current research interest is to develop a smart learning environment to promote peer learning.*

Dr. Naomi Malone *was awarded a Ph.D. in Instructional Design and Technology from the University of Central Florida in 2017. She is currently a Senior Instructional Designer at Cybermedia Technologies, designing online training for the DOD. Her primary area of interest is promoting self-regulated learning practices in online learning environments using evidence-based learning strategies and distributed learning technologies.*