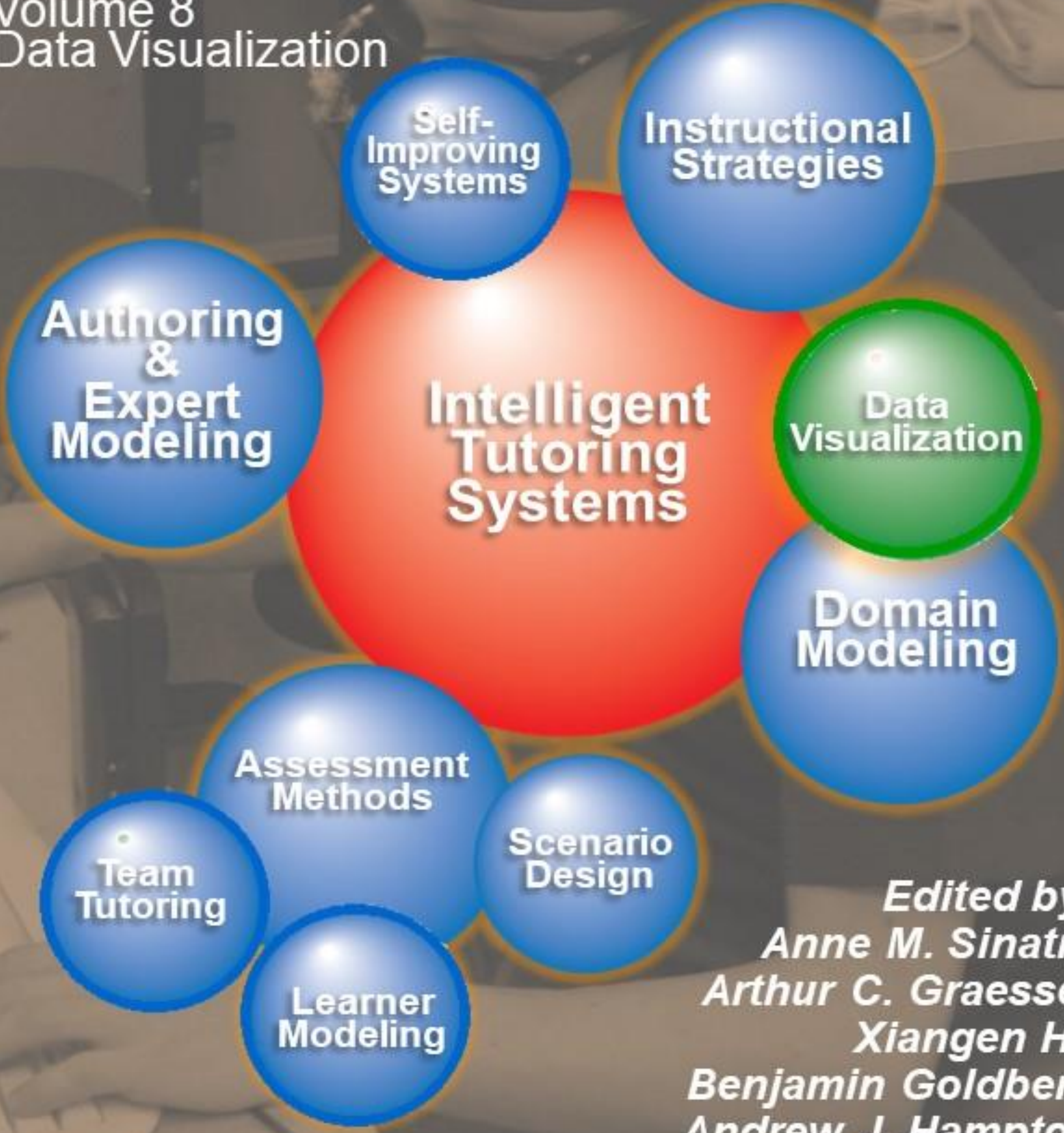
Design Recommendations for Intelligent Tutoring Systems

Volume 8
Data Visualization



Edited by:
Anne M. Sinatra
Arthur C. Graesser
Xiangen Hu
Benjamin Goldberg
Andrew J. Hampton

A Book in the Adaptive Tutoring Series

Design Recommendations for Intelligent Tutoring Systems

Volume 8
Data Visualization

Edited by:
Anne M. Sinatra
Arthur C. Graesser
Xiangen Hu
Benjamin Goldberg
Andrew J. Hampton

A Book in the Adaptive Tutoring Series

The views expressed herein are those of the authors and do not necessarily reflect the views of the US Army Combat Capabilities Development Command - Soldier Center.

Use of trade names or names of commercial sources is for information only and does not imply endorsement by the US Army Combat Capabilities Development Command - Soldier Center.

This publication is intended to provide accurate information regarding the subject matter addressed herein. The information in this publication is subject to change at any time without notice. The US Army Combat Capabilities Development Command - Soldier Center, nor the authors of the publication, makes any guarantees or warranties concerning the information contained herein.

Printed in the United States of America
First Printing, December 2020

US Army Combat Capabilities Development Command - Soldier Center
Simulation and Training Technology Center
Orlando, Florida

International Standard Book Number:
978-0-9977257-8-0

Special thanks to Jody Cockroft, University of Memphis, for her efforts in coordinating the workshop that led to this volume.

Dedicated to current and future scientists and developers of adaptive learning technologies

Contents

INTRODUCTION TO DATA VISUALIZATION & GIFT

*Anne M. Sinatra[1], Arthur C. Graesser[2], Xiangen Hu[2],
Benjamin Goldberg[1], and Andrew J. Hampton[2] Eds.*

*[1]U.S. Army Combat Capabilities Development Command – Soldier Center –
Simulation and Training Technology Center
[2]University of Memphis Institute for Intelligent Systems*

This book on data visualization is the eighth in a planned series of books that examine key topics (e.g., learner modeling, instructional strategies, authoring, domain modeling, assessment, team tutoring, self-improving systems, data visualization, and competency based scenario design) in intelligent tutoring system (ITS) design. This book focuses on data visualization and how it is applied in ITSs. The chapters within this book specifically examine topics in relationship to the Generalized Intelligent Framework for Tutoring (GIFT) (Sottilare, Brawner, Goldberg & Holden, 2012; Sottilare, Brawner, Sinatra, & Johnston, 2017). GIFT is an open-source, domain-independent, modular, service-oriented architecture for ITSs. The design of GIFT allows for reusability, reduction in authoring time, and reducing the skill level needed to create an ITS. GIFT provides functionality to create ITSs, distribute ITSs to learners through the Cloud, conduct research to evaluate ITSs, and to examine instructional outcomes.

Along with this volume, the first seven books in this series, Learner Modeling (ISBN 978-0-9893923-0-3), Instructional Management (ISBN 978-0-9893923-2-7), Authoring Tools (ISBN 978-0-9893923-6-5), Domain Modeling (978-0-9893923-9-6), Assessment Methods (ISBN 978-0-9977257-2-8), Team Tutoring (ISBN 978-0-9977257-4-2), and Self-Improving Systems (978-0-9977257-7-3) are freely available at www.GIFTtutoring.org.

Data visualization is an important topic for ITSs, as there are many different users of the systems (including learners, instructors, researchers, subject matter experts). The data that is collected by the ITS can be organized and displayed in a number of different ways. The current book includes a general discussion of how data visualizations can be applied in ITSs, as well as detailed specific examples of existing implementations, and technical details related to incorporating data visualization in ITSs. We believe this book can be used as a design tool for data visualization interfaces in ITSs.

GIFT and Expert Workshops

In 2012, Army Research Laboratory (ARL) with the University of Memphis developed expert workshops of senior tutoring system scientists from academia and government to influence the GIFT design goals moving forward. Expert workshops have been held each year since 2012 resulting in volumes in the ***Design Recommendations for Intelligent Tutoring Systems*** series the following year. In 2018, parts of ARL, including the GIFT team, were reorganized into another organization, Soldier Center. Research into applied adaptive tutoring and team tutoring have continued with Soldier Center. Additionally, the expert workshops and books have continued with topics in line with the relevant research gaps. Table 1 lists the expert workshop topics, the locations of the workshops, as well as the dates of the workshops and associated volume publications.

Table 1. Historical List of Expert Workshops, Locations, Dates, and the Book Publication Output.

Expert Workshop Topic	Expert Workshop Location	Expert Workshop Date	Book Publication
Learner Modeling	Memphis, TN	September 2012	Volume 1 – July 2013
Instructional Management	Memphis, TN	July 2013	Volume 2 – June 2014
Authoring Tools	Pittsburgh, PA	June 2014	Volume 3 – June 2015
Domain Modeling	Orlando, FL	June 2015	Volume 4 – July 2016
Assessment Methods	Princeton, NJ	May 2016	Volume 5 – June 2017
Team Tutoring	Ames, IA	May 2017	Volume 6 – August 2018
Self-Improving Systems	Nashville, TN	May 2018	Volume 7 – October 2019
Data Visualization	Orlando, FL	August 2019	Volume 8 – December 2020
Competency Based Scenario Design	Virtual	September 2020	In Progress

Design Goals and Anticipated Uses of GIFT

GIFT was designed with multiple functions in mind:

1. An architectural framework that is modular, and has components that can be replaced and customized by ITS authors for their specific tutor.

2. A set of authoring tools which allows subject matter experts, and those without a background in computer science to easily create customized ITSs.

3. A testbed for experimental research, which allows for the examination of research questions relevant to the continued development of ITSs.

The chapters within the book provide recommendations for how to implement the methods within the GIFT architecture with the above functions in mind.

Sections of the Book

This book is organized into three sections covering three related but diverse topics:

I. General Applications of Data Visualization

II. Data Visualization in Specific Domains and Applications

III. Technical Applications for Data Visualization

Chapter authors in each section were carefully selected for participation in this project based on their expertise in the field as ITS scientists, developers, and practitioners. *Design Recommendations for Intelligent Tutoring Systems: Volume 8 – Data Visualization* is intended to be a design resource as well as a community research resource.

We believe that Volume 8 can serve as an educational guide for developing ITS scientists and as a roadmap for ITS research opportunities. The authors of the chapters contained herein are experts in their area and the references provided (their own and those of others) compose a rich web of working professionals in the ITS field.

References

Sottilare, R.A., Brawner, K.W., Goldberg, B.S. & Holden, H.K. (2012). *The Generalized Intelligent Framework for Tutoring (GIFT)*. Orlando, FL: U.S. Army Research Laboratory Human Research & Engineering Directorate (ARL-HRED).

Sottilare, R., Brawner, K., Sinatra, A. & Johnston, J. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Orlando, FL: *US Army Research Laboratory*. May 2017.

Section I – General Applications of Data Visualization

Dr. Arthur C. Graesser and Dr. Andrew J. Hampton, Eds.

CHAPTER 1 – INTRODUCTION TO GENERAL APPLICATIONS OF DATA VISUALIZATION

Andrew J. Hampton and Arthur C. Graesser
University of Memphis

Core Ideas

The chapters in this section focus broadly on opportunities for advancing data visualization, design approaches that leverage various fields of engineering and psychology, and best practices to objectively validate data visualization approaches and principles.

General characteristics of good visualization already exist, including those that have been developed after decades of research in human factors, human-computer interaction, and cognitive science. For example, it is a good policy to have an overview data visualization for the user to start out with and quickly return to, with the opportunity to drill down to examine aspects of the data or even raw data in more detail. Empirical research also compares the relative ease of interpretation for various standard representation classes (e.g., bar charts, scatterplots, etc.). This research remains far from complete, especially with the expansion into multi-modal and interactive representations. Consequently, there needs to be additional basic research in data visualization design to fill such gaps in the literature.

Considerations for the intended audience play a critical role in guiding data visualization design. Today's learning technologies collect a deluge of information in an era of progressively bigger data. These data vary in format to satisfy different goals, grow dynamically over time, are often distributed in different locations, and are periodically reorganized for different purposes. Different groups (stakeholders) want to see aspects of the data that are important for their distinct goals and tasks. This influences not just which data to display, but also grain size, emphasis, and order of presentation. The complexity increases when the emphasis lies on collaborative learning that values metrics of both individual and team performance.

Given the current and inevitable increases in both the volume and complexity of data, guiding principles must be acknowledged, evaluated, and potentially new ones developed. Technological convergence offers one avenue. By capitalizing on transformative technologies in disparate fields, the learning technology community can minimize the time and investment needed to dramatically improve how we provide stakeholders with data. The chapters in this section explore alternative frameworks for approaching data visualization, at a time when data visualization standards must rise to a new level through a combination of established principles, appreciation for audience characteristics, convergent technology, innovative theoretical approaches, and adoption of empirically successful commercial products.

Individual Chapters

Goodell and Thomas take a learning engineering perspective to developing effective data visualization. Data visualization interfaces strive for the same essential goal as learning technologies, namely, to address learning objectives. Learners from this perspective can be designers evaluating an intervention, teachers evaluating a class, or the individual learner trying to understand his own progress. The design of the data visualization should proceed from structured questions regarding desired outcomes for the learner (e.g., enabling action, change in behavior, novel application of obtained knowledge). Learning engineering also emphasizes human-centered design, informed by cognitive science principles. It remains sensitive to feedback so that objective analysis of the visualization iteratively informs improvements and optimizations relative to the established learning objectives.

Kay, Zapata-Rivera, and Conati highlight the need to create intelligent systems that are explainable, transparent, and accountable, forming a foundation for ensuring that users can understand and control the intelligent systems they use. This chapter presents *scrutable* learner models, enabling learners to understand and control learner models in intelligent tutors created in an authoring system like GIFT. A user-centered approach defines scrutable learner modeling in terms of questions that learners should be able to answer using an Open Learner Model (OLM) interface. Scrutibility requires a whole-system approach to the authoring of personalized tutors— essential for capturing the relevant information during the authoring process.

Harrison and Hampton explore the potential of representations to encourage dynamic, kinesthetic interpretation from the viewer. This approach, a hybrid of narrative psychology and human factors psychology, embraces the complexity of learning data by organizing it through visual storytelling. Datasets naturally have predictable areas of greater and lesser importance, and often information that needs context to meaningfully absorb it. These properties suggest an order of presentation that cognitive and perceptual principles can encourage (or dissuade alternatives). Interactive visualizations expand the toolbox derived from these principles, and multi-model displays (auditory, tactile) hold a great potential.

DeFalco and DeFalco discuss the opportunities available in leveraging existing data that have conventionally been too costly for mid-stream analysis. Technologies to realize this potential already proliferate private industry to such an extent that capitalizing on them for the educational field could be accomplished without having to invent new platforms, processes, or techniques. The change in time scale afforded by this shift can change the purpose of data visualization from a reflective analysis to a mid-course decision-making tool. This applies to learners, teachers, and instructional designers. Similar techniques at various levels of analysis can ensure that the data presented to those stakeholders is not just timely, but targeted—data visualizations informed by the interaction of expertise and empirical validation.

Swiecki, Misfeldt, Hu, and Shaffer focus on the need to dynamically model connectivity among collaborative actors. With the increasing importance of teamwork and collaborative problem solving, data visualizations must keep pace in representing the added layer of complexity. Further, those visualizations must accommodate the distinct levels of analysis required by researchers, educators, and team members in action. Epistemic Network Analysis has demonstrated utility in modeling complex, interconnected team performance in educational domains. It does so by emphasizing connections between and among individuals and key concepts of the domain over time, which allows a visual representation of team members' contributions to specific aspects of team performance.

Zapata-Rivera, Graesser, Kay, Hu, and Ososky offer principles for evaluating the validity of data visualizations in the educational domain that consider the perspective of different stakeholders (students, teachers, policy experts, the public). The authors address concerns ranging from perceptual constraints on organization and complexity of the data to consideration of audience (a considerable challenge given the diversity of stakeholders in intelligent tutoring systems). These principles, drawn primarily from human–computer interaction, human factors psychology, and education, have a proven track record across digital environments, including learning technologies.

Harrison discusses both the challenges and opportunities that are involved with implementing data visualizations in intelligent tutoring systems. Among these topics for consideration are the scale of the system, characterizing the data, interpreting the logs of the data, and thinking carefully about the design implementation of the data. The author addresses not only these challenges, but also highlights the importance of evaluation of data visualization in ITSs.

Chapter 2 - A Learning Engineering Approach to Data Visualization

Jim Goodell and Bridget E. Thomas
Quality Information Partners

Introduction

This chapter presents the development of data visualizations as a learning engineering exercise. The approach is based on the theory that a viewer's interaction with a data visualization is a learning experience intended to address specific learning objectives. The process starts with questions like: What do you want the viewer/learner to know, be able to do, or do after experiencing the data visualization? Is it a call to action? Do you want to change the viewer's behavior? Do you want the viewer to be able to apply the new knowledge in some way? The approach applies human-centered design, incorporates key elements of cognitive science, uses data to inform design decisions, and iteratively develops and tests aspects of the visualization to optimize it for the desired outcomes.

Learning Engineering

More than 50 years ago, Carnegie Mellon University professor Herbert A. Simon coined the term "learning engineering." (Simon, 1967). More recently, the IEEE Standards Association's Industry Connections Industry Consortium on Learning Engineering (IEEE ICICLE) defined learning engineering as "a process and practice that applies the learning sciences using human-centered engineering design methodologies and data-informed decision making to support learners and their development" (IEEE ICICLE, 2019). Bror Saxberg, vice president of learning science at the Chan Zuckerberg Initiative and member of ICICLE's advisory board, described a learning engineer as "someone who draws from evidence-based information about human development — including learning — and seeks to apply these results at scale, within con-texts, to create affordable, reliable, data-rich learning environments" (Blake-Plock, 2018).

The concepts behind learning engineering have developed over the last half century through progress in three key areas:

- human-centered design,
- cognitive science, and
- data-informed decision-making.

These three fields of study form the foundations of learning engineering processes and practices. Learning engineering also builds on and is supported by existing professional fields such as instructional design, software engineering, research, and social sciences.

The increasing complexity of the problems that learning engineering attempts to solve calls for openness to different means of communication and dissemination of information, such that the learning needs and comprehension skills of different learners are supported and facilitated. Data visualization is an essential and dynamic means of sharing information in a learning engineering context.

> According to a quote from Goodell, Kessler, Kurzweil, and Kolodner (2019):
>
> "The kinds of engineering problems associated with supporting, enhancing, and creating equitable, effective life-long learning opportunities and conditions possible in the future have not yet been fully identified. The types of problems that previous generations were tasked with solving could be addressed in particular silos (e.g., software design, instructional design, teaching, educational technologists). We've never designed educational opportunities that meet the needs of all learners, and we're at a time in history when that needs to be the goal. Technology offers new opportunities for addressing these goals, but we still need to discover the best ways of using it. Thus, we are at a crossroad where problem solving for learning requires the integration of expertise from across a variety of team members who can work together to solve the problems of today and in the future. Multiple types of expertise are required from across numerous professions that requires new sets of team competencies and processes" (2019).

Data Visualization

Data visualization is the presentation of data in a pictorial or graphical format. It allows learners to see a large amount of potentially complex information visually, so they can grasp difficult concepts or identify new patterns. Though the concept of using pictures to understand data has been around for centuries, advances in technology and its ability to analyze and present vast amounts of complicated information have been the catalyst for significant advances in methods and uses of data visualization. "Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports" (SAS, 2019).

The National Forum on Education Statistics (2016) states that "data visualization for communications purposes boils down to the following four principles that serve as the foundation for helping viewers more readily understand information:

1. Show the data.
2. Reduce the clutter.
3. Integrate text and images.
4. Portray data meaning accurately and ethically."[1]

Creators of data visualizations often have objectives that go beyond clear communication of information: they want the viewer to learn something, gain a new perspective, or change behavior based on the information presented. At its core, development of data visualizations is a learning engineering problem. Creators of data visualizations want viewers to learn something about the data that they might not be able to learn as

[1] This Forum Guide list highlights information from Schwabish, J. A. (2014).

effectively, as easily, as efficiently, or at all without visualization. A viewer's interaction with a data visualization is a learning experience intended to address specific learning objectives.

Approaching the development of a data visualization from within a learning engineering context allows the creators to consider key questions about the information they are trying to convey, as well as the best way to present it to an intended audience. They must consider: What do we want the viewer to learn? What is the story we are trying to tell with this data? How do we want the viewer to respond, from both an intellectual and an emotional perspective? These questions all help the creators to define the learning objectives. If the visualization is not carefully and deliberately designed to produce the intended objectives, then the "engineering of the learning"—which one can think of as the structure and presentation of relevant data with the intent of producing a particular understanding or behavior—is not working as designed.

Data Visualization in the Learning Engineering Context

To develop data visualizations within a learning engineering context, creators should work within the three central concepts of learning engineering referenced previously: (1) human-centered design; (2) cognitive science; (3) and data-informed decision-making.

Human Centered Design

Human-centered design is an approach to problem-solving that aims to incorporate the human perspective in all steps of solution inquiry and development. Its intent is to involve the end-user throughout all parts of the design process, through brainstorming, conceptualizing, developing, offering feedback, and implementing solutions. Typically, human-centered design is focused on iterative development that validates the problem to be solved with representative end-users and tests designs in iterative cycles. It balances the realities of time and cost budgets to determine the right level of end-user involvement and iteration cycles. California-based design and innovation firm IDEO states that human-centered design "sits at the intersection of empathy and creativity," and that their process requires getting to know the individuals they are designing for because "without them, we wouldn't know what to design, how it should work, or why it matters" (IDEO, 2019).

Within a learning engineering context, human-centered design is an important concept to incorporate into the development of data visualizations. Rather than focusing solely on the information or data to be conveyed, human-centered design of data visualizations encourages the creator to use an iterative development process, seeking formative feedback from representative users at multiple points in the process and using this feedback to improve the visualization. The creator works through a process that begins with understanding the problem to be solved, as well as the nature of the learner(s), to develop the initial visualization design, and then remains responsive to feedback sought from users and adjust the visual display accordingly.

Understanding the Problem to be Solved

Creators of data visualizations must first understand the problems to be solved by the visualizations they are designing. Put another way, what is the underlying purpose of presenting the information via a visualization?

Answering this question involves understanding the objectives for the learner(s), as well as the overall theory of action and design objectives.

Identifying the Learning Objectives—What are the designated learning objectives for this presentation of information—what should the viewer/learner know, believe, or be able to do after experiencing the data visualization? Is it a call to action? Is the intent to change the viewer's behavior? Should the viewer now be able to apply the new knowledge in a particular way? The creator needs to consider how they hope to influence the viewer's knowledge and understanding, and design the visualization in the way that best aligns with these objectives.

Meeting the Theory of Action through Design Objectives—The creator of the visualization should not solely consider the outcome objectives—that is, how the data presented is received and interpreted by the learner. They must also consider design objectives for themselves, as the presenters of the information: in short, they must think about how the design choices they make allow them to meet their objectives for the visual presentation. What are they trying to do with the information, and how does this affect the ways in which they present it? Are they aiming to create sustained attention? Do they wish to create an emotional reaction? Are they hoping to spark controversy or inspire discussion? The creator needs to consider what they want to happen as a result of the visualization, and engineer its design based on these expectations.

The creators must consider the goals of the data visualization for the particular learners. They need to ask whether the goal is to:

- Present basic or introductory information on a topic
- Advance the audience's prior knowledge
- Allow understanding of a process
- Convey a sense of scale
- Tell a story
- Evoke emotion

In many cases, the goal may combine some or all of these possibilities.

Understanding the Learners—Part of understanding the problems to be solved by the data visualization is understanding the intended audience of a given visualization and particular attributes of the targeted viewers that may affect not only those viewers' understanding of the information, but also the most effective means of presenting the data. The intended audience may have more or less background knowledge on the topic or may be more likely to be engaged by a particular type of visualization. Different personal backgrounds, whether educational, cultural, or experiential, could create challenges for some learners, leading creators to consider how variations in the visualization could mitigate some of these challenges.

Using a learning engineering approach, creators might then consider the possible ways that the intended audience might achieve the learning objectives—while incorporating the knowledge they have about the learners.

- Will the learners simply receive the information from the visualization, or will they interact with it in some way?
- Would interacting with the data presented (e.g., with a slider to show progression of a trend) help

increase understanding?
- Is there a way to evoke emotion with the visualization such that the information becomes more salient for the viewers?
- How does the visualization fit into the overall learning experience, and how does this affect the choices made about its presentation?

Soliciting and Incorporating User Feedback

After the creator has designed the initial visualization, they need to solicit feedback and make adjustments that bring the model closer to its learning and conceptual goals. Depending on the scale of the effort, testing of the design may be informal and representative of the target audience or a more formal sampling of end users. It is possible, for example, that test users would not have the expected level of prior knowledge about the topic, and more background information would need to be included. Alternatively, test users could find the presentation too simple, and need a more complex and involved display of information with which to engage. The creator could find that the information is not being interpreted by users in the manner expected and need to alter its presentation to ensure accurate understanding. In short, incorporating a human-centered feedback process is a key element of developing data visualizations within a learning engineering context.

Cognitive Science

The creators of data visualizations must have an understanding of the elements of cognitive science that provide the foundation for learners' understanding of and response to data visualization. Cognitive science draws evidence from multiple disciplines, including psychology, linguistics, philosophy, and computer modeling, and addresses the questions of how people come to know what they know, and how they apply the knowledge they have. It seeks to clarify issues such as how knowledge is organized in the mind; how people develop conceptual understanding of different subjects; how people acquire expertise in specific subjects; and how the physical structures of the brain function when learning, storing, and retrieving information (National Research Council, 2001).

Cognitive science is also informed by developmental and social psychology, as well as anthropology (National Academies of Sciences, Engineering, and Medicine, 2018). Studies in areas such as behavioral economics (the study of psychology as it relates to economic decision-making) can be applied to better understand learner motivation and engagement, and thus improve learning engineering practices. Additionally, cognitive science and learning engineering increasingly depend on the application of data science—an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data. Data science has been called the "fourth paradigm for scientific exploration" (Hey, Tansley, & Tolle, 2009, p. xix). Whereas scientists once spent much of their time observing scientific phenomenon through direct observation (such as through telescopes and microscopes), the vast expansion of available data—in varied forms—now means that they are now able to observe and/or discover scientific phenomenon through trends and patterns in the data.

Multiple principles drawn from cognitive science offer a foundation for understanding how learners understand information presented in data visualizations. After understanding the problem to be solved and the intended learners, creators of data visualizations who are using a learning engineering approach should consider some key concepts from cognitive science that are relevant to the design of data visualizations.

Table 1 lists several cognitive science concepts, adapted from Goodell, Kolodner, and Kessler (2019), and questions related to each that creators might ask when creating a data visualization.

Table 1. Cognitive Science Concepts and Considerations for Data Visualizations (adapted from Goodell, Kolodner, & Kessler, 2019).

Concept from Cognitive Science	Consideration for Data Visualization
All Learning Builds on Prior Knowledge Learning is an active process in which the learner constructs meaning by mapping new information onto prior knowledge. As we experience life, our brains try to fit new information into what we already know about the world. Learning works best when we have the background knowledge that helps us make sense of the new information. Adaptive and adaptable learning experiences offer multiple modes and levels that consider differences in the population of learners to be served.	What do we assume the viewer already knows? How do we expect experiencing the data visualization to build upon existing learner competencies? What other resources can we ask learners to consider if they don't have the prerequisite competencies?
Providing an Appropriate Level of Challenge When learners work at the edge of their mastery while maintaining high expectations, it pushes them past their current abilities. Like in athletics, a deliberate training process leads to improvement. The right level of challenge with the right level of support (scaffolding) promotes effective learning while offsetting the motivational and cognitive costs of floundering and frustration.	What level of complexity/challenge fits the learner's current knowledge and abilities? How can we incorporate elements into the visualization that appropriately align with the learner's current status?
Scaffolding Cognitive scaffolding provides the structure and support learners need to construct their own learning, moving toward tasks more complex than the learner could handle without that support. Scaffolding includes hints, prompts, conceptual frameworks (e.g. advance organizers), process support, focus support, and strategic guidance. The goal is to provide just enough support to make tasks achievable without losing useful complexity or context. Incorporating Vygotsky's concept of the zone of proximal development—in which learners are guided through tasks just beyond their current abilities—effective scaffolding helps learners succeed by supporting them to move past what they can do by themselves. It also helps learners to grow their capabilities over time.	What structures and labels within the data visualization can provide scaffolding (e.g., roll-over definitions or explanations)? What scaffolding can be provided outside of the visualization itself?
Engagement and Motivation Humans are motivated by a wide range of factors, including curiosity, meaningfulness, accomplishment, empowerment, ownership, social influence, scarcity, unpredictability, avoidance. Intrinsic motivators—internal goals—are often more powerful than extrinsic rewards. Good learning experiences and environments tap into these drives so that learners will be motivated to engage deeply, remain engaged over sustained time, and re-engage after	What aspects of the visualization are included in the design to stimulate engagement? What we know about the learner(s) and their core drives that can influence the data visualization design? Will an interactive visualization provide greater engagement or interest?

becoming disengaged. Learning experience designs can leverage “game mechanics” to tap into core drives.	
Cognitive Load Cognitive load is the effort used in working memory. Short-term or working memory, where new learning starts, has limited capacity and requires greater mental effort. In some cases, a heavy cognitive load can impede learning. Cognitive load theory suggests that learning experiences should be designed so that they don’t overload working memory.	How do we determine an appropriate and effective amount of information for the learner(s)? How can we design the learning experience not to overload (e.g., switching modes, checking for understanding, breaking up the information)?
Moving from Concrete to Abstract Understanding Learning starts with concrete information and gradually moves to more abstract understanding. Learners often need to start with imperfect but concrete conceptual models, anchored in a single context, and then over time replace them with more accurate and abstract understanding that can be applied to more than one problem or context.	What concrete example can we use to anchor learner(s) understanding? How can the visualization be used to build on the anchor toward more complex and abstract understanding of the data?
Transfer Transfer refers to the ability of a learner to apply existing knowledge to new contexts. It requires recognizing that knowledge from another situation is applicable to a new situation, knowing how to apply it, and having the desire to apply it. Fully becoming able to transfer requires practice in a variety of circumstances, as well as being able to extract what might be useful another time.	Do we want the learner to be able to apply the new learning beyond the current context? How can we help the learner make cognitive connections between the current context and new contexts?
Metacognition Metacognition is thinking about one’s own thinking and learning. A metacognitive learner knows when she is on the right track and when she is having difficulties. Metacognitive strategies include self-explanation, asking for help, starting over, and persevering. A good way of helping learners develop metacognition is to prompt them to reflect on or articulate what they are learning—metacognitive scaffolding.	In what ways can we prompt the learner(s) to think about what the data visualization shows? How can we encourage the learner(s) to question previous assumptions/knowledge that may be contrary to what the data shows?
Effective Feedback Descriptive feedback gives learners information about the quality of their ideas and performance that they can use to make adjustments. Effective feedback is specific, understandable, timely, non-threatening, and actionable.	How can we check for understanding? How will we use what we know about the learners’ understanding to provide tailored formative feedback?
Making Learning Authentic and Personally Meaningful When asking learners to carry out a set of tasks or address some challenge, they are more likely to become intrinsically motivated and engaged over long periods of time if what they are asked to do is personally meaningful and authentic. Learners can be more successful if they believe they are capable, if they can relate to a particular task, if they have available the necessary resources, and if they can see how associated activities and assessment fit into the flow of the learning.	What do we know about the learners that could direct design decisions for the data visualization? Should there be alternate versions of the data visualization or parameterized presentations that vary for different learners? How can we make the interaction with the data visualization an active learning experience?

Data-Informed Decision-Making

A learning engineering approach to data visualization is informed by data. The only way to know how well a data visualization achieves the learning objectives is to measure, and the only way to know what might be improved is to test with people that represent the intended audience. This is part of human-centered design and the iterative approach to data visualization development.

Iteratively Solving Problems

Engineering data visualizations should be an iterative process. The creators measure the effectiveness of the visualization with an audience and learn from each iteration more about areas of "friction" that might hinder a viewer's experience and prevent reaching the objective. In some cases, the data visualization will be in large scale use over time and can be iteratively improved. In other cases, the data visualization is for one-time use and the pre-release trial may be simply showing it to a few people and measuring the effectiveness. Whether a simple test or part of an ongoing program, the trial(s) add to our deeper understanding of the problem to be solved. The problem to be solved may also relate to how this information has been communicated—or not communicated—in the previous trials. Put another way, what ends will be met by presenting this visualization that have not been met on previous attempts to share this information? This consideration may relate to the nature of the data itself, e.g., the data appears dry or irrelevant to the average consumer of information, making it unlikely that they would choose to engage with it. The consideration may also relate to how the data has been presented before, e.g., it may have been presented in ways that were confusing or unclear to the typical consumer, and they mistakenly aligned their opinion of the unhelpful medium to the data itself.

Once the creators have determined whether the problem is inherent to the data itself, earlier presentations of the data, or both, he or she can now use that insight as they approach a set of more practical questions about the data visualization in order to solve the problem. For example, what about this particular data or information makes it conducive to being presented visually and creatively, and what particular aspects of the data will be most salient to viewers, thus enhancing their learning?

Using Data to Inform Data Visualization Design

Collecting data to inform design need not be an expensive or highly technical endeavor. The learning engineering approach is much more agile and formative in nature than formal long-cycle efficacy studies. Data visualization design budgets rarely will support the cost and time required to do large scale randomized control trials, but even speaking to one or two representative users can inform refinements in how a visualization is presented. In some cases, however, the data visualization may be hosted on a platform with automated collection of user experience data and opportunities for supplemental formative assessment.

The purpose is to iteratively improve the data visualization rather than prove its efficacy. Therefore, the focus is on collecting data that might serve as leading indicators for comparison of design choices. Approaches may be used that collect just enough data to make incremental improvements to the design. Early in the human-centered design process, we are concerned about understanding the learner and the problem to be solved. Later in the process, during an implementation, we want to collect data to indicate if the solution was successful: that is, did the data visualization achieve its objectives with the people who experienced it? Data collection approaches may include:

1. Informal discussions and brainstorming with people as proxies for the intended audience
2. Anecdotal feedback from people on initial mockups
3. Informal questions to indicate if the learning objectives are being met and what barriers still exist
4. If the data visualization is presented online, the user experience can be "instrumented" to collect

data about usage and outcomes.

Instrumented data collection can be as simple and low cost as a link to a single Surveymonkey™ question. More technical approaches such as presentations within intelligent tutoring systems (ITSs) may collect click-stream and sensor data about a user's interaction with an interactive data visualization and post-interaction assessment activities. These approaches may stream that data to a learning record store using interoperability standards such as the Experience API (xAPI; www.xapi.com).

Every data visualization should undergo at least one iteration of defining the problem, brainstorming approaches to solve the problem, implementing one or more approaches, and testing the approach(es) with people. If more than one approach is used, the testing may involve A/B testing to determine the best of two approaches. An agile approach to iterative design is to test ideas in very simple ways, such as first producing a simple wireframe of the design, then a more formal prototype, and then a fully functional and graphically designed iteration of the visualization. This allows for iterative design and testing of concepts without multiplying the cost.

Another way to keep the cost and effort down but still allow for data-informed iteration is to use informal focus groups, such as asking colleagues not connected to the project or subject matter experts to look at a conceptual design and give informal feedback or answer a question. This is not evaluation data but helps to check assumptions.

Process Model

Figure 1 shows an example process model for an iterative learning engineering approach to data visualization development.

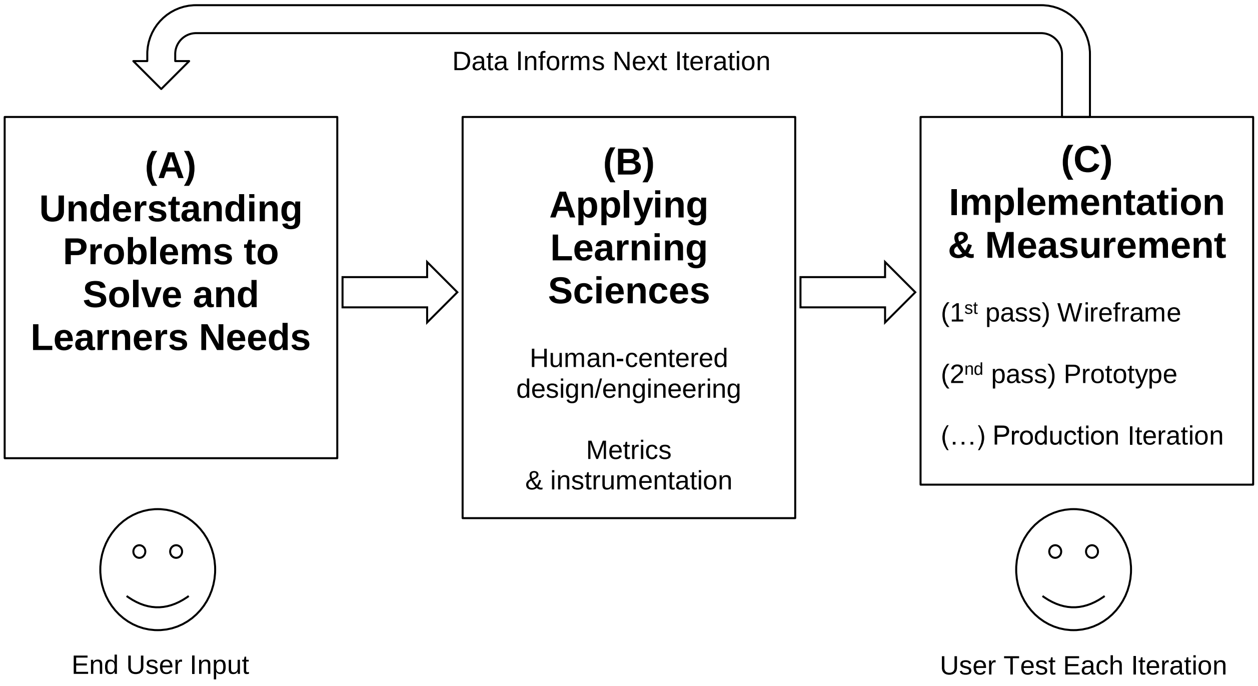

Data Informs Next Iteration
(A) Understanding Problems to Solve and Learners Needs
(B) Applying Learning Sciences
Human-centered design/engineering
Metrics & instrumentation
(C) Implementation & Measurement
(1st pass) Wireframe
(2nd pass) Prototype
(…) Production Iteration
End User Input
User Test Each Iteration

Figure 1. Process model for learning engineering approach to data visualization

The approach starts with understanding the problem to be solved by the data visualization and the people who will interact with it. Then learning sciences concepts and human-centered engineering methodologies are employed for an initial solution design (e.g. wireframe or prototype). Part of this step is instrumentation design, i.e. deciding what data to collect to inform iterative improvement and how it will be collected. Finally, the design is implemented and tested with representative end users and data collected to inform decisions about the next iteration. The process may employ one feedback cycle or many depending on the nature of the project. Analysis of the feedback gives new understanding of problems to be solved in the next design iteration to further optimize the data visualization for the intended purpose and audience.

Conclusion

When designing data visualizations within the context of learning engineering, it is important to remember that data visualization is not the only tool in the toolbox. Visualizations are excellent for capturing learners' attention, engaging viewers' emotions and reactions, and making key points salient, but they are rarely intended to function alone. Instead, they are an integral and influential part of the larger whole—in this case, the specific lesson or overall learning environment. Creators should ask themselves, "what other learning experience(s) might be combined with the data visualization to enhance or complete the experience?"

These considerations and additional learning elements are not meant to be determined or created outside of the learning engineering context for data visualizations, but instead following the same path and intent. Thus, these elements should also be developed via incorporation of human-centered design, key cognitive science considerations, and data-informed decision-making. Designing the full learning experience within the context of these crucial learning engineering concepts allows the focus to stay on the needs of the learners, and to approach the presentation of information as an iterative and dynamic process.

References

Blake-Plock, S. (2018, January 29.) "Learning engineering: Merging science and data to design powerful learning experiences." Getting Smart. Retrieved from https://www.gettingsmart.com/2018/01/learning-engineering-merging-science-and-data-to-design-powerful-learning/

Goodell, J., Kessler, A., Kurzweil, D., & Kolodner, J (2019). Competencies of learning engineering teams and team members. Presentation at the 2019 IEEE ICICLE Conference, Arlington, VA.

Goodell, J., Kolodner, J., & Kessler, A. (2019). Learning sciences concept cards. Presentation at the 2019 IEEE ICICLE Conference, Arlington, VA.

Hey, T.; Tansley, S. & Tolle, K. (2009). The Fourth PARADIGM: Data-Intensive Scientific Discovery. Microsoft Research. Redmond, Washington. Retrieved December 17, 2019 from https://www.immagic.com/eLibrary/ARCHIVES/EBOOKS/M091000H.pdf

IDEO. (2019). "IDEO: Tools." Retrieved December 17, 2019 from https://www.ideo.org/tools.

IEEE ICICLE. (2019). "What is learning engineering?" Retrieved December 17, 2019 from https://www.ieeeicicle.org/

National Academies of Sciences, Engineering, and Medicine. (2018). How People Learn II: Learners, Contexts, and Cultures. Washington, DC: The National Academies Press.

National Forum on Education Statistics. (2016). Forum Guide to Data Visualization: A Resource for Education Agencies. (NFES 2017-016). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

National Research Council. (2001). Knowing What Students Know: The Science and Design of Educational Assessment. Washington, DC: The National Academies Press.

SAS. (2019). "Data visualization: What it is and why it matters." SAS Insights: Big Data Insights. Retrieved December 17, 2019 from https://www.sas.com/en_us/insights/big-data/data-visualization.html.

Schwabish, J. A. (2014). An economist's guide to visualizing data. Journal of Economic Perspectives, 28(1), 209–234. Retrieved from http://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.1.209

Simon, H. (1967). The Job of a College President. The Educational Record, Winter, 1967, 48: 68-78. (Reprinted from the Winter 1967 issue of The Educational Record, published by the American Council on Education, Washington, D.C.) Retrieved from http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=33692

CHAPTER 3 - THE GIFT OF SCRUTABLE LEARNER MODELS: WHY AND HOW

Judy Kay[1], Diego Zapata-Rivera[2], and Cristina Conati[3]
University of Sydney[1], Educational Testing Service[2], University of British Columbia[3]

Introduction

There is a growing call to build intelligent systems in ways that ensure the user can understand and control them. This is particularly clear in Europe's General Data Protection Regulation (GDPR) that came into effect in 2018 (European Union, 2020). In this chapter, the GDPR is important, not so much for the legal details, but rather for its spirit. This reflects the wishes of many people who want to control their own data, including the right to be forgotten, the right to explanations of intelligent systems that use their data and the ability to control that use.

These important needs have recently generated a large and fast-growing body of work that aims to address these concerns in the context of systems based on algorithmic decision making and Artificial Intelligence (AI). It goes under various names, such as explainable (including XAI, for explainable AI), intelligible, interpretable, comprehensible, human-interpretable machine learning, transparent, glass box, understandable, as well as scrutable and the work is linked with goals of fairness, accountability and trust (Abdul, Vermeulen, Wang, Lim & Kankanhalli, 2018; Arrieta et al., 2020; Bunt, McGrenere, & Conati, 2007; Millecamp, Htun, Conati, & Verbert, 2019). That work has covered a huge scope, with a strong focus on decision making systems and ways to enable people to understand the decisions and predictions made by intelligent systems (Eiband, Buschek & Hussmann, 2020; Miller, 2019; Rader, Cotter & Cho, 2018). An analysis of trends and trajectories in this work, by its many names, highlighted "a fundamental challenge in balancing these powerful capabilities provided by machine learning with designing technology that people feel empowered by" (Abdul, Vermeulen, Wang, Lim & Kankanhalli, 2018, p.1). With so many terms, often used in different ways, and often interchangeably, this chapter uses the term scrutable to describe a particular class of learner model visualization.

The Artificial Intelligence in Education (AIED) community has a long track record in learner modeling and in that work, interfaces onto these models are called Open Learner Models (OLMs). One synthesis of OLM work is described in the framework called SMILI - Student Models that Invite the Student In (Bull & Kay, 2007, 2016). SMILI introduces additional terms, such as negotiable models, which relates to the nature of the learner's interaction with a model. Since this chapter is intended to inform recommendations for building authoring frameworks for Intelligent Tutoring Systems (ITSs), we focus on scrutability which we now define.

Scrutability, as we use it, has two key elements which distinguish it from all the many other terms. The first is its learner-centered view, based on the questions that a learner should be able to answer by careful study (scruting) of a scrutable learner model. The term, scrutability, reflects the effort a learner will need to make to scrutinize a system if they are to understand its elements – which include potentially multiple sources of data the ITS keeps in the learner model, the data provenance and reliability for the individual learner as well as the ways the learner model is interpreted and then used.

The second key element is that the term scrutability requires that we build the ITS with all the infrastructure needed for the interfaces that can support user scrutiny and control. Scrutability must also be reflected in the commitment of ITS designers, programmers and authors to build an ITS so that it can support the learner in scrutinizing the ITS personalization. An authoring tool like the Generalized Intelligent Framework for Tutoring (GIFT) can play a critical role in making this manageable.

We now introduce a high level, intuitive definition of scrutability in terms of a set of "Competency Questions" that competently scrutable systems should enable a learner to answer about their learner model:

- What does the system "believe" about me?
- Why does it "believe" this? What data and processes determine the system's beliefs?
- Where does that data come from?
- How is that data used to make inferences about me?
- Where does my data go?

We consider such scrutability as important in an ITS for the following broad reasons:

- to facilitate learner responsibility;
- to enable learner control;
- to allow the learner to be a partner in the modeling process;
- to provide a foundation for sharing the ITS author's view of the learner modeling processes;
- to accommodate the fact that learning data and machine models are incomplete and inaccurate;
- to increase learner trust in the system and support accountability.

Such scrutability is particularly important at this point in time when the maturity of technology is opening new possibilities for lifelong and life-wide learning. This means we can create ITSs for long-term use, with student data kept in long-term learner models that can drive the personalization. New technology provides a growing range of rich "sensors" that can collect data about the learner, for example, based on speech and image analysis as well as data from worn sensors. These can be combined in new ways with evidence from interaction with the ITS. This can provide a rich learner model to drive personalized learning activities and assessment. It can also support broader explainability of the ITS behavior (e.g. Conati, Barral, Putnam, & Riegel, 2020), in line with the aforementioned general objective of enabling people to understand the decisions and predictions made by intelligent systems.

The next section introduces some examples of scrutable systems to help the reader build an intuitive sense of the nature of scrutability. Then we define scrutability and discuss how an ITS authoring platform like GIFT can provide the essential infrastructure for authoring scrutable learner models and ITSs.

Some Examples of Scrutable Learner Models

This section introduces three examples of OLMs. Each has been chosen to illustrate a different form of OLM and different forms of scrutability and lessons about how to create OLMs that are scrutable.

Example 1: OLM for group contributions in multi-week group collaboration

Our first example of a scrutable learner model is the Narcissus Open Learner Model (Upton & Kay, 2009). This was created to support groups of students in a semester-long software engineering capstone project subject. Students used an agile programming approach where teams of around five created a substantial piece of software for an authentic client. In the agile method, students have various roles, including a team manager, a tracker, client liaison and programmer. They used a tool called trac, similar to github[2]. In tools like these, there is a wiki for information like minutes of meetings. Another key form of collaboration support is called issue-tracking in Github. This enables a team to define each task/issue identified as needing to be done, allocate it to a team member and track the progress to completion. The code is managed with a version control tool, git; such version control systems are widely used by programmers because this helps them manage the evolving software in its many versions, making it easy to see when changes were made, by whom and one can go back to earlier versions.

Figure 1 shows a stylized version of the OLM. It has one row per day; in the figure the user has configured the display to show January 12, at the bottom, to January 25 at the top. As indicated by the legend at the right of the figure, the leftmost cell in each row is for wiki activity. This user had a burst of high activity for their first four days, from Jan 12-15 and then a lower level, indicated by a lighter red on Jan 16, the 4 days with no activity. The blue column is for activity on the Git code version control (described above) and the third orange column is for activity on issues.

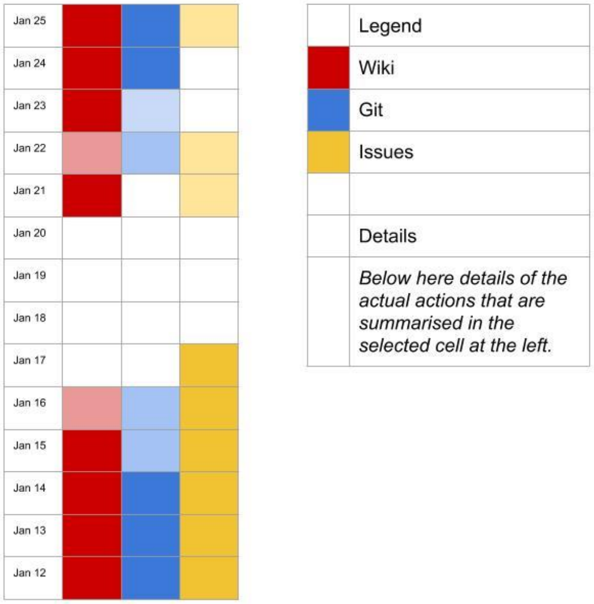


Figure 1. Example of OLM showing an individual student's daily contribution to group work in terms of their contributions to the wiki, version control system and issue-tracking.

We note that this visualization makes use of very simple evidence about each student's activity. For example, the amount of programming work is measured simply as the lines of code added to the version control system. While this is clearly a very crude measure, it is used often in industry. But importantly, to meet the design goal for scrutability, Narcissus allowed the individual student to decide the threshold for line counts to appear, as low, medium, high or very high levels. The student could alter such thresholds at any time for

[2] https://github.com/

each of the media, wiki, issues and version control. While we did not collect detailed data on this, individuals needed to alter these values to match their roles. For example, the team leader should have far more activity on the wiki, the person responsible for tracking the progress of the group against the plans should have a lot more activity on the ticket system, and those with primary responsibility for the programming should have had more activity on the version control system. So, people with these roles had reason to alter these values when they presented their progress to the group and their tutor.

Even more important is the context of the design and use of Narcissus. The driver for its design was to help the user identify problems in the group functioning and track the impact of initiatives to tackle these. For example, such group work often needs to deal with individuals who fail to contribute their fair share of work and to fulfill tasks allocated to them. Narcissus was used in the weekly standup meetings with the tutor, with each team member using it when explaining their contributions and any problems. It was also used in weekly group manager meetings with the course coordinator. Each manager used the Narcissus interface to explain how their group was progressing and to point to problems for discussion. This context meant that the people using the OLM knew the important additional information such as the role of each person and the challenges the group was facing. Different group roles should have been reflected in different profiles in the OLM. For example, a person with the job of typing up the minutes for weekly meetings on the day of the meeting should have had wiki activity on that day each week.

We now explain the panel at the right in figure 1. This starts with the legend. Below that, the interface provided additional details of any cell the user selects. For example, if this was the wiki, the interface showed a list of links to the wiki actions. So, it was a matter of one click on a cell to drill down to the detailed actions of the user. In practice, students used this in weekly meetings to quickly show what they had done and to illustrate their problems.

The interface also provided an explanation of the mapping from the count of lines of code to the color gradient. A separate configuration interface enabled the user to set these thresholds. So, for example, a person whose job was to write a substantial body of code over 2 weeks should set quite high threshold counts in line with the amount of code they planned to write.

This class of interface can enable a student to answer several important questions:

- How much did I contribute on the wiki, code version control and issues over the time period?
- How much did others contribute, compared with me and with what our group had planned?
- How does last week compare with the week before?
- As the group manager, I would expect to be the most active person on the issue system since I allocated each task to a team member and mark it off as complete - is this what is happening?

At any stage, a user can scrutinize any unexpected activity using the panel at the right to see the details. When we used this interface, the teaching staff were able to identify group problems early enough to help groups address them. We were also able to identify individuals who needed additional follow up to determine how to address apparent problems.

Narcissus is a simple OLM visualization that was iteratively designed (Upton & Kay, 2009). It is scrutable in the sense that a user can determine:

- what evidence about them was used to determine the value of each part of the visualization (for example, in figure 1, the precise code committed to give the shade of blue in a particular cell);

- how the evidence is interpreted (such as the example labeled F in figure 1);
- how to change the interpretation (for example, altering how many lines of code they wrote and committed in a day).

Example 2: Group OLM for teachers

We now consider a very different learner model for group work, now in a tabletop classroom. We chose this case study for two reasons. First, it has a link with the Narcissus OLM in that both provide visualizations of group work activities. But the second reason is that it introduces several important differences. In this case, the OLM was designed for use by the teacher. But even more important is its evolution over the research project. It began in the lab (Martinez-Maldonado, Kay & Yacef, 2013) with sophisticated machine learning based on touches at the tabletop as well as audio of the speech during collaboration. In moving to real classrooms, it became simpler.

The needs that drove the design of the visualization for the authentic classrooms came from academics responsible for a business subject (Martinez-Maldonado, Dimitriadis, Kay, Yacef & Edbauer, 2012) and two human-computer interaction subjects (Martinez-Maldonado, Clayphan, Yacef & Kay, 2014). The classroom teachers wanted to be confident they understood the OLM and its internal inference mechanisms, so scrutability concerns drove the design. In both cases, students worked in a tabletop classroom, with 5 tabletops, each with a group of 3-5 students. Data about student work came from the tabletop hardware plus Kinect cameras which analyzed the tabletop screen image to map each touch to the student who did that action. This was used to support the interface and to build the model of what each student did. In our lab studies, we also had directional microphones which distinguished each student's utterances.

Figure 2 shows an anonymized version of the visualization designed for the first dashboard, based on lab studies with groups of three students. We built an OLM based on analysis of speech audio as well as touches at the interface as students did a concept mapping task. Figure 2 shows information about the collaboration in three groups, each with three visualizations. The leftmost shows a measure of the level of collaboration. Group 1 has a very high score, Group 2 has a score that puts it just above the middle of the scale of co-operativeness and Group 3 is less co-operative. This co-operation score was based on a complex model from machine learning on data for the learner actions and utterances (Martinez-Maldonado, Kay & Yacef, 2013). The interface provided no way to scrutinize how these scores were determined and a meaningful explanation would be difficult to create. In the user studies, teachers commented about this and said that they were unsure they could trust such a score.

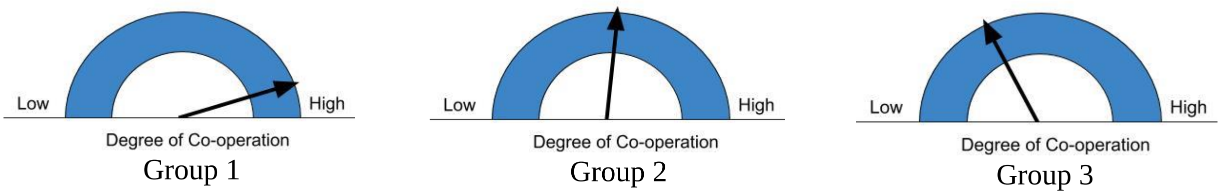


Figure 2. The first OLM designed for lab studies where groups of three students created a collaborative concept map.

Figure 3 shows a stylized form of the OLM we created later for teachers in an authentic tabletop (Martinez-Maldonado, Dimitriadis, Kay, Yacef & Edbauer, 2012) in which the teacher had designed concept mapping activities. Students worked to create concept maps that were automatically analyzed to identify core elements the teacher had expected students to incorporate in their answers. Each table was color coded, and the visualizations matched that color. The OLM had one histogram for each group – figure 3 shows just two examples of these. Each histogram has one bar for each student in the group – the top Red-group has five students and the lower Orange-group has six. The length of the bar for a student is a measure of their contributions. The bright-colored ones matched elements defined by the teacher. The lighter part beyond this may be correct or not; if there are many such contributions in a group, the teacher should spend time with them to determine if they have conceptualized the problem in a different but valid way or if they have misconceptions that the teacher should discuss. In figure 3, it is easy for the teacher to see that the five students of the red group have each created some elements and most of these match ones the teacher defined. The orange group has created a similar number of elements but most of them do not match those the teacher expected. It is notable that two of the students created many non-matching elements. At a glance, the teacher can see that the Orange-group needs more attention at this time.

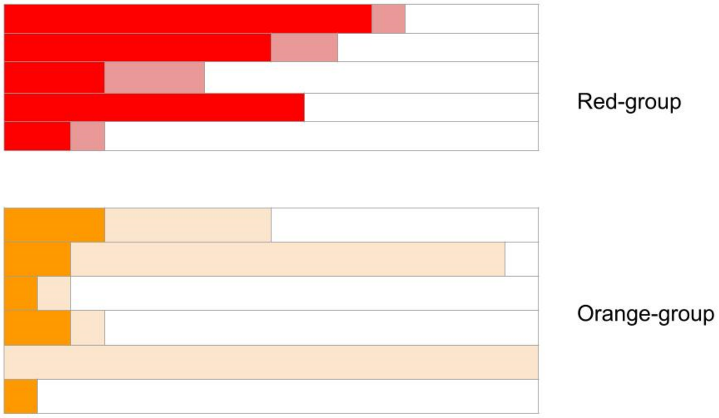


Figure 3. Visualisation for two of the groups in a tabletop classroom control for groups of up to 6 students working on collaborative concept mapping activities.

We note that this OLM is characterized by quite simple measures. The only data it has available comes from the touch actions at the tabletop. From these, the software can assess what part of the concept map was created by each student. It can then compare each such contribution with a set the teacher defined. Clearly, a student could have made very valuable contributions in the group discussions of concepts, relationships and how to improve the group's concept map; the OLM had no data about such verbal contributions and so it does not represent that at all.

This OLM enables a teacher to see a quick overview of each groups' activities to answer several questions:

- How much is each group adding to their concept map?
- How much of the activity matches the map features the teacher expected in good answers?
- How much is each individual contributing to the activity by adding to the concept map?

The next level of scrutability would be to make the actual detailed contributions available to the teachers.

However, this was not needed for the classroom role for the OLM, as an overview for group and individual activity as a complement to the actual concept map.

Example 3: OLM to visualize Knowledge Components in a large learner model

We now introduce the Scrutable Inference Viewer (SIV) (Kay & Kummerfeld, 2013; Kay & Lum, 2005; Kay, Li, & Fekete, 2007) as an example of an OLM interface where scrutability was a foundational design goal. The work aimed to create an OLM for large learner models based on a visualization to enable a learner to answer several questions about the system's model of their knowledge:

- What aspects of my knowledge does this system model?
- What is my overall progress?
- What does the model show that I know?
- And how high does it rate my knowledge of each aspect?
- How confident is the system of its rating of my knowledge?

Figure 4 shows a stylized and simplified version of the SIV visualization for a C programming course where students did weekly programming tasks, homework and quizzes. All these could provide evidence about each student's progress. There are three key aspects for interpreting the model: the size of the text describing the component of the model, the color of the text, and its horizontal position. For the SIV OLM in figure 4, the student has selected a focus learning topic, in this case "Memory models" and so, SIV made this the largest font, as in the figure.

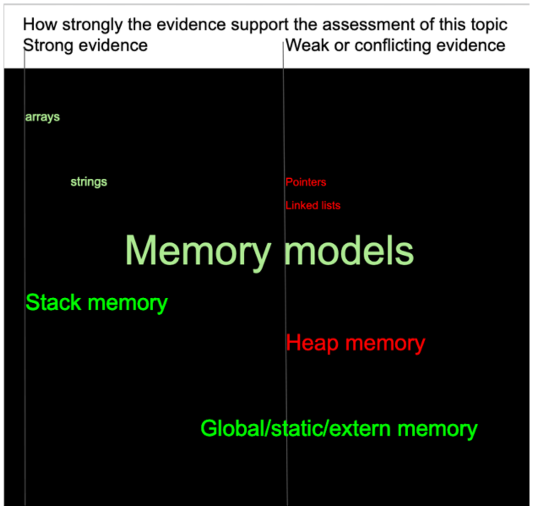

How strongly the evidence support the assessment of this topic
Strong evidence
Weak or conflicting evidence
arrays
strings
Pointers
Linked lists
Memory models
Stack memory
Heap memory
Global/static/extern memory

Figure 4. An example of the SIV in the OLM on a C programming subject.

The text that is the next smaller size is the aspect that is closely related to the focus concept in the learner model ontology. In this example, the three aspects of memory models are "Stack memory", "Heap memory" and "Glocal/static/extern memory". Concepts that are more indirectly related to the focus concept are shown even smaller. SIV is an animation and if the student selected "Stack memory" it would become the focus and its text would become the largest.

The color of each concept reflects the learner's knowledge, with bright green indicating high levels of knowledge and bright red, as in “Heap memory”, indicating the student does not know that aspect. Lighter green, as in “Memory models” indicates lower levels of knowledge demonstrated by the available evidence.

The horizontal position indicates how confident the system is about the assessment of the student's knowledge. The labels at the top of figure 4 indicate the alignment in terms of system confidence, with topics at the left, such as “Stack memory” having consistent evidence from multiple sources indicating the student knows about this topic. The topics at the right, like “Heap memory” and more distantly related topics, “Pointers” and “Linked lists” are modeled with low confidence. This could be because there is very little evidence, for example the student has done one week’s work and it was all incorrect. A low confidence value that is light green or red may be due to conflicting evidence as the student has some correct and some incorrect work.

One implementation of this visualization approach was evaluated with large models (100, 300, 500 and 700 concepts) and users were able to answer questions about the system’s assessment of them and the system's certainty. User accuracy and speed was high for up to 500 concepts, dropping somewhat at 700.

The SIV visualization was incorporated in various support interfaces. For example, one of these provided the details of the evidence used to infer the value shown. This was presented as a list of class activities, such as playing to online lecture content, performance on weekly homework and class activities. The list provided direct links to these. For example, for lecture content, the link took the student to that lecture material.

Defining Scrutable Learner Models for a GIFT Authoring Context

The examples above explored various aspects of scrutability and were chosen to give the reader an intuitive foundation for a more formal definition of a scrutable learner model. We do this for the case of ITSs such as those that GIFT authors create. The one essential aspect is that they have a learner model and a suitable OLM interface could provide many benefits, scrutability being one. We begin this section with some of the history for the first use of the term scrutable for learner models. This is important because some authors have used the term to mean different things, usually ones that are equivalent to other terms such as transparency and taking a completely different view from Kay’s initial vision for whole-system, user-driven design as a comprehensive foundation for learner control and responsibility.

Kay chose the term *scrutable* to describe the driving design approach underlying a series of user modeling systems. These began with the um toolkit (Kay, 1994, 1998), then the Personis user model server (Kay, Kummerfeld & Lauder, 2002), and the PersonisAD user modeling framework for rich sensor data and ubiquitous personalized systems (Assad, Carmichael, Kay & Kummerfeld, 2007) such as the personalized system for scrutable modeling of an indoor location (Niu & Kay, 2010). This was a foundation for a body of work exploring various challenges of building scrutable learner and broader user models (Kay & Kummerfeld, 2013) as well as work that positioned scrutability in the landscape of OLMs (Bull & Kay, 2007, 2016). While the idea of a scrutable learner model is abstract, this chapter is based on the actual user and learner modeling systems created by Kay and colleagues. This describes the smallest unit modeled as a component. The component types are:

- Knowledge Components (KCs), such as whether the learner knows how to read simple loops in the C programming language - this term matches work in Koedinger, Corbett, and Perfetti (2012);
- Belief, to represent aspects that the ITS author considers to be misconceptions;
- Preferences, such as the learner’s preference for mathematical examples to illustrate code;
- Goals, such as whether the learner is aiming for minimal competence or high expertise mastery;
- Attributes for other aspects such as the learner’s height or having color blindness.

This section deals just with KCs. It uses an illustrative example from the context of an ITS for the C programming language for the student's knowledge of how to read simple C loops. An ITS could collect evidence about this in many ways. For example, students could do Multiple Choice Questions (MCQs) that test their ability to trace code segments. To take one other example, the student could be asked to write code.

Tables 1 and 2 define a scrutable learner model in terms of the core types of questions that it should enable a learner to answer. Table 1 has the questions about ***what*** the system believes, whereas Table 2 deals with questions about ***why*** the system holds a belief. After this, we consider ***how*** the learner might contribute to the system's reasoning and scrutinize use of it.

Each table has three columns. The first column has the *abstract* questions. These use the specialized terminology of the field, making it easier for a reader to link this chapter to the literature. The second column has a more user-oriented version of each question. These take an anthropomorphic view of the ITS where the learner asks what the system believes. The final column shows what information the OLM needs to make available at the OLM interface for the learner to be able to answer their questions. Our approach to defining scrutability reflects the deeply user-centered design needed for scrutable learner model visualizations and other scrutability interfaces linked to the OLM.

Table 1 has the two core learner questions to understand what the system believes about them, about the components modeled and their value. To ensure an OLM can provide the information to answer them, the ITS author needs to create each component with:

- *an identifier* for internal use in the ITS authoring system e.g., *read-loops;*
- *a short description* e.g., *reading loops;*
- *a detailed description* e.g., *your ability to read C loops with a single loop, both by tracing loops code and by reading at a higher level, for example to read a loop that repeats 10,000 times*.

An ITS authoring system normally supports definition of components, including creating an internal identifier. A scrutable OLM needs to be able to also have a short string that has been written to make sense to learners in an overview OLM, like those discussed above as well as in the many OLMs with a small set of skill-meters. The author may decide on an identifier identical to the short description. However, it is preferable to avoid this. The identifier is for use by the ITS authors and should be chosen to follow programming style guidelines (Kernighan & Pike, 1999). The short description should be designed for the learner and user studies, and could evaluate whether students actually can understand it in the context of the learning environment.

The longer description should give richer information to the learner. It provides a valuable way for the author to communicate the thinking about the domain to the learner. Both descriptions should be designed for the context(s) of use in an independent OLM (Bull & Mabbott, 2006; Bull & Kay, 2013; Kay, 1994) or in one or more ITSs, such as our earliest use for teaching about unix and a text editor (Kay, 1994). For example, in Figure 4, some of the reddest components are *Core*, *Memory Models*, *Pointers* and *Similar concepts in both C and Java*. These are four of the seven highest level learning areas that were shown to students throughout the course and should have been meaningful for students using this OLM.

Table 1. Support for a core scrutability question: What does the system "believe" about me?

Abstract questions	User questions	Information in the answers
What is the *meaning* of the components in the learner model ontology?	What aspects of my knowledge, misconceptions, goals, preferences and attributes does this system model about me?	Each OLM component needs a short identifier and a *detailed description* of the meaning, expressed so that the target student for the ITS can understand it.
What is the *value* of each component of my learner model?	Does the system think I know?	The OLM shows the learner a meaningful value for each component. (e.g., that the system rates this component as known).

The *detailed description* should minimally use terms that are consistent with those in the rest of the teaching materials. One could also go beyond this to make the scrutable learner model serve as a teaching tool. We did this in a long term project modeling knowledge of both a text editor (sam) and the unix operating system (Cook & Kay, 1994). The OLM had a personalized explanation sub-system. This was a useful standalone tool for use over the years that a student needs to build a deep knowledge of these tools. Its explanations of components took into account the terms that the learner understood. So, for example, a novice who did not know the term *current window* would see it described as *the window with the darker border*. This approach meant that the student could use the scrutiny sub-system to learn about components they did not know. The system was used over several years; some students who used the OLM in one subject returned to use it in this way in the following two years to master new aspects of the text editor.

The first row of Table 1 refers to the learner model ontology. We use this description because it is important to recognize that the design of the learner model requires the ITS author to create an explicit specification of his or her conceptualization of the learning domain. The author must determine the concepts modeled and the relationships between them (Gruber, 1995). In many ITSs, the learner model is a simple overlay on a list of KCs, so there is no need to define relationships. But even in very simple learner models, it is very common to have at least the structure of ***prerequisite*** and ***is-part*** relationships. For example, for students learning the C language, a ***prerequisite*** for *Pointers* is knowledge of variables (modeled as ***part-of*** *Similar concepts in both C and Java*). In the SIV OLM in the above figures, the ontology for the learner model was automatically generated by analysis of an online dictionary (Apted, Kay & Lum, 2004). For the C course of Figure 4, the teacher created this as a glossary for the subject. In the HCI course described above, we used an online dictionary (Usability First). We used a tool, MECUREO, that analyzed such online dictionaries to automatically construct the ontology by using the main terms as concepts and mining the entries to identify relationships between them.

A broadly similar approach would be appropriate for an authoring tool like GIFT. The author could write the glossary for the subject. In a very simple form, this would have one entry for each component in the learner model and this could serve as the *detailed description*. The authoring interface would support creation of each glossary definition. The student should be able to access the glossary in two ways. It could be available as a glossary page in the ITS. In addition, each *short description* for a KC in an OLM could be linked to the relevant glossary item. In an interface like in the figures above, where the learner sets a focus concept, its glossary description could appear at the bottom of the screen.

We now consider the second row of Table 1. This is a core function of OLMs. They enable a learner to answer a question like, “Does the system think I know how to read simple C loops?” This basic information is a starting point for scrutability as we now consider in Table 2 which deals with learner’s questions about the whole process of collecting data, processing it and then using the processed form to conclude how well the learner knows an aspect.

The first row in Table 2 addresses the evidence sources that the GIFT author creates. For scrutability, we need to add a small overhead to that process so that the author includes a simple string that describes each source. This needs to be written in a manner that will make sense to the learner. This step may provide a useful way for the author to capture the rationale for using a particular data source. For example, to model whether a learner knows how to read C loops, some examples of evidence sources are:

- self-rating questions where the learner is asked to assess their own knowledge - those could be automatically generated from the long description described above;
- short answer code reading tasks where the student is presented with code and asked to type in what is printed or the value of a particular variable at a nominated line;
- camera-images captured as the student does the tasks above so that these can be analyzed to recognize when the student was frustrated.

Table 2. Supporting core scrutability question: Why does it “believe” a component has this value?

Abstract questions	**User questions**	**Information in the answers**
What evidence *sources* are used to inform the value of this component?	What raw data did the system use?	Describe each evidence source so that the learner can judge its reliability for assessing their knowledge.
How are the raw data streams filtered, interpreted, wrangled, munged?	How was raw data processed?	This is evidence used in the current ITS and may not include all evidence in the learner model.
How is evidence *interpreted* at this *time*?	How did the system use that processed data to determine how well I know this at a particular time?	Explanation of the process used to interpret the current set of evidence about this component at the time it was used.

The second row relates to the processing used to translate raw data into the processed form kept in the model. In the examples above, the raw camera images may not be kept at all. Instead, the data may be processed to give time-stamped frustration scores. There are many ways that raw data is processed, such as dropping outlier values, filtering out noise, reducing its volume and building sophisticated machine learning models. For scrutability, the author needs to capture the details of what has been done and why so that this can be made available to the learner.

While the second row relates to learning evidence that is stored in the model, the third row calls for explanations of the process used to conclude a value at a particular time. For example, suppose we have the following evidence about our running example for the read-loops KC:

- yesterday: the learner rated their knowledge of reading C as very low;
- this morning: they correctly answered 5 code tracing reading tasks of increasing difficulty;
- this afternoon: they correctly answered 2 easy abstract code reading tasks and then had incorrect answers to 3 harder ones;
- camera-based frustrating ratings were low in all cases above.

There are many ways to interpret this collection of evidence to conclude a value for the KC. In a scrutable OLM, the learner could delve down into this component to see an explanation for the method the system used. Many ITSs and OLMs have rather simple interpretations of evidence that are easy to explain. For example, a very simple approach interprets a KC as known when the learner correctly completes three assessments in a row. Importantly, at different times, the same component will have different values. In the example above, the only evidence available about this KC yesterday was the learner’s self-rating.

The underlying vision for making learner models scrutable is to provide a foundation for broader learner control with interfaces designed to enable the learner to:

- control what evidence each application can *add* to the learner model;
- control what information an application can *access* from the model;
- *add new evidence* about any aspect modeled;
- *delete evidence* about any aspect modeled;
- and discover how the learner model is used in a teaching system.

Discussion and recommendations for GIFT

There are different levels of scrutability. The level of scrutability implemented in the system depends on characteristics of the learning activity and the users who will be interacting with the learner model. For example, the types of interactions with the learner model may be very different during a training exercise compared to exploring the learning model during an after-action review session. Also, different user groups have different learner model information needs and may benefit from particular external representations and guidance (Zapata-Rivera, Graesser, Kay, Hu, & Ososky, 2020; Zapata-Rivera, Hansen, Shute, Underwood, & Bauer, 2007).

We now propose several very simple ways to add scrutability to the GIFT authoring system. Our goal is to provide pointers to the ways that scrutable OLMs could be added to ITSs created using GIFT. We organize these around the core abstract questions described above.

What is the *meaning* of the components in the learner model ontology?

GIFT’s interface for creating the learner model should have interfaces for the author to define the *identifier*, *short description* and a glossary entry for the *long description*. Each new glossary item should be incorporated into a glossary that serves as a learning resource. The entries should also be used as long descriptions available from OLMs. This minimal form of support for this core question has modest cost and high benefit in terms of supporting communication between the learner and author about the core learning goals in the course. The glossary could be written to include explanations of the importance of each aspect and this too could help the learner understand why the author included each aspect. Our descriptions above touched on more sophisticated approaches. For example, just as a teacher could create personalized teaching materials in an ITS, so too, the glossary could, like any other learning resource, be personalized. This is likely to have the highest pay-off by personalizing it so that a novice is presented with simple and minimal information; as they progress, more detail could be revealed and descriptions could make use of concepts already mastered.

What is the *value* of each component of my learner model?

One of the core questions that an OLM should enable a learner to answer is:

- What does the model show that I know?

Designing a visualization for a learner model requires a choice about the way to present this information. One very simple option is to report a binary value for each KC, indicating it is known or not. There are many other options.

This chapter has presented several examples of OLMs that present the values of KCs in a learner model. The first example, Narcissus, shows cells, each of which can have one of 5 possible values, where the learner sets the thresholds that determine these. The second set of examples, for tabletop learner models, included a gauge with a continuous set of values, shown in Figure 2. Figure 3 shows the learner model as histograms. This is such a widely used visualization that we can be confident that learners will be able to easily understand it. Figure 4 illustrates the approach of VLUM which uses color to indicate whether each KC is known (bright green), not known (bright red) or has an intermediate value. If the whole model is green, the learner can readily see they are doing well. VLUM also shows the system's confidence in the value. If there is more evidence about a component, it is at the left and if there is no evidence, or the evidence available is conflicting, it is at the right.

Considerable work has explored many other ways to show the value of KCs in an OLM, notably in the body of work by Susan Bull, such as (Bull, 2016; Guerra-Hollstein, Barria-Pineda, Schunn, Bull, & Brusilovsky, 2017). There are many options for conveying the uncertainty of the values, such as those described in (Al-Shanfari, Baber & Epp, 2017; Al-Shanfari, Epp & Baber, 2017; Al-Shanfari, Epp & Bull, 2016; Bull, 2020; Hooshyar, Pedaste, Saks, Leijen, Bardoneet et al., 2020) and we need careful evaluations to determine whether learners can understand them and, more importantly, how they impact learning, as in recent work (Al-Shanfari, Epp, Baber & Nazir, 2020).

What evidence *sources* are used to inform the value of this component?

An authoring interface could readily include support for authors to add a user-oriented description of each raw data source that is used by an ITS. This can also provide a way for the ITS author to share the rationale for using each source.

How are the raw data streams filtered, interpreted, wrangled, and munged?

This is the form of the data that is actually used by the ITS and, like the raw data sources, an authoring system could provide support for creating user-oriented descriptions of these processes.

How is evidence interpreted at this *time*?

A key idea in this work is that there are typically multiple ways to interpret learning data. Interfaces could allow learners to set their own standards. We saw this in the Narcissus configuration above. We conducted exploratory user studies with a SIV interface. Medical students saw an OLM based on their answers to a large MCQ bank covering over 600 learning topics. Some altered the 50% correct answer rate for making a concept green, preferring a higher standard so that only topics where they scored at least 80% were green. We would like to see exploration of OLM interfaces that allow learners to take such responsibility for judging their own learning. More broadly, in a scrutable system, the learner should be able to determine how the system interpreted the available evidence of learning. Several approaches for making available the internal mechanisms used by the system to determine learner model values have been explored (Bull, 2016; Van Labeke, Brna & Morales, 2007; Zapata-Rivera, et al., 2007). One step beyond this is to offer the learner choices in interpretations. This includes access to the algorithm along with details of the raw data, its processing and then interpretation. This may open the possibility of some learner control of some aspects and that may improve trust (Dietvorst, Simmons, & Massey, 2018). This may be particularly important if the user considers the learner model to be inaccurate (Yu et al., 2017). At the same time, such systems need to be evaluated to assess their impact in practice (Poursabzi-Sangdeh, Goldstein, Hofman, Vaughan & Wallach, 2018).

Conclusions

We have drawn on several pieces of work that have explored parts of the whole-system puzzle we need to put together to create scrutable learner models. We have defined the questions that a scrutable system should enable a learner to answer about their learner model. We have presented case studies of OLM interfaces that enable learners to gain answers. We have extracted basic recommendations for ways that an authoring system in GIFT can support scrutability at a modest additional cost to the author, but with multiple benefits based on both giving learners understanding of the use of their data in an ITS and mechanisms for the authors to share their design rationale with the learner.

References

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In Proceedings of the 2018 CHI conference on human factors in computing systems (pp. 1-18).

Al-Shanfari, L., Baber, C., & Epp, C. D. (2017). Student Preferences for Visualising Uncertainty in Open Learner Models. In International Conference on Artificial Intelligence in Education (pp. 445-449). Springer, Cham.

Al-Shanfari, L., Epp, C. D., & Baber, C. (2017). Evaluating the effect of uncertainty visualisation in open learner models on students' metacognitive skills. In International Conference on Artificial Intelligence in Education (pp. 15-27). Springer, Cham.

Al-Shanfari, L., Epp, C. D., Baber, C., & Nazir, M. (2020). Visualising alignment to support students' judgment of confidence in open learner models. *User Modeling and User-Adapted Interaction, 30*(1), 159-194.

Al-Shanfari, L., Epp, C. D., & Bull, S. (2016). Uncertainty in Open Learner Models: Visualising Inconsistencies in the Underlying Data. In LAL@ LAK (pp. 23-30).

Apted, T., Kay, J., & Lum, A. (2004). Supporting metadata creation with an ontology built from an extensible dictionary. In International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (pp. 4-13). Springer, Berlin, Heidelberg.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. and Chatila, R., (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82-115.

Assad, M., Carmichael, D. J., Kay, J., & Kummerfeld, B. (2007). PersonisAD: Distributed, active, scrutable model framework for context-aware services. In International Conference on Pervasive Computing (pp. 55-72). Springer, Berlin, Heidelberg.

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. (2016). Negotiated learner modelling to maintain today's learner models. *Research and Practice in Technology Enhanced Learning*, *11*(1), 10.

Bull, S. (2020). There Are Open Learner Models About! IEEE Transactions on Learning Technologies.

Bull, S., & Kay, J. (2013). Open learner models as drivers for metacognitive processes. In International handbook of metacognition and learning technologies (pp. 349-365). Springer, New York, NY.

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., & Mabbott, A. (2006). 20000 inspections of a domain-independent open learner model with individual and comparison views. In International conference on intelligent tutoring systems (pp. 422-432). Springer, Berlin, Heidelberg.

Bunt, A., McGrenere, J. & Conati, C. (2007). Understanding the Utility of Rationale in a Mixed-Initiative System for GUI Customization. User Modeling 2007: 147-156.

Conati C. Barral O., Putnam V. & Riegel L. (2020). Toward Personalized XAI: A Case Study in Intelligent Tutoring Systems.

Cook, R., & Kay, J. (1994). The justified user model: a viewable, explained user model. In Proceedings of the Fourth International Conference on User Modeling.

Cook, R., Kay, J., & Kummerfeld, B. (2015). MOOClm: user modelling for MOOCs. In International conference on user modeling, adaptation, and personalization (pp. 80-91). Springer, Cham.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155-1170.

Eiband, M., Buschek, D., & Hussmann, H. (2020). How to Support Users in Understanding Intelligent Systems? Structuring the Discussion. arXiv preprint arXiv:2001.08301.

European Union (2020) General Data Protection Regulation GDPR, visited June 2020.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, *43*(5-6), 907-928.

Guerra-Hollstein, J., Barria-Pineda, J., Schunn, C. D., Bull, S., & Brusilovsky, P. (2017). Fine-grained open learner models: Complexity versus support. In Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization (pp. 41-49).

Hooshyar, D., Pedaste, M., Saks, K., Leijen, Ä., Bardone, E., & Wang, M. (2020). Open learner models in supporting self-regulated learning in higher education: A systematic literature review. *Computers & Education*, 103878.

Kay, J. (1994). The um toolkit for cooperative user modelling. *User Modeling and User-Adapted Interaction*, *4*(3), 149-196.

Kay, J. (1998). A scrutable user modelling shell for user-adapted interaction (Doctoral dissertation, Basser Department of Computer Science, Faculty of Science, University of Sydney).

Kay, J., & Kummerfeld, B. (2013). Creating personalized systems that people can scrutinize and control: Drivers, principles and experience. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(4), 1-42.

Kay, J., Kummerfeld, B., & Lauder, P. (2002). Personis: a server for user models. In International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (pp. 203-212). Springer, Berlin, Heidelberg.

Kay, J., Li, L., & Fekete, A. (2007). Learner reflection in student self-assessment. In Proceedings of the ninth Australasian conference on Computing education-Volume 66 (pp. 89-95).

Kay, J., & Lum, A. (2005). Exploiting Readily Available Web Data for Scrutable Student Models. In AIED (pp. 338-345).

Kernighan, B. W., & Pike, R. (1999). The practice of programming. Addison-Wesley Professional.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757-798.

Martinez-Maldonado, R., Clayphan, A., Yacef, K., & Kay, J. (2014). MTFeedback: providing notifications to enhance teacher awareness of small group work in the classroom. *IEEE Transactions on Learning Technologies, 8*(2), 187-200.

Martinez-Maldonado, R., Dimitriadis, Y., Kay, J., Yacef, K., & Edbauer, M. T. (2012). Orchestrating a multi-tabletop classroom: from activity design to enactment and reflection. In Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces (pp. 119-128).

Martinez-Maldonado, R., Kay, J., & Yacef, K. (2013). An automatic approach for mining patterns of collaboration around an interactive tabletop. In International Conference on Artificial Intelligence in Education (pp. 101-110). Springer, Berlin, Heidelberg.

Millecamp, M., Htun, N. N., Conati, C., & Verbert, K. (2019, March). To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In Proceedings of the 24th International Conference on Intelligent User Interfaces (pp. 397-407).

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1-38.

Niu, W. T., & Kay, J. (2010). PERSONAF: framework for personalised ontological reasoning in pervasive computing. *User Modeling and User-Adapted Interaction, 20*(1), 1-40.

Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Vaughan, J. W., & Wallach, H. (2018). Manipulating and measuring model interpretability. arXiv preprint arXiv:1802.07810.

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. In Proceedings of the 2018 CHI conference on human factors in computing systems (pp. 1-13).

Upton, K., & Kay, J. (2009). Narcissus: group and individual models to support small group work. In International Conference on User Modeling, Adaptation, and Personalization (pp. 54-65). Springer, Berlin, Heidelberg.

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*(3), 305-338.

Yu, K., Berkovsky, S., Taib, R., Conway, D., Zhou, J., & Chen, F. (2017). User trust dynamics: An investigation driven by differences in system performance. In Proceedings of the 22nd International Conference on Intelligent User Interfaces (pp. 307-317).

Zapata-Rivera, D., Graesser, A., Kay, J., Hu, X., & Ososky, S. J. (2020). Visualization implications for the validity of its. in design recommendations for intelligent tutoring systems. in Design recommendations for intelligent tutoring systems: volume 8 – Data Visualization. US Army Combat Capabilities Development Command Soldier Center.

Zapata-Rivera, D., Hansen, E., Shute, V. J., Underwood, J. S., & Bauer, M. (2007). Evidence-based approach to interacting with open student models. *International Journal of Artificial Intelligence in Education, 17*(3), 273-303.

Chapter 4 – Walk Me Through This: Utilizing Kinesthetic Effect in Data Storytelling

Lane T. Harrison[1] and Andrew J. Hampton[2]
Worcester Polytechnic Institute[1], University of Memphis[2]

Introduction

Visualizing complex data requires consideration of the processing abilities and behavioral patterns of the intended audience. Though this may present as a constraint, accommodating the reality of human strengths and limitations in perception, by the ordered presentation of information, has led to advancements in graphical storytelling. Properly constructed, consumers of data visualization proceed through meaningful phases of comprehension in route to full absorption—like chapters in a book. The interaction of an engaged viewer with visualizations that anticipate cognitive tendencies, constraints, and affordances can create a kinesthetic effect, where people engage with and learn from data through both cognition and action, embedded within the nominally static medium.

The Generalized Intelligent Framework for Tutoring (GIFT) has demonstrated the feasibility of advancing data visualization beyond static depictions and into dynamic, interactive media. Systems like the 'Sandbox' (Goldberg & Hoffman, 2019) utilize multiple projectors, real-time updates, and three-dimensional maps large enough to walk from point to point. This paradigm shift in representation of actors and variables requires a concurrent expansion of techniques for conveying optimal kinesthetic interaction. While the dynamic model solves conventional problems such as depicting change over time, it creates new forms of potential information overload.

In this chapter, we examine cognitive and visual processing, techniques for organizing information, and strategies for applying these concepts to aid users. In doing so, we aim to highlight ways in which designers can aid their audiences in grasping complex relationships such as relative importance of findings, temporal advancement, or any other progression through a given narrative that a set of data intrinsically contains. Finally, we will make recommendations specific to the GIFT platform to leverage these ideas to more effectively inform and engage users.

Storytelling

Visualization enables people to uncover and communicate insights hidden in data by combining the graphical and interactive capabilities of computing with our powerful visual processing and cognitive systems. In practice, however, creating effective visualizations is often challenging, due to realities such as limitations in peoples' attention or memory, and the difficulties of navigating the large and complex design space of visualization alternatives. Fortunately, decades of research and practice has yielded not only exemplars of excellent visualizations, but also defined forms of visualization that are broadly effective. One such form is storytelling.

In a 2006 TED talk, health statistician Hans Rosling delivered a presentation that many would consider to be a key example demonstrating the value of visualization as a means for data-driven storytelling (Rosling, 2006). Using a relatively simple visualization—a scatterplot of countries showing life expectancy and infant mortality—Rosling vividly illustrated events spanning hundreds of years of human history, with accompanying animations and narration. The positive reception to Rosling's efforts raised questions for researchers focusing on data communication and visualization: how and when are interactive visualizations an effective way to deliver stories and insights about data to audiences?

In the subsequent years, advances in visualization authoring tools coupled with the rise of digital media led to an increase in the number of storytelling visualizations available on the web. These web-based visualizations span a wide range of topics, techniques, and intended audiences, serving as a valuable resource for researchers seeking to explore new perspectives and to innovate at the intersection of storytelling and data visualization.

Too often, forms of presentation rely on sparsely annotated tables of data or open-ended exploratory interfaces that fail to provide a framework within which to process information. Storytelling draws on established, familiar traditions of presenting causally related events in sequence. Through artful articulation, readers interact with content in ways that support retention of facts and overall understanding. Beyond storytelling's presence in traditional arts and the written word, it has established its place as a key communicative form supporting data exploration.

In data visualization, specifically, storytelling has a history beyond current research efforts that seek to define and measure its effectiveness. For decades, full page graphics in newspapers and magazines have guided the reader's eye from thought to thought via expertly crafted graphics and text. With technological advancement, we have seen custom interfaces using animation and other tools of shaping the user experience, communicating insights through personal devices. From these and other trends, storytelling appears to form not only a substantial part of the history of visualization, but also a dominant aspect in its future.

Here, we aim to summarize some of the research perspectives and advances in storytelling and visualization. At a high level, we discuss several studies examining forms and models which add language and structure to the broad design space that intersects storytelling and visualization. Tools for crafting stories with data and visualizations are another focus, given the absence of any widely accepted toolsets in practice and the need to support creators who may not have formal training in visualization engineering. Given the many alternative designs that may be considered during visualization creation, we consider storytelling-related evaluation methodologies that shape our understanding of which visualizations best communicate their stories, and why. Throughout, we bend our framings towards considerations that may relate to Intelligent Tutoring Systems (ITSs) in general and GIFT in particular. As GIFT appears to be moving beyond the desktop, we present some recent work in visualization that illustrates creating experiences in Virtual Reality (VR)/Augmented Reality (AR) and other forms of presentation, including a discussion of emerging challenges in these spaces.

Discussion

Forms and Models of Storytelling Visualizations

Much like written stories, an important consideration for the design of a storytelling visualization is its structure, or the form that ties its events together. Drawing from visualizations available in the media and on the web, Segel and Heer (2010) illustrate and define high-level forms of storytelling visualization. Examples include magazine-style depictions, annotated charts, partitioned posters, flow charts, comic strips, slide shows, and films or animations. Recognizing the interactive capabilities of computers, Segel and Heer (2010) also characterize storytelling visualizations according to how they balance author- versus reader-driven control, which leads to either passive or active experiences on the part of the reader. Interactive slideshows might be primarily author-driven, for example, in the sense that the author controls the sequencing of information and the reader controls when it appears. In contrast, drill-down stories might provide the reader with more control over the sequence of information available to them. One particularly salient form identified in their work is the Martini-glass structure, in which an author guides the reader through a series of key points in the data before handing over full control of the interactive components of the visualization for more personalized exploration.

Forms of storytelling visualizations have also been examined at lower, more granular levels, focusing on techniques and possibilities in the design space. Hullman et al. (2013), for example, have proposed that the individual transitions within a storytelling visualization play a large part in shaping the experience of the reader. Examples of transition types identified in their work include dialogue, or questions asked by the author/visualization to the reader, and transitions that move the reader to alternate realities that are intended to facilitate creative thinking or reflection. Moving beyond the individual experience of a visualization toward considerations for entire audiences, Willett, Heer, Hellerstein, and Agrawala (2011) design and evaluate several methods for integrating reader comments and insights directly within visualizations, achieving a form of social data analysis.

Storytelling Visualization Beyond the Desktop

The primary modalities of the storytelling visualizations and research discussed so far are largely confined to traditional desktop-style computers and related screen sizes. However, the rise of the Internet-of-Things, augmented and virtual reality, more diverse screen sizes, and interaction modalities such as touch and voice have spurred the development of visualization beyond the desktop. While these are still burgeoning areas for visualization practice and research, recent work has aimed to lay some groundwork for navigating this landscape and its future directions. Badam, Srinivasan, Elmqvist, and Stasko (2017), for example, covered a range of modalities and corresponding considerations for immersive visualization, such as challenges related to mid-air gesturing and possibilities for gaze-based input. Besides VR/AR and diverse screen sizes, the concurrent advances in natural-language processing and artificial intelligence have led to innovations in visualization authoring and exploration, where readers can use simple voice commands to create and navigate interactive visualizations (Gao, Dontcheva, Adar, Liu, & Karahalios, 2015; Setlur, Battersby, Tory, Gossweiler, & Chang, 2016). In the future, we expect to see more research that investigates the examples briefly mentioned in this chapter. This future research might include reflective characterizations of different forms of immersive visualization for learning, to evaluation considerations for storytelling visualizations targeting learning beyond the desktop.

Recommendations and Future Research

Parallels with GIFT and ITSs

While this chapter has covered only a sampling of work in the visualization community, the examples are intended to align with concerns shared by the GIFT community specifically and the broader ITS community in general. For example, work from Rowe, Shores, Mott, and Lester (2011) has already explored the use of storytelling and narratives in learner-focused systems. Similarly, both GIFT (e.g., Sandbox) and related initiatives (e.g. the tablet-based activities from Bull & Kay, 2016) appear to already be moving beyond the desktop and into areas with multiple novel displays and accompanying challenges.

Aside from existing parallels, however, we offer a few prospective directions that might emerge from an intersection of storytelling-focused visualization research and the GIFT and ITS communities.

One possibility would be to combine visualization and storytelling specific evaluation methodologies in the creation of a visual learner-focused dashboard, as successful approaches combining these areas will likely require considerations of both learner- and visualization-focused evaluation criteria. For example, one could imagine evaluating a visualization-enabled learning focused dashboard both in terms of how it supports domain (i.e. teaching related) tasks, and in terms of the effectiveness of the visualizations used. Consider a visualization showing learner progress as measured by some quantitative variable: a domain-focused evaluation might center on how a given visualization leads to insights about learner progress, while a visualization-focused evaluation might compare alternative encodings such as bar charts or color-scale encodings to determine which leads to more accurate judgments about the underlying learner state.

Tools for crafting visual narratives in learner-focused contexts is another direction which may align with goals held by both communities. For example, one might envision a highly interactive visualization-authoring tool like Lyra (Satyanarayan & Heer, 2014), but focused on crafting visual learning experiences. Lyra uses a formal underlying grammar called Vega to construct a range of interactive visualizations—a model which could plausibly be extended to construct learning experiences and possibly even learning assessments. One key advantage of using an underlying formalism for visualization, that may also be useful in learning environments, is that a formal specification can be used to "transport" the visualization across platforms and device types. Specifics aside, there are clearly overlapping goals and challenges between storytelling, visualization, and learning systems, which members of each community should consider when planning for the future.

Immersive Representation and Tactile Interaction

As always, the primary task of actors (in this case learners) interacting with data visualizations remains the identification and integration of critical variables into meaningful (or actionable) conceptions of the domain (as through storytelling). In multi-model dynamic representations this integration becomes potentially orders of magnitude more difficult as the actor must incorporate not only visual, but auditory and temporal factors. The physical magnitude of systems like the Sandbox creates concern for fundamental issues such as human processing capacity. And though the system allows for minute examination by physically moving toward articles of interest, this creates literal blind spots as the actor's orientation changes, as pieces of information can be obscured by other pieces of information or other elements in the environment.

Tactile representation represents one method of mitigating these concerns. Tactile cues could provide redundancy or emphasis to critical aspects of the display. Designers could deliver these cues through peripheral devices already in place, given that they are typically necessary for control of expansive interactive displays. Directional rumbles using haptic techniques could alert actors of critical developments in real time, as these control "wands" (or similar) typically have built-in orientation calibration relative to the display. Utilizing a separate modality as a redundancy should demand comparably little cognitive resources relative to in-kind alarms that may compete in modalities (visual and auditory) already near capacity (Garcia et al., 2012).

Future research into this augmentation should evaluate its potential along several dimensions. Prominent among these is the integration of new perceptual channels into cohesive data storytelling that now goes beyond "visualization".

References

Badam, S. K., Srinivasan, A., Elmqvist, N., & Stasko, J. (2017). Affordances of input modalities for visual data exploration in immersive environments. In *2nd Workshop on Immersive Analytics,* Phoenix, AZ.

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.

Gao, T., Dontcheva, M., Adar, E., Liu, Z., & Karahalios, K. G. (2015). Datatone: Managing ambiguity in natural language interfaces for data visualization. In C. Latulipe & B. Hartmann (Eds.) *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (pp. 489–500). Charlotte, NC. Association for Computing Machinery.

Garcia, A., Finomore, V., Burnett, G., Calvo, A., Baldwin, C., & Brill, C. (2012). Evaluation of multimodal displays for way-point navigation. In *2012 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support* (pp. 134–137). IEEE.

Goldberg, B., & Hoffman, M. (2019, May). *Intelligent exercise control: Integrating GIFT and battle space visualization*. GIFTSym7, Orlando, FL.

Hullman, J., Drucker, S., Riche, N. H., Lee, B., Fisher, D., & Adar, E. (2013). A deeper understanding of sequence in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2406–2415.

Rosling, H. (February, 2006). TED Talk: Hans Rosling shows the best stats you've ever seen [Video file]. Retrieved from https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education*, *21*(1-2), 115–133.

Satyanarayan, A., & Heer, J. (2014). Lyra: An interactive visualization design environment. In *Computer Graphics Forum* (Vol. 33, 3, pp. 351–360). Wiley Online Library.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics, 16*(6), 1139–1148.

Setlur, V., Battersby, S. E., Tory, M., Gossweiler, R., & Chang, A. X. (2016). Eviza: A natural language interface for visual analysis. In J. Rekimoto, & T. Igarashi (Eds.) *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 365–377). Tokyo, Japan. Association for Computing Machinery.

Willett, W., Heer, J., Hellerstein, J., & Agrawala, M. (2011). Commentspace: Structured support for collaborative visual analysis. In D. Tan (Ed.) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3131–3140). Vancouver, B.C., Canada. Association for Computing Machinery.

Chapter 5 – First Measure Everything: Engineering Trends in Data Visualization

Darian J. DeFalco[1] and Jeanine A. DeFalco[2,3]
[1] Transfix, Inc.
[2] U.S. Army Combat Capabilities Development Command (DEVCOM) – Soldier Center –
Simulation and Training Technology Center (STTC)
[3]Oak Ridge Institute for Science and Education (ORISE) at United States Military Academy

Introduction

Adaptive learning systems produce volumes of detailed data on learning processes, but much of this collected data is simply discarded. The tools and techniques needed to mine all of the potential value out of all data collected have historically been too costly. By combining typically neglected, real-time data with modern data science tools and techniques from the private sector, focused, purpose-built dashboards can be composed using informed decisions relevant to interested stakeholders. Having detailed custom-access to such data in real-time enables the opportunity to make timely decisions based on that data, for students, instructors, course designers, and decision-makers.

The opportunity to align current engineering trends in private industry opens the door for performant, reliable, sophisticated platforms -- without having to invent new platforms, processes, or techniques. The versatility and breadth of these patterns lend themselves to making visualized data more than a tool to reflect on results after the fact, but rather a tool to guide collective decision making by providing shared, tangible, understandable insights into complex data in real time.

Modern Data Platforms

For hundreds of years, learning systems were built on paper-driven processes, resulting in what a modern data scientist today might consider to be extremely high latency data pipelines — such as weekly, monthly, or daily updates on reports on only limited amounts of measured data. The time needed to aggregate sufficient data for meaningful analysis was then complemented by additional high-latency sets of practices to analyze the data, distill it into relevant sets of patterns, and identify ways to effectively communicate to relevant stakeholders interested in the data (e.g., students, instructors, curricular personnel, policy makers, etc.). With the advent of modern tools and techniques, competencies grew around the visual presentation of quantifiable data.

Much like architecture, this field grew into a practical fusion of math and art, with goals reflecting the medium of its day. Given the high latency transmission of data and the labor involved in analyzing and rendering it, maximizing the “data-density” of a given take-away in a report makes perfect sense. More and more ornate visualizations were made possible, enabling greater volumes of data to be distilled into messages consumed by a glance. Edward Tufte’s classic volume, *The Visual Display of Quantitative Information*, staked out a world view where he felt “the task of the designer is to give visual access to the subtle and the difficult — that is, the revelation of the complex,” (Tufte, 2001, p. 191).

However, the past ten years has witnessed an even greater explosion in computational data for a host of reasons including more detailed instruments, new sources of data, faster networks, and more affordable yet performant computation and data storage. This has enabled near real-time aggregation, rendering, and distribution of massive piles of data, creating an increased amount of opportunities for informed decisions to be made at the highest levels, in ways never before possible. These dynamics have impacted many unrelated traditional disciplines--including life science, energy, financial markets, and market research--as well as new roles in emerging industries, e.g., social media and internet search. The effects of these evolving dynamics can be seen in how these disparate domains acquire, store, manage, and communicate their information.

Open Source and Data Science

The technical requirements to build and manage platforms gave rise to new open source infrastructure tools and protocols designed to facilitate the increasing data needs of real-time systems at scale. Over time, this has in turn resulted in increasingly robust, open source data tools and protocols, including the evolution of techniques and strategies to harness, manage, and mine this data.

The emergence and popularity of disciplines such as Data Science, Dev Ops, and Site Reliability Engineering, have increasingly gravitated to these modern open source toolsets, and are themselves evolutions of disciplines which preceded them, yet they share a common goal of adapting to new ways of working with data and its infrastructure at scale.

Whereas in the past, traditional commercial vendors (such as IBM or Microsoft) offered commercial tools at additional cost to visualize their data, the scale and funding model of many of these endeavors prohibited such expenses. However, more recently open source efforts, along with cross-industry broad adoption, resulted in very usable toolsets that are highly performant, reliable, and adaptable (Davis & Daniels, 2016). Open source tools exist for every step of the information pipeline, enabling both dynamic dashboards targeting easy consumption for specific audiences, as well as functioning effectively as playgrounds for data-scientists to play with the data, through which they can identify meaningful patterns that can be shared and analyzed.

Even though free databases have existed for many years, the open source community has produced an increasing number of unique and specialized datastores reflecting the needs of data scientists; these include graph databases, document databases, and timeseries databases. These specialized datastores deviate from more conventional Structured Query Language (SQL) relational databases in order to optimize storage and performance of specific types of data.

In addition to open source code, open source projects have often given way to *de facto* standards which emerge due to community adoption. Of specific interest to data collection of quantifiable data within a time-series at scale is StatsD. Originally its own open source project developed at Etsy, its simple design is built on top of pushing metrics using the existing http standard and has made it a popular protocol to adopt for sending time-series data (Pomel, 2013).

In cases where code cannot be extended, agents that handle StatsD can be added, such as *telegraf* which can be used to pull data from code running locally, which then is pushed to a timeseries database (Netsil, 2017). Regardless of whether the code is augmented to push metrics itself or if an agent is used to facilitate this, pushing the data (as it is created) to a central aggregation point is a model that scales easily. Understanding the affordances and limitations of data extraction is key when considering how to implement and design effective data visualization in complex adaptive instructional systems, for example in the Generalized Intelligent Framework for Tutoring (GIFT) (Sinatra, 2019).

Extending GIFT

GIFT provides interfaces to access data collected by sensors, event logs, and learner state, via its monitor module and the event report tool (Sinatra, 2019). The monitor module is in real-time, whereas the event report tool can produce data exports to be used in other tools. While the monitor serves a valuable role in active monitoring of training sessions, it does not provide any means to compose a customized dashboard of the data for relevant stakeholders. Similarly, the event report tool is limited insofar as data needs to be manually exported — not streamed to an external data source.

Within that context, then, when considering how to optimize data visualization functionality in GIFT, we suggest the first step includes pushing all recorded data to a timeseries database. A timeseries database is a specialized database optimized for the creation and recall of a high volume of time-recorded data. This would enable use of the same tools widely deployed in private industry, thereby capitalizing on investments that have already been made available for common data-visualization needs.

By relegating this data to on-demand exporting or direct monitoring of experiments in progress, the data set becomes what is referred to colloquially as "write-once - read maybe." This means the data is captured and recorded, but not necessarily recalled for analysis. Any analysis is by definition already stale because the data has been exported. Identifying changes as they occur deep within the data is less likely to be identified simply because it is only accessed on demand.

By aggregating metrics in real time to a timeseries database, analyses can be performed against multiple GIFT instances, and multiple training sessions, as they happen or after the fact. Current performances can be compared with past performances using visual cues such as dashed lines to distinguish between current and past events.

In short, opening access to the data stored in currently existent silos creates the potential for both deeper understanding of patterns hidden within the data and translating that data into clearly and more immediately actionable messages. Specifically, by enabling the pushing of timeseries data, this data can be aggregated in large timeseries databases that are optimized for handling large volumes of measurements over time.

Finding Appropriate Signals

While the collection of data is essential, its presentation has a singular purpose: to promote decisive, informed action. Each visualization should be held to the same critical challenge: what does this show and what can be done with it? By restricting the scope of data (the way it is aggregated within groupings, or the relationships it shares within dimensions of time) this becomes a very effective way of helping decision makers accelerate informed and effective courses of action, rather than baffle them with clusters of noise.

For a dashboard to be effective, its consumers' needs must be clearly identified. Let's explore a scenario where we have identified our targets as policy-makers or instructors responsible for a team of learners. The individual learner or team of learners, instructional staff, team leaders, and others in the hierarchical leadership chain, will all have their own discrete needs for data visualization, all of which can potentially be fed by the same data. In addition, learners will have dashboards reflecting how they are performing relative to their peers across all teams, as well as within their own specific team; this helps give context to both individual and team success.

In this scenario, instructional staff may have dashboards emphasizing the 95th percentile and 99th percentile performance, to identify outliers both in terms of individual learners and teams of learners. Policy makers and instructors will likely want to identify the delta between median and 99th percentile, to identify when teams are performing close to parity. Similarly, policy makers and instructors will want to identify 95th percentile and 99th percentile outliers over time, so as to reflect on a learner's ability to adapt the lessons provided by performance data.

Essentially, it is important to remember that different stakeholders will be looking for different signals to boost their actions. Relative rankings to one's peers over time can inform an individual learner where he or she needs to focus their efforts relative to their team's performance. In parallel, dashboards designed for instructional staff can be composed to readily identify training scenarios in need of revision. This can be accomplished by aggregating data by sizes of teams, reporting means of learning frustration at 95th percentile and 99th percentile for all groups, and identifying changes over time.

Suggested Tools

Just as the content and composition of visualized data needs to reflect the different stakeholders, it follows that this is facilitated by having an appropriate range of tools to serve those needs. Specifically, there is a need for a strong dashboarding platform to present on-going operational data as reflected by signals with high confidence that have been appropriately validated. There is also a need for robust backends, appropriately sized, to sustain not just the volume of data needed to be ingested, but also complimentary loads for vending that data. The maturing of new open source toolsets has provided a solid foundation for much of this growth, which in turn gives way to complementary toolsets and methodologies, enabling smaller and smaller teams to effectively manage and monitor an ever-increasing scale of data and resources.

For example, the emergence of reliable, open-source, timeseries databases, such as InfluxDB and TimescaleDB, have made them indispensable tools upon which to build metric-driven workflows (Assay, 2019). The value in these specialized data-stores is in their singular focus around ingesting and vending massive amounts of tagged measurements at specific points in time. The optimization and specificity to common issues related to measurement in general — as well as their open-source licenses — have made them indispensable solutions for a wide range of needs requiring measurement at scale. IBM, who is cited as one of their users, sells its own proprietary database, db2 — a reflection of the special focus of InfluxDB (InfluxData, n.d.).

On top of these timeseries databases, tools like Grafana allow the creation of elaborate dashboards, reflecting the contents of these time-series databases in real-time (Grafana, 2020). One can interact with Grafana's dashboard to focus in on specific data. In addition to being very easy to use to create complex dashboards, Grafana also includes tools to encourage sharing and group editing. For an indication of its industry adoption, Grafana's user list includes Intel, Ebay, Staples, and Fermilab. Their Github code repository reflects 190 releases with 1,104 contributors to its codebase (Github, 2020).

Importantly, Grafana allows a user to compose a dashboard, drawing from an expandable set of premade visual elements, such as heatmaps, histograms, and pie charts. These graphs are visually connected to timeseries databases — such as InfluxDB — filtered with queries. These dashboards can then be saved and shared in a list of dashboards or by sharing a URL. All edits automatically maintain a history of edits, so changes can be reverted (Grafana, n.d.).

Additionally, it is prudent to provide tools for data-exploration. This creative exploration is vital to enabling data-scientists the opportunity to unlock more accurate signals on a macro scale, as well as identifying trends hidden within the data. Tools such as Jupyter and Streamlit have been embraced by many data-science teams as ways to democratize data-exploration, making it easier to present profound insights to key research stakeholders, who often have both the urgent need for timely insights and the lack of time to mine such insights themselves.

Jupyter provides data scientists with web pages which they can create and share among colleagues, where they use data science languages (e.g., Python, Julia, and R) to write code that can visualize remote datasets interactively. Because a Jupyter "notepad" can be composed of many different sections, it is a very convenient way to experiment with data to be shared among a working group (Jupyter, 2020).

The Jupyter notebook is accessed via a web browser. There are controls at the top to control execution for the code below. There are two snippets of Julia code, followed by the visualizations that the code produces. This format makes it very easy for researchers to explore the same data and to share it with stakeholders who are not data scientists. Since the data is able to draw from external data-sources, including SQL and InfluxDB, Jupyter notebooks can capture real-time data, making this a robust platform to validate and test hypotheses against data (Jupyter, 2020).

Streamlit is an example of a data science application framework that makes it very easy to create interactive web pages enabling end users to explore visualizations of data using Python (Streamlit, 2020). Since Streamlit is based in Python, it works well with other common Python data science frameworks such as Pandas and Numpy. It also means that it can access common SQL databases as well as timeseries databases, such as InfluxDB. Streamlit makes it very easy to enable data-scientists to compose polished user interfaces enabling sophisticated visualization and interaction with the data.

Being able to directly explore and experiment with data stored in the same sources used for dashboards means these are perfect platforms to backtest dashboard changes, validate hypotheses that are driving current dashboards, and formulate queries that can then be translated into dashboards using a tool, such as Grafana, to be shared at scale. If a dashboard establishes a scale of an axis based upon a baseline determined with a subset of data, such as localized performance versus generalized performance across training sites, more appropriate baselines can be identified and applied by analyzing this same data in parallel.

In sum, the hidden risk in relying on dashboards at scale is the ease by which an organization can get locked into tunnel vision, with a set of dashboards built on top of assumptions that may no longer apply. While the virtues of a dashboarding platform like Grafana lie in their ability to continuously visualize data filtered through meaningful lenses in real-time at scale, it needs to be understood that the wealth of data must be challenged and validated in order to realize its value. Such analysis is how organizations learn about their failures and capitalize on their hidden successes by better understanding what made them successful.

Conclusion

Modern, open-source data platforms include new tools to solve problems related to large volumes of quantifiable data at scale. The near real time insights they enable, were once the domain of highly specialized applications, such as stock exchanges. By extending GIFT to aggregate data already being collected, these platforms present an opportunity to capitalize on the promise of adaptive instructional systems by ensuring visibility into the present at scale, as well as a provide a contextualized view of current outcomes. Following best practices in industry, our suggestion is to focus on pushing data into a timeseries database, which can then be used to drive dashboards and extend the abilities for real-time data-science to test and validate hypotheses. This in turn provides an opportunity to capture historically overlooked data that can inform policy-makers and instructors on how to make efficient and informed decisions in shaping the delivery and flow of adaptive learning to address learning objectives.

References

Assay, M. (2019). *Why time series databases are exploding in popularity.* Retrieved from https://web.archive.org/web/20190626143018/https://www.techrepublic.com/article/why-time-series-databases-are-exploding-in-popularity/

Davis, J. &, Daniels, R. (2016). *Effective Dev Ops*, O'Reilly Media.

Grafana, (ND). *Grafana/Grafana.* Retrieved from https://github.com/grafana/grafana.

Grafana, (2020). *Grafana Play Home*. Retrieved from https://play.grafana.org/d/000000012/grafana-play-home?orgId=1

InfluxData. (ND). Retrieved from https://www.influxdata.com

Jupyter. (2020). *Jupyter Binder Examples: Demo Julia*. Retrieved from https://jupyter.org/try

Netsil (2017). *Collector Comparison: Telegraf vs Collectd vs DD-agent*. Retrieved from https://blog.netsil.com/collector-comparison-telegraf-vs-collectd-vs-dd-agent-f49c866657b0

Pomel, O. (2013) *StatsD, what it is and how it can help you*. https://www.datadoghq.com/blog/statsd/#what-problem-does-statsd-solve

Sinatra, A. M. (2019). The 2019 Instructor's Guide to GIFT. In *Proceedings of the 7th Annual GIFT Users Symposium* (p. 19). US Army Combat Capabilities Development Command–Soldier Center. Retrieved at https://gifttutoring.org/documents/149

Streamlit. (2020). *demo-uber-nyc-pickups*. Retrieved from https://github.com/streamlit/demo-uber-nyc-pickups

Tufte, E. (2001) *The Visual Display of Quantitative Information*. Graphics Press LLC.

Chapter 6 – Visualizing Team Processes using Epistemic Network Analyis: Affordances for Researchers, Educators, and Teams

Zachari Swiecki[1], Morten Misfeldt[2], Xiangen Hu[3], and David Williamson Shaffer[1,4]
University of Wisconsin–Madison[1], University of Copenhagen[2], University of Memphis[3],
Aalborg University Copenhagen[4]

Introduction

Complex problems in domains such as engineering, medicine, and the military often require the coordinated efforts of multiple individuals. As a result, teamwork and collaborative problem solving have become critical 21st century skills to study, teach, and assess (Hesse, Care, Buder, Sassenberg, & Griffin, 2015). With an emphasis on the military domain, a key goal of the Generalized Intelligent Framework for Tutoring (GIFT) project is to scaffold the development of tutoring scenarios that support team-training and assessment (Sinatra, 2018). Central to this endeavor is the ability to model team processes (Ruis, Hampton, Goldberg, & Shaffer, 2018), and in this chapter, we argue that modeling *connectivity*—or the ways in which team processes relate to and build upon one another—is critical to doing so.

As we describe more below, developing effective team-training and assessment will require models of connectivity that are sensitive to the goals of multiple audiences. Specifically, they should afford (a) researchers the ability to visualize, compare, and make predictions, (b) educators the ability to monitor team processes, intervene, and make assessments, and (c) teams the ability to maintain awareness. In this chapter, we describe Epistemic Network Analysis (ENA) (Shaffer, Collier, & Ruis, 2016) as one approach to modeling and visualizing connectivity, and provide examples of the affordances of the approach for researchers, educators, and teams.

Modeling Connections

When teams solve problems, their processes include actions toward accomplishing a task and actions toward managing the processes of collaboration, such as communicating information and coordinating behavior (Marks, Mathieu, & Zaccaro, 2001). In turn, team processes are not simply the sum of individual actions; rather, individual actions interact with one another. An important consequence of this interaction is that the actions of individuals on teams occur in relation to and are thus influenced by one another. In other words, teams are interdependent systems. Interdependence suggests that an important feature of team processes is the idea of *connections*, or the relationships that exist between individuals and their actions as they work together, not simply whether particular actions occur in isolation (Csanadi, Eagan, Kollar, Shaffer, & Fischer, 2018).

Network models are powerful tools for modeling the connections that exist between individuals as they work together. For example, Social Network Analysis (SNA) has been widely used to understand how teams interact as they solve problems (Fincham, Gašević, & Pardo, 2018). While SNA is a powerful technique, it is limited due to its focus on the structure of team interactions, such as who communicated with whom. However, team interactions are always *about* something, meaning that they have content as well as structure. One network technique for understanding the content of team interactions is ENA. ENA models team activity by identifying categories of action, communication, cognition, and other relevant features and characterizing them using coding schemes into smaller sets of domain-relevant nodes. The weights of the connections among network nodes (i.e., the connection structure of key elements in the domain) are then computed and visualized. Critically, ENA models team interactions in such a way that it is possible to extract information about each team member's contributions to team performance while accounting for the interdependent nature of team processes (Siebert-Evenstone et al., 2017)

ENA has been successfully used to study teams in domains such as engineering education and medicine (Arastoopour, Shaffer, Swiecki, Ruis, & Chesler, 2016; Sullivan et al., 2018). Moreover, prior work using ENA to model the performance of military teams has shown that it has statistical and interpretive advantages compared to models that use coding-and-counting (Swiecki, Ruis, Farrell, & Shaffer, 2020) and sequential pattern mining (Swiecki, Lian, Ruis, & Shaffer, 2019) approaches to modeling team processes.

Connections for Different Audiences

The arguments above suggest that measuring connectivity is critical to understanding team processes and that ENA is a powerful tool for doing so. However, different audiences need to understand connectivity for different reasons. For example, researchers need to understand which patterns of connections are related to team performance in order to build theory and make predictions. In turn, they need to be able to visualize and compare connections, as well summarize them for use in predictive models.

Educators, on the other hand, may not need the same level of sophistication in models, nor will they necessarily have the technical background or time to meaningfully engage with the models in the ways that researchers do. Instead, they need ways to quickly and easily understand the patterns of connections that teams are making to monitor team performance, deliver pedagogical interventions, and make assessments.

Similarly, teams may lack the need or ability to engage with connectivity models as researchers do. But, research on teams has shown that coordinating activity is essential to good team performance (Salas, Sims, & Burke, 2005). This work suggests that teams need ways of developing awareness of team interactions to better coordinate their efforts. Research on Group Awareness Tools has found that visualizations of cognitive and social information about teams, such as the knowledge that individuals have or the activities they are doing, can promote awareness, help coordinate activity, and improve performance (Bodemer, Janssen, & Schnaubert, 2018). However, few of these tools represent connections (Swiecki & Shaffer, 2018). Given the importance of connections for understanding team processes, such tools could potentially benefit from connectivity models.

In what follows, we use ENA models of teams in a military training context to illustrate how connectivity models can be adapted to the needs of researchers, educators, and teams.

As part of the Tactical Decision Making Under Stress project, sixteen teams participated in training scenarios to test the impact of a decision-support system and teamwork training on team performance in the context of air defense warfare (Swiecki & Shaffer, 2018). During the scenarios, teams needed to detect and identify ships and aircraft via radar, assess whether they were threats, and decide how to respond. Each team consisted of six members who held either a command role, such as the Commanding Officer (CO), or a supporting role, such as the Electronic Warfare Supervisor (EWS). The dataset consists of transcripts of team communications and performance scores for each team.

The ENA algorithm and our ENA analysis of this dataset is described in detail in Swiecki and colleagues (2020). In brief, the analysis entailed segmenting the transcripts by turns of talk. Next, we developed and validated an automated coding scheme that captured critical features of the air defense warfare process, such as seeking information about radar contacts or giving orders to defend the ship from attack. Finally, we analyzed the coded data using ENA to model the connections—operationalized as the co-occurrence between codes in turns of talk—that individuals made as they worked together.

Researchers

To afford researchers the ability to visualize connections, ENA produces *weighted network graphs* in which the nodes correspond to codes and the edges are proportional to the relative frequency of connection between two codes. Each unit of analysis has an associated network, and to compare networks, ENA affords the ability to subtract two networks to show the connections that are stronger for one unit relative to the other.

To afford comparisons between more than two units and their networks, ENA produces statistics, or *ENA scores*, that summarize the connections in their networks via their location in a lower dimensional space. This space is created when the ENA algorithm performs a dimensional reduction of the connection counts of all units in the data. ENA scores can be used to conduct statistical tests between samples of units—for example, high and low performers—and they can be included in subsequent predictive models of team performance.

Figure 1 below highlights these affordances. The figure includes the ENA scores for the commanders on high (green circles) and low (red circles) performing teams determined using a median split on the performance scores. The green and red squares correspond to the mean position in the space of the high and low performers, and the network shows the subtraction between the mean networks of the high and low performers. The subtraction indicates that high performing commanders focused more on tactical actions—as evidenced by stronger connections among codes such as Deterrent Orders, Recommendation, and Track Behavior—while low performers focused more on seeking information to understand the tactical situation—as evidenced by stronger connections between Seeking Information and TrackBehavior, DetectIdentify, and AssessmentPrioritization. A *t*-test between positions of the commanders on the first dimension of the space found that commanders on high and low performing teams were statistically significantly different in terms of the connections they made: $t(26.34) = -3.34$, $p < 0.05$, $d = 1.17$.

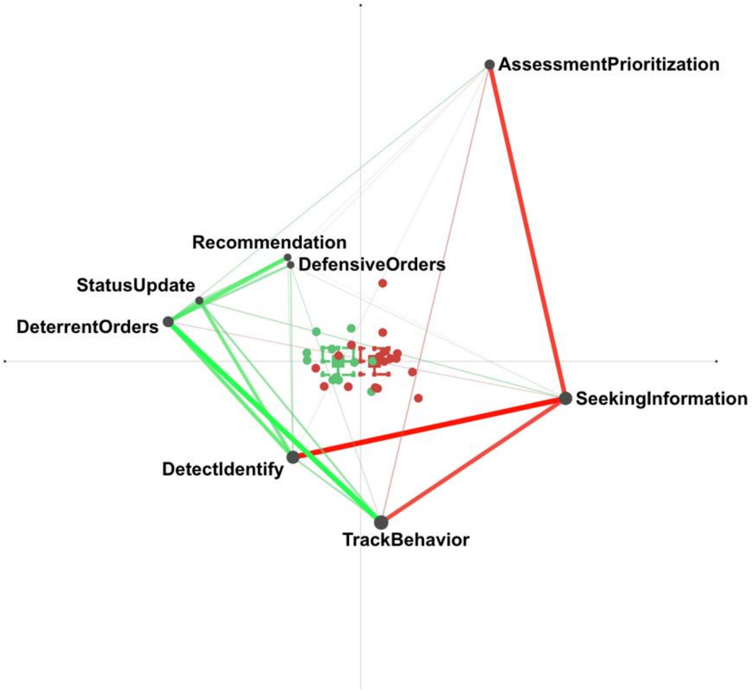


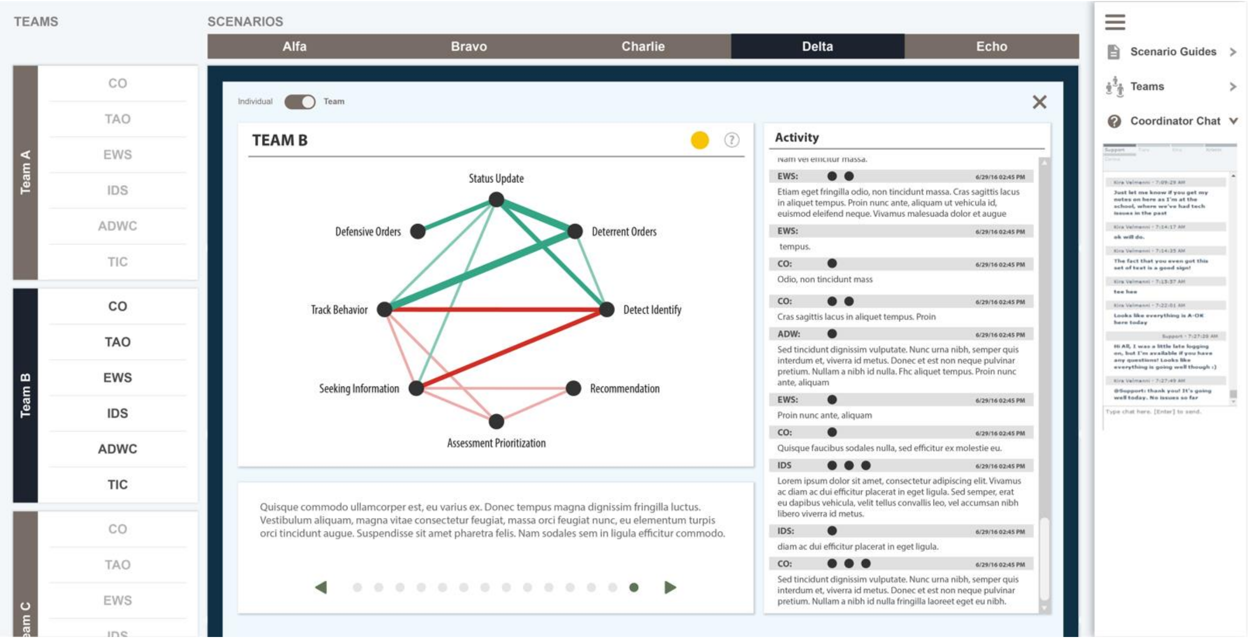
Figure 1. ENA visualization for commanders on high (green) versus low (red) performing teams.

Educators

To afford educators the ability to visualize connections, monitor and assess team processes, and conduct interventions, we developed a technique for simplifying ENA models and integrating them into a real-time dashboard. The design was based on an interface we co-designed and implemented with teachers who used educational simulations in their classrooms (Herder et al., 2018; Misfeldt et al., 2020, this volume).

The dashboard features high-level performance overviews of teams and individuals, as well as simplified network diagram designed to show connections as they are made in real-time. Figure 2 shows an example of this design for ENA models in the air defense warfare context. On the left are identifiers for each team and individual (by role). At the top are identifiers for the training scenario being examined. The middle of the figure shows the simplified network visualization of one team's performance at the end of the training scenario. Connections indicative of high performance are in green and low performance in red. These connections were identified as the variables that most distinguished high and low performing teams/individuals (as indicated by the team performance score described above) in an ENA space. The mixed nature of this network (having both high and low connections present) is reflected by the yellow performance summary icon at the top right. This view also shows the team activity represented in the network model—in this case, the coded team transcript—on the right. Below the network is a narrative description of the connections present in the network (note that placeholder text is used in the design for the activity record and description).

While the dashboard visualizations are designed to update in real-time, educators can also use the arrows below the network to step through the scenario at their own pace and review each connection that occurred. In addition, they can highlight and examine the activity contributing to a connection by clicking the connection in the network model.



Quisque commodo ullamcorper est, eu varius ex. Donec tempus magna dignissim fringilla luctus. Vestibulum aliquam, magna vitae consectetur feugiat, massa orci feugiat nunc, eu elementum turpis orci tincidunt augue. Suspendisse sit amet pharetra felis. Nam sodales sem in ligula efficitur commodo.

Nam vel efficitur massa.

EWS: ● ● 6/29/16 02:45 PM
Etiam eget fringilla odio, non tincidunt massa. Cras sagittis lacus in aliquet tempus. Proin nunc ante, aliquam ut vehicula id, euismod eleifend neque. Vivamus malesuada dolor et augue

EWS: 6/29/16 02:45 PM
tempus.

CO: ● 6/29/16 02:45 PM
Odio, non tincidunt mass

CO: ● ● 6/29/16 02:45 PM
Cras sagittis lacus in aliquet tempus. Proin

ADW: ● 6/29/16 02:45 PM
Sed tincidunt dignissim vulputate. Nunc urna nibh, semper quis interdum et, viverra id metus. Donec et est non neque pulvinar pretium. Nullam a nibh id nulla. Fhc aliquet tempus. Proin nunc ante, aliquam

EWS: ● 6/29/16 02:45 PM
Proin nunc ante, aliquam

CO: ● 6/29/16 02:45 PM
Quisque faucibus sodales nulla, sed efficitur ex molestie eu.

IDS ● ● ● 6/29/16 02:45 PM
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus ac diam ac dui efficitur placerat in eget ligula. Sed semper, erat eu dapibus vehicula, velit tellus convallis leo, vel accumsan nibh libero viverra id metus.

IDS: ● 6/29/16 02:45 PM
diam ac dui efficitur placerat in eget ligula.

CO: ● ● ● 6/29/16 02:45 PM
Sed tincidunt dignissim vulputate. Nunc urna nibh, semper quis interdum et, viverra id metus. Donec et est non neque pulvinar pretium. Nullam a nibh id nulla fringilla laoreet eget eu nibh.

Kira Velmanni · 7:09:29 AM
Just let me know if you get my notes on here as I'm at the school, where we've had tech issues in the past

Kira Velmanni · 7:14:17 AM
ok will do.

Kira Velmanni · 7:14:35 AM
The fact that you even got this set of text is a good sign!

Kira Velmanni · 7:15:57 AM
tee hee

Kira Velmanni · 7:22:01 AM
Looks like everything is A-OK here today

Support · 7:27:29 AM
Hi All, I was a little late logging on, but I'm available if you have any questions! Looks like everything is going well though :)

Kira Velmanni · 7:27:49 AM
@Support: thank you! It's going well today. No issues so far

Type chat here. [Enter] to send.

Figure 2. ENA-based dashboard for educators.

Teams

To afford teams the ability to maintain awareness in real-time, we refined the educator dashboard design for use by teams (Figure 3). In temporally and cognitively demanding tasks such as the air defense warfare scenarios described here, teams likely only have the bandwidth to focus on a small number of representations. As such, we focused this design on the simplified network diagram and made the activity record and network description collapsible. Here, training scenarios are along the left, and individual team member identifiers (by role) are along the top. Each individual has a performance summary icon (reflecting the connections in their network) next to their role name. The design has also been updated to direct attention to the recent connections made via color saturation: more saturated connections occurred more recently, while less saturated connections occurred earlier in the scenario. In this way, teams can quickly see the history of team activity. Team members can also switch between the diagrams of the full team and any given individual on their team to get more specific information. We hypothesized that these features would support team awareness and also provide a representation that teams could use as part of after action reviews of their performance (see Johnston et al., 2020, this volume).

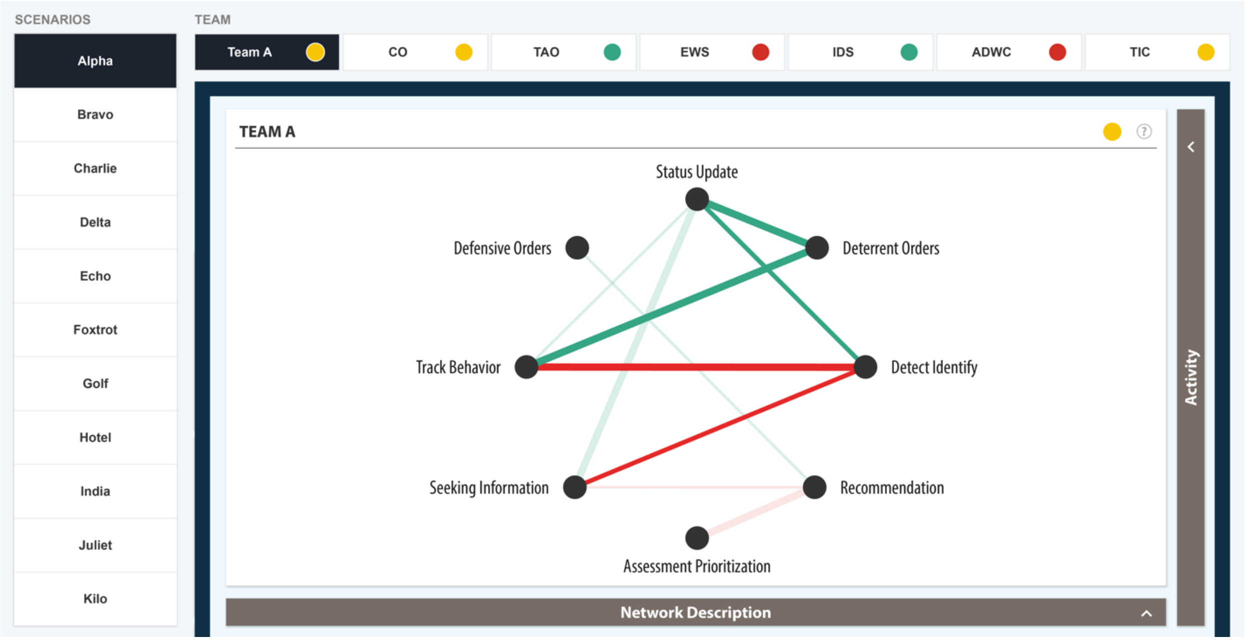


Figure 3. ENA-based dashboard for teams.

Discussion

While the discussion above suggests that ENA is a powerful tool for investigating the connections that characterize team processes, the method is not without limitations. First, the designs presented here have yet to be implemented as part of military training scenarios. While similar designs have shown promising results for educational simulations designed for civilians (Herder et al. 2018), future work is needed to test their utility for educators and teams in this context.

Second, there are of course other more sophisticated techniques that could be used to model and visualize the connections that are critical to team processes. For example, ENA can be combined with SNA to produce models that simultaneously capture both social connections of the team—for example, who communicated with whom and how much—and the cognitive connections of the team (Swiecki & Shaffer, 2020). Future work will explore adapting such techniques for use by educators and teams.

Finally, the examples provided above relied on data from a particular context, and are thus limited by the nature of the data. In other contexts, data sources might come from a variety of modalities, including positional data, gestures, interface clicks, biometrics, and so on, and the team structures may be quite different—for example, containing many more members, or containing sub-teams that interact with one another. However, ENA has been successfully used to study teams in a variety of domains, so its affordances for researchers have been shown to generalize to other contexts. Moreover, ENA is agnostic to the type of data used, provided it is formatted correctly. Future work will need to explore the use of ENA in new contexts with a particular focus on how to represent networks developed from multimodal data to educators and teams.

Conclusions and Recommendations and Future Research

In this chapter, we argued that modeling connectivity is critical to understanding team processes. In addition, we argued that while researchers, educators, and teams can each benefit from models of connectivity, they interact with those models in different ways, and thus need different affordances in each case.

The work presented here suggests two recommendations for GIFT and future Intelligent Tutoring Systems (ITSs) for teams. First, given the importance of modeling connections for understanding team processes, GIFT managed ITSs could benefit from integrating models of team connectivity into their systems. Such a project would require mechanisms for automatically collecting data in a machine readable format, model development in targeted domains, and the development of a version of the model that would integrate with the existing GIFT architecture during the run-time of the system. The integration of connectivity models into GIFT would also allow further exploration of the effectiveness of such models in a larger variety of team contexts, improving our understanding of how educator and team processes may be improved through visualizing connectivity.

Second, this work suggests that any efforts to include models of team connectivity into GIFT managed ITSs should carefully consider the audiences that will interact with the system, as this will impact its design and usability. Given that GIFT managed ITSs are generally used for training purposes, systems that afford educators and teams the ability to monitor connections as they occur and review connections after training could be particularly useful.

Acknowledgements

This work was funded in part by the National Science Foundation (DRL-1661036, DRL-1713110), the U.S. Army Research Laboratory (W911NF-18-2-0039), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

References

Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, *32*(3B), 1492–1501.

Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group awareness tools for computer-supported collaborative learning. In F. Fischer, C. Hmelo-Silver, S. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 351–359). Routledge.

Csanadi, A., Eagan, B., Kollar, I., Shaffer, D. W., & Fischer, F. (2018). When coding-and-counting is not enough: Using epistemic network analysis (ENA) to analyze verbal data in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, *13*(4), 419–438.

Fincham, E., Gašević, D., & Pardo, A. (2018). From Social Ties to Network Processes: Do Tie Definitions Matter? *Journal of Learning Analytics*, *5*(2), 9–28–29–28. https://doi.org/10.18608/jla.2018.52.2

Herder, T., Swiecki, Z., Fougt, S. S., Tamborg, A. L., Allsopp, B. B., Shaffer, D. W., & Misfeldt, M. (2018). Supporting teacher's intervention in student's virtual collaboration using a network based model. *Proceedings of the International Conference on Learning Analytics*, 21–25.

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In *In P. Griffin & E. Care (Eds.), Assessment and teaching of 21st century skills: Methods and approach* (pp. 37–56). Springer.

Johnston, J.H., Sinatra, A.M., & Swiecki, Z. (2020, this volume). Application of Team Training Principles to Visualizations for After-Action Reviews. In Sinatra, A.M., Graesser, A., Hu, X., Goldberg, B., & Hampton, A. (Eds.), *Design Recommendations for Intelligent Tutoring System: Volume 8—Data Visualization*. U.S. Army Combat Capabilities Development Command - Soldier Center.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A Temporally Based Framework and Taxonomy of Team Processes. *The Academy of Management Review*, *26*(3), 356. https://doi.org/10.2307/259182

Misfeldt, M., Swiecki, Z., Zapata-Rivera, D., & Hu, X. (2020, this volume). Pedagogical Use Scenarios for Data Visualizations: Prepare, Conduct and Evaluate. In Sinatra, A.M., Graesser, A., Hu, X., Goldberg, B., & Hampton, A. (Eds.), *Design Recommendations for Intelligent Tutoring System: Volume 8—Data Visualization*. U.S. Army Combat Capabilities Development Command - Soldier Center.

Ruis, A. R., Hampton, A. J., Goldberg, B. S., & Shaffer, D. W. (2018). Modeling processes of enculturation in team training. In R. Sottialre, A. Graesser, & A. M. Sinatra (Eds.), *Design Recommendations for Intelligent Tutoring System: Volume 6—Team Tutoring* (pp. 45–51). U.S. Army Research Laboratory.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in Teamwork? *Small Group Research*, *36*(5), 555–599. https://doi.org/10.1177/1046496405277134

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9–45.

Siebert-Evenstone, A. L., Irgens, G. A., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2017). In search of conversational grain size: Modelling semantic structure using moving stanza windows. *Journal of Learning Analytics*, *4*(3), 123–139.

Sinatra, A. M. (2018). Team Models in the Generalized Intelligent Framework for Tutoring: 2018 Update. In R. Sottialre (Ed.), *Proceedings of the Sixth Annual GIFT Users Symposium* (pp. 157–161).

Sullivan, S. A., Warner-Hillard, C., Eagan, B. R., Thompson, R., Ruis, A. R., Haines, K., Pugh, C. M., Shaffer, D. W., & Jung, H. S. (2018). Using epistemic network analysis to identify targets for educational interventions in trauma team communication. *Surgery*, *163*(4), 938–943.

Swiecki, Z., Lian, Z., Ruis, A. R., & Shaffer, D. W. (2019). Does order matter? Investigating sequential and cotemporal models of collaboration. In K. Lund, G. P. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer-Supported Collaborative Learning (CSCL) 2019* (Vol. 1, pp. 112–120). International Society of the Learning Sciences.

Swiecki, Z., Ruis, A. R., Farrell, C., & Shaffer, D. W. (2020). Assessing individual contributions to collaborative problem solving: A network analysis approach. *Computers in Human Behavior*, *104*.

Swiecki, Z., & Shaffer, D. W. (2020). ISENS: An integrated approach to combining epistemic and social network analyses. *10th International Conference on Learning Analytics and Knowledge (LAK'20)*, 9.

Swiecki, Z., & Shaffer, D. W. (2018). Toward a taxonomy of team performance visualization tools. In J. Kay & R. Luckin (Eds.), *Rethinking Learning in the Digital Age: Making the Learning Sciences Count: Vol. III* (pp. 144–151).

Chapter 7 – Visualization Implications for the Validity of Intelligent Tutoring Systems

Diego Zapata-Rivera[1], Arthur C. Graesser[2], Judy Kay[3], Xiangen Hu[2], and Scott J. Ososky[4]
Educational Testing Service[1], University of Memphis[2], University of Sydney[3], Microsoft[4]

Introduction

Learner Models (LMs) keep information about students' knowledge, skills, abilities, and attitudes. These LMs are used to support the adaptive features of Intelligent Tutoring Systems (ITSs; e.g., adaptive sequencing and adaptive scaffolding features (Graesser, Hu, & Sottilare, 2018; Greer & McCalla, 1994)). LMs are also known as student models, student/learner profiles, or user models (Kay & Kummerfeld, 2019) based on the context of use. LMs are referred as the Learner Module in the Generalized Intelligent Framework for Tutoring (GIFT; gifttutoring.org; Sottilare, Brawner, Sinatra, & Johnston, 2017).

A variety of educational stakeholders (e.g., students, teachers, parents, and researchers) make decisions based on LM information that is made available to them in the form of dashboards, open learner modeling tools, or other reporting systems. The types of decisions they make can be compromised depending on their understanding and use of this information. The validity of the interpretations made by users depends on a clear communication of learner model information with the intended audience.

Visualization principles can be used to design systems that clearly communicate LM information to the intended audience. Many of these principles have been articulated in models developed in the fields of human-computer interaction and human factors that apply to a variety of digital environments designed for different purposes (Card, Moran, & Newell, 1983; Hegarty, 2018; John & Kieras, 2006), including learning technologies (Roscoe, Craig, & Douglas, 2018). Issues such as cognitive bias, accessibility, possible misunderstandings, and misuse of results should be considered when designing and evaluating graphical interfaces used in these interactive reporting systems. Thus, the type of visualization and amount of detail provided are likely to differ for different types of users.

In this chapter we elaborate on validity issues and their implications for the design and evaluation of graphical representations used to share LM information. We review principles of effective visualization, a framework for designing and evaluating reporting systems, and provide recommendations for the use of graphical representations in GIFT.

Validity and Visualization

Assessments are usually designed and used to support human decision making. Assessments can be thought of as composed of assessment arguments (Mislevy, 2012). The quality of these assessment arguments depends on the type of evidence collected and used to support assessment claims. Validity is a property of the proposed interpretations and uses of assessment results. An assessment may be designed to support particular interpretations and uses (Kane, 2013). Katz et al. (2017) describe how validity theory and practice can inform research and development of ITSs.

Supports for clear communication, appropriate interpretation, and use of assessment results by the intended audience contribute to the overall argument that supports test validity (Hambleton & Zenisky, 2013; O'Leary, Hattie, & Griffin, 2017; Tannenbaum, 2018; Zapata-Rivera & Katz, 2014). The potential benefits of well-designed assessments can be compromised if the results are not clearly understood by the intended users or if these results are misused. Reporting systems are usually employed to inform users about the results of an assessment. To create graphical representations that clearly communicate assessment results, it is important that their design is based on the application of empirically-derived principles from areas such as cognitive science, human-computer interaction, human factors, and information visualization.

Tannenbaum (2018) emphasizes the importance of an alignment between the construct of the assessment (what it is intended to measure) and the report. In general, assessment results should be provided to users only for claims for which there is enough evidence to support the intended uses. If there is a lack of such alignment, this could result in users making decisions based on assessment results without enough supporting evidence. Examples of this misalignment occur when making decisions is based on evidence from a few multiple-choice questions or when there is conflicting evidence from various sources. Alignments in assessments are also needed in the design of an ITS that collects process and response data about students' interactions with the system. These data are used by the system to maintain internal representations of the student's cognitive and non-cognitive skills (i.e., LMs). LMs can be seen as a set of claims about student knowledge and skills (e.g., student x mastered concept y). Alignment between the assertions in the LM and data collected to support these assertions is key in the development of valid ITSs.

In addition, once assessment information is available to be presented to the intended audience, it is important to present this information in a way that supports clear communication of the intended message and appropriate use of assessment results. External representations (e.g., graphical, verbal) should be selected to minimize biased interpretations, possible misunderstandings, misuse of results, and support accessibility and comprehension of the assessment results (Zwick, Zapata-Rivera, & Hegarty, 2014).

A framework for designing and evaluating score reporting systems is described in Zapata-Rivera, VanWinkle, and Zwick (2012). This framework includes the following activities: gathering assessment result needs for the intended audience; developing prototypes of the report; evaluating the prototypes with both experts and the intended audience; and iterating these steps in light of the evaluation in a development cycle. Table 1 shows examples of assessment information needs for a variety of users including teachers, students, parents, administrators, and researchers (Zapata-Rivera, 2019). Reporting systems should be designed considering the knowledge, needs, and attitudes of the target audience (Zapata-Rivera & Katz, 2014).

The next section describes principles of effective visualization and related research in the area of Open Learner Modeling, and the use of graphical displays to represent uncertainty. These principles and related research can be used in the design of clear graphical representations to support the use of learner model information by different types of users.

Table 1. Sample assessment information needs for various types of users.

Audience	Assessment information needs
Teachers, tutors, and mentors	• **Student performance at the individual, sub-group and class levels** What are my students' strengths and weaknesses? How did the class perform on a task or a group of tasks? How does a student's performance compare to that of other students? • **Progress information at the individual, subgroup, and class levels** How much progress have my students made towards mastery? • **Information that helps understand current performance** Were my students engaged in the task(s)? Did my students try to game the system? How reliable are the knowledge estimates calculated by the system? • **Information that can help inform future teaching** How difficult were the tasks for my students? What were the most frequent errors and misconceptions? What should I do next to help an individual student or the class as a whole?
Students	• **Actionable feedback that they can use to guide their learning** What are my strengths and weaknesses? How can I improve? • **Progress and performance information** How much progress have I made towards mastery? How does my performance compare to that of other students? • **Evidence supporting assessment claims** What type of information was used to calculate my knowledge levels? Can I provide additional evidence to update my knowledge levels in the system?
Parents	• **Actionable information on how to help the student** What are my child's strengths and weaknesses? How can I help my child? Should I talk to the teacher? • **Progress and performance information** How did my child perform? How much progress has my child made towards mastery? How does my child's performance compare to that of other students?

Administrators and policy makers	• **Aggregate data to inform decisions in areas such as evaluation of current educational policies, school improvement plans, professional development, program selection and evaluation, curriculum selection, improving student achievement, and staff allocation** How does student performance compare to other students at the classroom, school or district levels? How do students from particular subgroups perform? How much progress did students from particular subgroups make in the last month?
Researchers	• **Information to evaluate and improve different aspects of the system** Were students engaged in the task(s)? How effective were adaptive mechanisms such as feedback and task sequencing in helping students learn? Do users understand and appropriately use the information in the reporting system or dashboard? To what extent do teachers and students use the information in the reporting system or dashboard to inform their teaching/learning? Which aspects of the system need to be improved?

Note about Table 1. **Adapted from "Supporting Human Inspection of Adaptive Instructional Systems", by Zapata-Rivera, D., 2019, Lecture Notes in Computer Science, vol. 11597. Springer, Cham. p. 487. Copyright by EducationalTesting Service, 2019 All rights reserved. Reprinted with permission.**

Principles for Effective Visualizations and Related Research

Hegarty (2011, 2018) discusses principles for the design of effective visualizations. These principles are based on work in fields such as information visualization, human-computer interaction, and cognitive science. These principles consider the roles of perception, attention, working memory, and prior knowledge on understanding graphical representations and the affordances of different types of displays to represent data. Thus, there is no such thing as the "best visualization." It depends on the type and purpose of visualization, as well as the type of data and characteristics of the user (e.g., knowledge of graphic conventions, mathematical knowledge, and domain knowledge).

Hegarty's principles of effective visualizations include:

- *Relevance*. No more, no less information than what is needed by the user (Kosslyn, 2006).
- *Capacity*. Taking into account limitations in working memory and attention. This principle is related to the concept of data-ink ratio (Tufte, 2001) and further revisions of this concept by Gillan and Richman (1994). This principle and the relevance principle complement each other since by highlighting important pieces of information and limiting the number of relevant components, it is possible to facilitate users' understanding of the intended message.
- *Apprehension*. A visual display has to be accurately perceived (Cleveland & McGill, 1983; Kosslyn, 2006). For example, it is important to use appropriate visual dimensions to accurately represent different relationships among variables and to avoid representations that lead to biased judgments (Wickens & Hollands, 2000).
- *Perceptual organization.* Display elements should be grouped into units. Placing display elements should facilitate comprehension (Wagemans et al., 2012).

- *Compatibility principle.* Display form should be consistent with its meaning (e.g., meaning that is consistent with cultural interpretations of spatial metaphors) (Kosslyn, 2006).
- *Matching dimensions. S*cales of measurement of visual variables should match with the ones they represent (e.g., shape—a categorical dimension, length—a ratio dimension) (Zhang, 1996).
- *Pragmatics principles* (Kosslyn, 2006).
    - *Salience*. The most important information should be the most salient.
    - *Informative changes*. Changes in a display are expected to carry information/meaning.
- *Usability principles*. Users have the knowledge or are provided with the supports to understand and correctly use the information in the display (e, g., graphic conventions and legends) (e.g., Tufte, 2001).

Relevant work in the area of Open Learner Models (OLMs) includes the use of a variety of external representations (e.g., text, progress bars, concept maps, mastery grids, hierarchical graphs, and directed acyclic graphs) and interactive approaches (e.g., interaction protocols, guided exploration, collaborating with a peer, and negotiating the model with a teacher or a virtual tutor) to support human interaction with the model (Bull & Kay, 2016; Zapata-Rivera & Greer, 2002). This work includes the use of visualization techniques to facilitate navigation through big learner models to help identify areas with different levels of supporting evidence (Uther & Kay, 2003) and the use of tree map-based representations to support the exploration of large, complex learner models (Kump, Seifert, Beham, Lindstaedt, & Ley, 2012).

Kay and Kummerfeld (2013) describe how different challenges of creating and deploying OLM have been addressed during more than twenty years of work on the Personis platform. They elaborate on issues such as privacy and control over personalization, making sense of large amounts of LM information, and dealing with errors or inaccuracies of the model. They describe principles for designing an OLM. Some of these principles include: the representation should be as simple as needed for scrutability; support for reasoning on personalization should be provided; and each aspect of the model should be explained and include links to sources of supporting evidence.

Related work also involves research exploring the use of graphical displays to represent uncertainty (e.g., Brodlie, Osorio, & Lopes, 2012; Epp, & Bull, 2015; Hullman, Qiao, Correll, Kale, & Kay, 2019). This includes work on educational assessment using various types of error bars to represent measurement error (Zwick, Zapata-Rivera, & Hegarty, 2014) and using additional materials (e.g., short video tutorials) to support user understanding of this concept (Zapata-Rivera, Zwick, & Vezzu, 2016).

Recommendations and Future Research

We now summarize the key lessons for GIFT in a set of recommendations for GIFT and future ITSs. These draw on established principles for designing visual interfaces. They particularly deal with systematic ways to present visualizations of information that are based on valid assessment.

Follow the principles of effective visualizations and relevant results from areas such as OLM and reporting systems in the design and evaluation of ITS visualizations, dashboards, and reporting systems. This could be implemented by offering training and materials that support the application of best practices (e.g., a tutorial on relevant principles and evaluation approaches), adding automated feedback on problematic or missing aspects of dashboards and reporting systems, and providing templates and use cases that showcase the use of best practices.

Follow an iterative design and evaluation framework that takes into account the information needs of the users of the system when designing graphical communication tools. Conduct continuous evaluation of graphical displays to ensure the appropriate use of assessment results.

Make use of a technological platform that facilitates the alignment between learner model information and available supporting evidence that can be used to develop effective visualizations.

The GIFT authoring environment has the potential to provide ITS creators with a systematic way to follow these recommendations. The interfaces could support all three aspects, as part of a systematic process for the iterative design and evaluation of feedback from assessments.

References

Brodlie, K., Osorio, R.A., & Lopes, A., (2012). A review of uncertainty in data visualization. In Expanding the Frontiers Visual Analytics and Visualization, J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong, Eds. London, U.K.: Springer. 81–109.

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.

Card, S. K., Moran, T. P., & Newell, A. (1983). The psychology of human-computer interaction. 1983.

Cleveland, W. S. & McGill, R. (1983). Graphical perception: Theory, experimentation and application to the development of graphical methods. *Journal of the American Statistical Association, 79*, 531-554.

Epp, C.D., & Bull, S. (2015). Uncertainty representation in visualizations of learning analytics for learners: Current approaches and opportunities. *IEEE Transactions on Learning Technologies, 8* (3). 242-260.

Gillan, D. J. & Richman, E. H. (1994). Minimalism and the syntax of graphs. *Human Factors, 36*, 619-644.

Graesser, A.C., Hu, X., & Sottilare, R. (2018). Intelligent tutoring systems. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, and P. Reimann (Eds.), International handbook of the learning sciences (pp. 246-255). New York: Routledge.

Greer, J. E., & McCalla, G. I. (Eds.). (1994). Student modelling: The key to individualized knowledge-based instruction, Berlin: Springer-Verlag.

Hegarty, M. (2011). The cognitive science of Visual-Spatial displays: Implications for design. *Topics in Cognitive Science, 3*(3), 446-474.

Hegarty, M. (2018). Advances in Cognitive Science and Information Visualization. In Zapata-Rivera, D. (Ed.). Score reporting research and applications. New York, NY: Routledge. 19–34.

Hambleton, R. K., & Zenisky, A. L. (2013). Reporting test scores in more meaningful ways: A research-based approach to score report design. In K. F. Geisinger (Ed.), APA handbook of testing and assessment in psychology: Vol 3. Testing and assessment in school psychology and education (pp. 479–494). Washington, DC: American Psychological Association.

Hullman, J., Qiao, X., Correll, M., Kale, A., & Kay, M. (2019). In pursuit of error: A survey of uncertainty visualization evaluation. In IEEE transactions on visualization and computer graphics, 25(1). 903-913.

John, B. A. & Kieras, D. E. (2006). Using GOMS for user interface design and evaluation: Which technique? *ACM Transactions on Computer-Human Interaction, 3*, 287-319.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50, 1–73.

Kay, J., & Kummerfeld, B. (2013). Creating personalized systems that people can scrutinize and control: Drivers, principles and experience. *ACM Transactions on Interactive Intelligent Systems*, *2*(4), 24–42.

Kay, J., & Kummerfeld, B. (2019). From data to personal user models for life-long, life-wide learners. *British Journal of Educational Technology, 50*(6), 2871-2884.

Katz, I.R., LaMar, M.M., Spain, R., Zapata-Rivera, D., Baird, J., & Greiff, S. (2017). Validity Issues and Concerns for Technology-based Performance Assessments. In R. Sottilare, A. Graesser, X. Hu, and G. Goodwin (Eds.). Design Recommendations for Intelligent Tutoring Systems: Volume 5 - Assessment Methods. Orlando, FL: U.S. Army Research Laboratory. ISBN 978-0-9893923-9-6. 209–224.

Kosslyn (2006). Graph design for the eye and mind. New York, NY: Oxford University Press.

Kump, B., Seifert, C., Beham, G., Lindstaedt, S. N., & Ley, T. (2012). Seeing what the system thinks you know: visualizing evidence in an open learner model. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 153-157). ACM.

Mislevy, R. J. (2012). Four metaphors we need to understand assessment. Commissioned paper for The Gordon Commission on the Future of Assessmentin Education. Princeton, NJ: Educational Testing Service. Retrieved April 13, 2020, from www.ets.org/Media/Research/pdf/mislevy_four_metaphors_understand_assessment.pdf

O'Leary, T. M., Hattie, J. A., & Griffin, P. (2017). Actual interpretations and use of scores as aspects of validity. *Educational Measurement: Issues and Practice, 36*, 16–23.

Roscoe, R.D., Craig, S.D., & Douglas, I. (2018) (Eds.), End-user considerations in educational technology design. Herschey, PA: IGA Global.

Sottilare, R., Brawner, K., Sinatra, A. & Johnston, J. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Orlando, FL: US Army Research Laboratory. May 2017.

Tannenbaum, R. J. (2019). Validity aspects of score reporting. In D. Zapata-Rivera (Ed.), Score reporting research and applications. New York and London: Routledge. 9–18.

Tufte, E. T. (2001). The visual display of quantitative information (2nd Edition). Cheshire, CT: Graphics Press.

Uther, J., & Kay, J. (2003). VlUM, a web-based visualisation of large user models. In International Conference on User Modeling (pp. 198-202). Springer, Berlin, Heidelberg.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin, 138*, 1172–1217.

Wickens, C. D. & Hollands, J. G. (2000). Engineering psychology and human performance. Upper Saddle River, NJ: Prentice Hall Inc.

Zapata-Rivera D. (2019) Supporting Human Inspection of Adaptive Instructional Systems. In: Sottilare R., Schwarz J. (eds) Adaptive Instructional Systems. HCII 2019. Lecture Notes in Computer Science, vol 11597. Springer, Cham. pp. 482-490.

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, D., & Katz, R. I. (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., & Zwick, R (2012). Applying Score Design Principles in the Design of Score Reports for CBAL™ Teachers. ETS Research Memorandum RM-12-20. Princeton, NJ: ETS.

Zapata-Rivera, D., Zwick, R., & Vezzu, M. (2016). Exploring the effectiveness of a measurement error tutorial in helping teachers understand score report results. *Educational Assessment, 21*(3), 215-229.

Zhang, J. (1996). A representational analysis of relational information displays. *International Journal of Human Computer Studies, 45*, 59-74.

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.

CHAPTER 8 – LOOKING AT LEARNING: CHALLENGES AND OPPORTUNITIES FOR VISUALIZATION DESIGN AND EVALUATION

Lane T. Harrison
Worcester Polytechnic Institute

Introduction

Intelligent tutoring systems (ITSs) aim to deliver quality educational experiences to students, while at the same time providing instructors with new and valuable opportunities to observe, understand, and shape student behavior and performance. However, difficult realities such as the scale and complexity of modern tutoring platforms such as the Generalized Intelligent Framwork for Tutoring (GIFT) are bringing new challenges to the continued growth and efficacy of ITSs. This chapter explores several ways in which information visualization techniques, theory, and practice might prove to be a useful means for addressing some of the emerging data-centered challenges in ITSs.

Data scale is a common challenge and a problem that can illustrate the value of information visualization as it relates to ITSs. Scale in terms of the number of learners, for example, may impact how readily instructors can assess and monitor learner progress. With hundreds or thousands of learners, coupled with modern features like learner-directed sequencing of instructional content, interface designers may find it difficult to craft tools for instructors that enable them to quickly, accurately, and reliably analyze learner behavior. Fortunately, issues of scale are a core topic in information visualization research and practice. In this case, a designer drawing on information visualization techniques and theory might adopt systematic design patterns for addressing scale, including visualization-specific prototyping methods for identifying possible scalable encodings, or data-focused task elicitation activities to explore possible summarization or overview strategies.

Because ITSs will face many challenges beyond scale in the coming years, the goal of this chapter is to highlight multiple perspectives of ongoing research threads in the field of visualization. For addressing challenges in analyzing the large and complex data that is already present in ITSs, we discuss several novel visualization techniques that move beyond basic, commonly recognized forms of displaying data. For better defining analytical goals and tasks in learning scenarios, we turn to the development and use of visualization design methodologies. Finally, to ensure that visualizations align with the abilities and needs of end-users, we discuss established visualization evaluation techniques.

In doing so, we seek to lay out a range of possible areas of inquiry at the intersection of visualization science and ITSs, coupled with considerations for which directions might be prioritized to impact theory and practice in the near and long term.

Discussion

Characterizing ITS Data for Novel Visualization Techniques

A unique advantage for ITSs like GIFT is that many forms of data are already collected, managed, and available for analysis and visualization. This data availability highlights several possible directions for research, for instance leveraging and experimenting with the vast number of existing novel visualization forms, many of which have already been successfully applied to similar datasets and domains.

Typically, exploring visualization alternatives requires a preliminary step of describing the structure and properties of the available data, commonly known as data characterization. At a high level, characterization involves describing the general form of the data, along with properties like completeness, variable types, ranges, number of unique values, and similar factors. For example, using language defined by Munzner (2014), learner activity captured in spreadsheets or relational databases would be described as tabular data, whereas nodes of instructional content and links between them would be described as network data. Following the classification of form, data characterization would also define properties of the individual observations (i.e., rows) of the dataset itself, including variable types (e.g., categorical, ordinal, ratio), and limitations such as missing or uncertain measurements.

Proper data characterization lays the groundwork for a systematic exploration of visualization alternatives for a given data source. To illustrate possibilities in this space, we explore novel visualization alternatives for two common forms of data available in many ITSs: log files and network data.

Logging is a pervasive and complex data source in many ITSs, which happens to be amenable to advanced visualization techniques. Logs might be captured for a variety of reasons, for example to observe learner activity, to track course module development and deployment, or for interface-level usage analytics and debugging. While commercial off the shelf tools exist for ingesting and visualizing log data, many of their visualization forms available are relatively simple, e.g., bar charts, pie charts, or line graphs. Basic chart forms remain valuable, as many important insights can be gathered from familiar views of data. However, basic visualization forms are often limited in that they display only a small portion of the available data, sometimes using aggregation due to screen-size constraints, or leaving out important variables due to limitations in the available primary encodings (shape, size, color, etc.).

Several novel visualization techniques address known challenges in analyzing and understanding log data. Recognizing the difficulties of displaying large and complex website clickstream log files, Zhao, Liu, Dontcheva, Hertzmann, and Wilson (2015) developed a technique called MatrixWave. MatrixWave uses a series of dense, heatmap-enabled matrix visualizations to display large clickstream log files. In a user study, MatrixWave was shown to enable users to perform not only basic count and comparison tasks, but also to help them gain insights about patterns of user behavior such as the paths they took while exploring the website. Another visualization system targeting log files, EventPad, addresses the longstanding challenge of mining log data for event sequences of interest (Cappers & van Wijk, 2017). Eventpad consists of a visualization interface that allows users to specify patterns using a visual "language" inspired by regular expressions, which are commonly used in computing for finding patterns in text data. In a user study, EventPad was shown to be an effective tool for exploring telecommunications data, for example to detect missed or repeated calls, and for mining hospital treatment logs for patterns of interest to clinicians.

In the case of MatrixWave, ITS researchers may be able to adapt similar visualization techniques to explore learner activity at scale, or to investigate how groups of instructors navigate the interfaces of a given ITS. Similarly, techniques like EventPad could be deployed to detect, mine, and explore learner activity patterns that might otherwise be difficult to find, particularly given how dense and large many application logging schemes tend to be in practice.

Network data is another form of information present in ITSs that may benefit from the application of novel visualization techniques. In the context of ITSs, network data might consist of links between content modules, or discussions between learners or between learners and instructors. In either case, common challenges in visualizing network data include scale as well as the complexity of the underlying nodes and edges. For example, a network visualization depicting learner-instructor interaction may need to represent complex node-level attributes such as the number of hours learners are spending in assignments, their average grades, or other variables. Nobre, Meyer, Streit, and Lex (2019) systematically classify many possible ways of visualizing multivariate network data, for example, through node-level visualizations or juxtaposition (e.g., where a tabular visualization would appear alongside a traditional network representation), and evaluate some of these alternatives in more recent work (Nobre, Wootton, Harrison, & Lex, 2020). Similarly, other visualization research has developed network visualization techniques designed to scale to larger networks, such as the NodeTrix technique which combines node-link and matrix diagrams (Henry, Fekete, & McGuffin, 2007), or GraphPrism which constructs network "fingerprints" that support the analysis and comparison of large networks (Kairam, MacLean, Savva, & Heer, 2012).

Visualization Design and ITSs

While there is clear value in applying existing, vetted visualization techniques to the data available in ITSs, there is also risk in that the visualizations explored may not align with the actual needs, skillsets, and desires of the people who would use these visualizations. Visualization design methodologies seek to fill this gap. As described by Meyer and Dykes (2019), visualization design studies include process models and validation methods that can be used to systematically explore how visualizations might address particular challenges in a domain. Recent research efforts have significantly expanded and solidified the role of design theory and practice in visualization. For example, in advancing theory, Walny and colleagues argue for moving beyond single "designers" and into team-based methodologies for approaching visualization design and development at scale (Walny et al., 2019). Such considerations seem to align with the realities of large-scale efforts such as GIFT, which spans multiple stakeholder groups and platforms. In advancing practice, Roberts and colleagues introduce the Five Design Sheet methodology, a visualization design exercise that can be used by both visualization developers and stakeholders (Roberts, Headleand, & Ritsos, 2015). In light of the short-term benefits of applying novel visualization techniques to data in ITSs, it may also be worth considering the potential long-term benefits of employing visualization design activities as a driving force to shape visualizations in service of ITSs.

Evaluating Visualizations for ITSs

A need for rigorous, transparent evaluation is common to both the technique-driven and design-driven approaches discussed thus far. Evaluating visualizations remains challenging, however, particularly in multi-stakeholder domains such as ITSs. Fortunately, perspectives on evaluating visualizations have matured in recent years, and there are widely used models for visualization evaluation that mitigate common pitfalls identified in past efforts. In particular, the nested model for visualization evaluation proposed by Munzner (2009) identifies different layers for consideration in evaluation: the domain problem, data/operation abstractions, encoding/interaction techniques, and algorithm design. Visualization research targeting ITSs might begin with activities focused on domain characterization, to identify data-oriented problems and stakeholders that could become topics for future technique exploration or design efforts.

Recommendations and Future Research

Considering the wealth of data already available in ITSs like GIFT, there appears to be several promising pathways towards applying not only novel visualization techniques, but also visualization design and evaluation methodologies in service of ITS stakeholder goals. This chapter highlights several of these opportunities, advocating for both short-term and long-term strategies that could extend ongoing research threads in visualization to serve the emerging challenges and growth of ITSs.

References

Cappers, B. C., & van Wijk, J. J. (2017). Exploring multivariate event sequences using rules, aggregations, and selections. *IEEE transactions on visualization and computer graphics, 24*(1), 532-541.

Henry, N., Fekete, J. D., & McGuffin, M. J. (2007). Nodetrix: a hybrid visualization of social networks. *IEEE transactions on visualization and computer graphics, 13*(6), 1302-1309.

Kairam, S., MacLean, D., Savva, M., & Heer, J. (2012, May). Graphprism: compact visualization of network structure. In Proceedings of the International Working Conference on Advanced Visual Interfaces (pp. 498-505).

Meyer, M., & Dykes, J. (2019). Criteria for Rigor in Visualization Design Study. *IEEE transactions on visualization and computer graphics, 26*(1), 87-97.

Munzner, T. (2009). A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics, 15*(6), 921-928.

Munzner, T. (2014). Visualization analysis and design. CRC press.

Nobre, C., Meyer, M., Streit, M., & Lex, A. (2019, June). The state of the art in visualizing multivariate networks. In Computer Graphics Forum (Vol. 38, No. 3, pp. 807-832).

Nobre, C., Wootton, D., Harrison, L., & Lex, A. (2020, April). Evaluating Multivariate Network Visualization Techniques Using a Validated Design and Crowdsourcing Approach. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-12).

Roberts, J. C., Headleand, C., & Ritsos, P. D. (2015). Sketching designs using the five design-sheet methodology. *IEEE transactions on visualization and computer graphics, 22*(1), 419-428.

Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S., & Willett, W. (2019). Data Changes Everything: Challenges and Opportunities in Data Visualization Design Handoff. *IEEE transactions on visualization and computer graphics, 26*(1), 12-22.

Zhao, J., Liu, Z., Dontcheva, M., Hertzmann, A., & Wilson, A. (2015, April). Matrixwave: Visual comparison of event sequence data. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 259-268).

Section II– Data Visualization in Specific Domains and Applications

Dr. Xiangen Hu and Dr. Benjamin Goldberg, Eds.

Chapter 9 – Introduction to Data Visualization In Specific Domains and Applications

Benjamin Goldberg[1] and Xiangen Hu[2]
U.S. Army Combat Capability Development Command – Soldier Center[1], University of Memphis[2]

Core Ideas

A common answer to a question regarding the best way to visualize information often entails an "it depends" type response. When considering the visualization of data generated from an interaction within an Adaptive Instructional System (AIS), the role and intent of the consumer of that information needs to be carefully considered when implementing visualization techniques. Ideally, we can generate a set of best practices that can guide the visualization design and implementation process across an ecosystem of resources.

In this book section, each chapter highlights the utility of data visualizations as they adhere to a specific domain or learning application. There is a diverse representation of domains and applications represented, by which each use case highlights a different role, function, and interaction with different types of data visualization.

Individual Chapters

Johnston, Sinatra, and Swiecki provide a good practitioner's perspective in applying data visualization best practices to drive efficient training outcomes. The chapter authors describe techniques and intent as they relate to the project Squad Overmatch (SOvM). SOvM has a large emphasis on team learning behaviors and using structured After Action Reviews (AAR) to promote optimal team development through guided discussion and reflection. In this context, visualization plays a critical role in communicating the measures and inferences collected during training, by assisting the training audience in narrowing down the topics and inputs for discussion. At the conclusion, the authors provide a set guidelines and implications for future adaptive training and tutoring design consideration.

Stevens, Mullins, Hu, Zapata-Rivera, and Galloway provide a hierarchical framework and applied use case for visualizing Electroencephalogram (EEG) derived data streams into representations of collective uncertainty through modeling neurodynamic organizations. The authors highlight the implications of team uncertainty on cognitive functions across varying contexts and timescales, and the importance of accounting for its effects in training and measurement. The chapter continues with a detailed study involving junior enlisted officers completing a set of collective submarine training simulations while donning EEG headsets. The results of the modeling efforts are presented, with guideline recommendations for using these techniques to validate and improve AIS, to support AAR processes and discussion, and to extend the methods to create a real-time predictive classifier for states of team uncertainty.

Misfeldt, Swiecki, Zapata-Rivera, and Hu present considerations for pedagogical data visualizations. The authors argue that the prepare, conduct, and evaluate (PCE) framework should be considered for understanding pedagogical situations. They have demonstrated how this framework would be implemented visually, in the spirit of *data visualization*, in two examples. They have pointed out challenges implementing PCE in their applications. They have suggested that PCE guided data visualization be implemented in the

Generalized Intelligent Framework for Tutoring (GIFT) as quality assurance guidelines for content creators, learning scientists, and teachers, when assessing the practical usability and value of solutions.

Sottilare, Folsom-Kovarik, Cockroft, and Hampton explore data visualization with “digital twins”. They propose to visually “duplicate” physical systems in training so learner’s performance as well as changes of scenarios can be monitored in real-time.

Hu, Rus, Cockroft, and Zhang consider the general form of “data” (digital or/and analog) in data visualization. In their chapter, data visualization is defined as the process to represent data in pictorial or graphical format for human cognition. They argue that facial expressions (of fear, happiness, etc.) are a form of real-time “data” visualization. With this analogy, they propose that data visualization for Intelligent Tutoring Systems (ITSs) should include facial expressions of avatars when interacting with learners in real-time.

Kay, Rus, Zapata-Rivera, and Durlach provide design guidelines for ITSs from a metacognitive processes point of view of the learners. They pointed out that when learners interact with ITSs, there are fast and slow metacognitive processes and visualization of ITSs should be optimized for each of the processes. They made a strong argument by examining several Open Learner Model (OLM) applications. They conclude that we should consider fast and slow metacognitive processes when designing and implementing visualization methods in ITSs.

CHAPTER 10 – APPLICATION OF TEAM TRAINING PRINCIPLES TO VISUALIZATIONS FOR AFTER-ACTION REVIEWS

Joan H. Johnston[1], Anne M. Sinatra[1], and Zachari Swiecki[2]
U.S. Army Combat Capabilities Development Command (DEVCOM) - Soldier Center – Simulation and Training Technology Center (STTC)[1]; University of Wisconsin-Madison[2]

Introduction

Research has shown that using effective after action review (AAR) methods can be crucial to improving team performance (Allen, Reiter-Palmon, Crowe, & Scott, 2018; Johnston et al., 2019; Lacerenza, Marlow, Tannenbaum, & Salas, 2018). The most recent Army Doctrine Publication (ADP) 7-0 (Headquarters, Department of the Army, 31 July 2019) indicates AARs have significantly evolved as the result of years of dedicated research to find the most effective strategies (Morrison & Meliza, 1999). The ADP 7-0 defines an AAR as:

> "…a guided analysis of an organization's performance, conducted at appropriate times during and at the conclusion of a training event or operation with the objective of improving future performance. It includes a facilitator, event participants, and other observers. An AAR enables an organization to objectively ascertain its mastery of tasks. AARs are conducted as needed during and following a training event. Participants record observations, insights, and lessons learned for future use to identify trends and prevent reoccurrences of improper practices" (pp. 4-11).

Recently, the Army specified a need for intelligent tutoring and AAR technologies in collective training simulations to accelerate learning and reduce overhead training costs (U.S. Army Common Synthetic Environment Statement of Need, 14 March 2019). A major capability gap is the need for AAR technologies that support team learning requirements (Salas, Reyes, & McDaniel, 2018). Currently, standard practice for AAR visualizations in simulation based training is to record player actions and movement through the virtual simulation and replay it as a visual aid to support discussions between instructors and trainees about what happened. An AAR for live training typically uses video/audio snippets of team actions such as is depicted in Figure 1 where a female roleplayer plays an injured civilian being taken to a casualty checkpoint by soldiers (Johnston et al., 2016). The review is usually supplemented by trainers or observer/controllers (OCs) describing their observations based on notes they made from task checklists and engaging the team in a discussion about it. Little research-based guidance exists for incorporating visualizations that depict for example, team performance processes and outcomes into AARs based on principles of team learning and training. Therefore, in this chapter we provide an overview of these principles to establish a framework for the design of AAR visualizations, describe a military team training use case that incorporated AAR visualizations based on these principles, and discuss the implications for designing and developing adaptive team training environments.

Science of Team Learning and Training

Over the last thirty years, military team researchers have focused on designing and developing simulation-based training technologies that develop complex decision making, resilience and teamwork skills critical to performing in high stress combat environments (Goodwin, Blacksmith, & Coats, 2018). To be effective, teams must develop both taskwork and teamwork skills. Such taskwork skills as situation awareness and decision making must be performed as a collective unit. Teamwork skills are the communication and coordination skills that enable teams to effectively perform the task. Decision making and teamwork skills involve employing "team learning behaviors" that enable team members to attend to, recognize, understand, and share information necessary to achieve the task objectives.

Figure 1. Snapshot from a video of a female roleplayer with a fake gunshot wound being taken to a casualty collection point by two Soldiers in a live training setting (Adapted from Johnston et al., 2016).

Team learning is a dynamic and emergent process that results in changes in a team's collective knowledge state; and involves three types of behaviors – fundamental, intra-team, and inter-team (Wiese & Burke, 2019). Fundamental team learning involves the individual actions team members perform to share and exchange, store in memory, and retrieve from memory the information that enables gaining situation awareness. Sharing involves making a team member aware of individually held information; memory storage behaviors are those that act to maintain collective knowledge over time; and retrieving information involves recalling information from knowledge repositories. Intra-team learning is the dialogue (whether electronic, verbal or non-verbal) among team members that involves asking questions; seeking feedback; seeking out new knowledge; discussing information together to test ideas and assumptions that may deviate from expectations; discussing errors; collaboratively discussing information that changes collective knowledge; engaging in dialog that involves confrontation, negotiation, and resolving divergences in interpretations and opinions; and reviewing/reflecting on previous team functioning (reflexivity). Inter-team learning involves scanning/querying the outside environment for necessary information; and actively creating partnerships or collaborations. Team training research has shown that incorporating team learning behaviors into the AAR process enables the development of team learning skills (Johnston et al., 2019).

Recent literature reviews describe how creating an efficient AAR feedback structure and supportive climate reinforces the use of team learning behaviors during AAR discussions while mitigating negative team dynamics that can derail the use of these behaviors (Allen et al., 2018; Lacerenza et al., 2018). Physical location, roles, and cognitive biases often result in team members developing different perspectives about what actually happened in a scenario. To mitigate these problems, an effective AAR feedback structure uses an event-based approach to elicit review, recall, and discussion about critical incidents that focus on supporting training objectives. To facilitate consideration of multiple viewpoints, leaders maintain the focus on discussions, listen attentively to conversations, encourage participation, and model such behaviors as reflection, sharing information, and respectful interactions.

This approach enables "sense making" activities through the use of team learning behaviors, focusing on behavioral outcomes, thus preventing the discussion from becoming sidetracked by issues that do not relate to learning objectives. Team leaders can effectively facilitate sense making by encouraging team members to openly reflect and discuss their different viewpoints. This is called using team self-correction. By discussing good and poor behaviors rather than blaming people for errors, team leaders can promote "psychological safety" which is the tendency to stop worrying, feel safer in the group, and be willing to take risks in voicing criticisms or concerns. When individual views are shared in a well-facilitated discussion, members not only share their perspectives, but are more open to views being challenged, supported, modified, and combined until some degree of consensus about the incident is developed. When teams engage in reflexivity, members share and discuss relevant information, elaborate on information shared, and use it to change preconceived notions when they are inappropriate. This approach can reduce ambiguity sufficiently to produce enough shared understanding to support group learning. To summarize we condensed the principles described above into five general AAR guidelines:

- Guideline 1 - Review learning objectives;
- Guideline 2 - Use an event-based approach to scenario review;
- Guideline 3 - Conduct behaviorally-based discussions focused on learning objectives;
- Guideline 4 - Facilitate participation, reflection, sharing multiple viewpoints, supporting challenges to views, revising views, and gaining consensus (team self-correction); and
- Guideline 5 - Facilitate sense-making through fundamental, inter, and intra team learning behaviors.

To illustrate application of these guidelines we describe a military training use case that applied the principles to the design of visualizations used to support team AARs (Johnston et al., 2019).

Visualization Use Case: Squad Overmatch for Tactical Combat Casualty Care

The Squad Overmatch (SOvM) for Tactical Combat Casualty Care (TC3) research program was a multi-year, joint US Army – US Navy research effort to improve team performance under stressful conditions using an integrated training approach (ITA) that included classroom, simulation-based training (SBT) using the Virtual Battlespace Simulation 3 (VBS3), and live training in an outdoor village of about 15 small buildings (Johnston et al., 2019). The ITA focused on improving five skill areas: 1) Understanding team member roles and priorities in response to medical tactical situations (TC3); 2) Observation and communication of critical environmental cues to identify hostile, friendly, or neutral actors (Advanced Situation Awareness or ASA); 3) Maintaining task focus and reducing physiological responses during high intensity scenario events (Stress management); 4) Teamwork through information exchange, communication delivery, supporting behavior, and team initiative/leadership; and 5) Conducting AARs using team self-correction following SBT and live training. Next we present examples of the visualization graphics used during the AAR and discuss how they supported the five guidelines.

SOvM TC3 AAR Framework

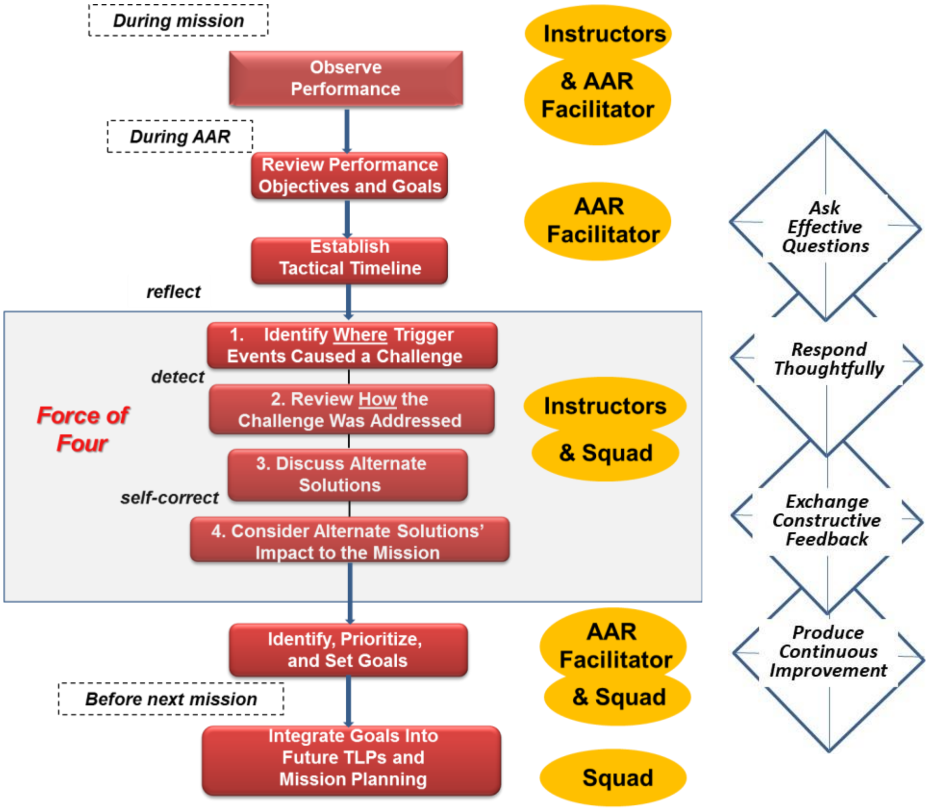
Figure 2 presents an outline of the AAR tasks and procedure that incorporated reminders from each of the guidelines (Townsend et al., 2018). It was developed by the SOvM research team to be used as a visual support aid in the SOvM Quick Reference Guide (QRG) that was issued to each Soldier (Figure 3) (Wolf & Johnston, 2016). Instructors/OCs and the Platoon Leader encouraged Soldiers to refer to the QRG throughout classroom and training exercises, and during the AARs. The yellow ovals in Figure 2 indicate responsibilities for each segment of the process. For example the Instructor and squad members used the Force of Four tasks during the AAR which reminded them to "reflect, detect, and self-correct" and the four diamonds on the right hand side of the figure were a reminder to participants to use facilitative behaviors to foster team learning behaviors (Guideline 4). Figure 2 was an effective prompt to keep the teams from getting derailed in their discussions by, for example, assigning blame, letting one person do all the talking, being overly critical of team members, or only discussing tactical performance and ignoring teamwork or situation awareness.

Observe Performance

The OCs/Instructors observed and took notes of squad actions during both the virtual and live scenario-based training. Then they huddled with the AAR facilitator (Platoon Leader) to choose the most important critical events and squad behaviors for review (Guideline 2). The facilitator asked instructors whether their training objectives were met, if there were any squad weaknesses, if goals were met, and when they wanted to talk in relation to the scenario event timeline during the AAR (Guideline 1, 2, and 3). Each instructor defined major training objectives, and then identified related errors and examples of good performance (Guideline 3). Finally, to ensure AAR organization and to optimally utilize the time allotted, the facilitator designated someone to take notes and present the learning objectives summary at the end of the AAR.

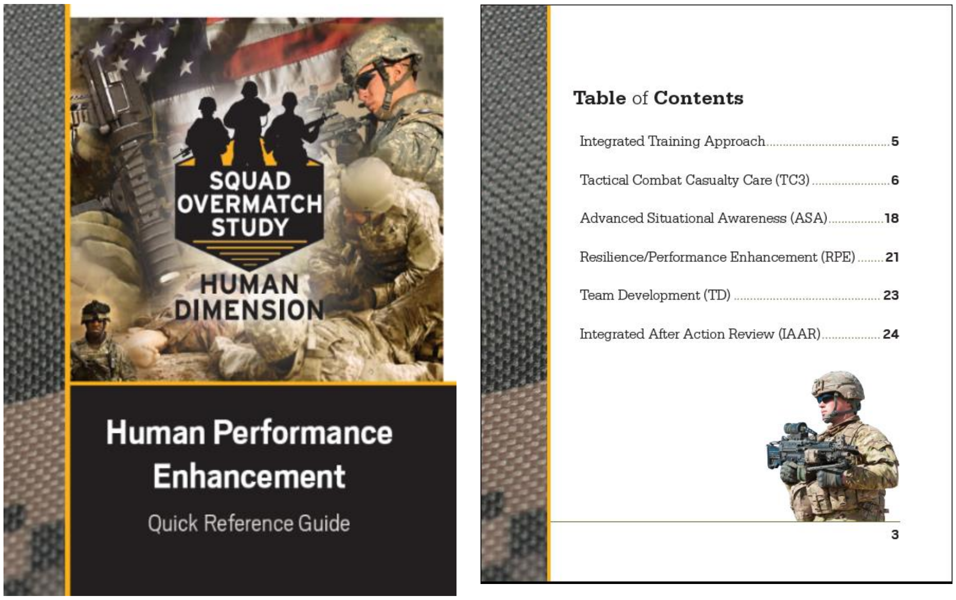
Review Performance Objectives and Goals

At the beginning of the AAR, Figure 2 was projected on a large screen display and team members were also directed to view the same graphic in their SOvM QRG (Figure 3). The squad's performance objectives were projected in a bulleted format on a second large screen display for the facilitator to review with the squad prior to detailed discussions (Guideline 1). For example, Figure 4 presents all the teamwork performance objectives (Wolf & Johnston, 2016).



During mission
Observe Performance
Instructors & AAR Facilitator
During AAR
Review Performance Objectives and Goals
Establish Tactical Timeline
AAR Facilitator
reflect
1. Identify Where Trigger Events Caused a Challenge
detect
2. Review How the Challenge Was Addressed
Force of Four
3. Discuss Alternate Solutions
self-correct
4. Consider Alternate Solutions' Impact to the Mission
Instructors & Squad
Identify, Prioritize, and Set Goals
AAR Facilitator & Squad
Before next mission
Integrate Goals Into Future TLPs and Mission Planning
Squad
Ask Effective Questions
Respond Thoughtfully
Exchange Constructive Feedback
Produce Continuous Improvement

Figure 2. SOvM TC3 performance observation and AAR process (Adapted from Townsend et al., 2018).



SQUAD OVERMATCH STUDY

HUMAN DIMENSION

Human Performance Enhancement

Quick Reference Guide

Table of **Contents**

Integrated Training Approach 5

Tactical Combat Casualty Care (TC3) 6

Advanced Situational Awareness (ASA) 18

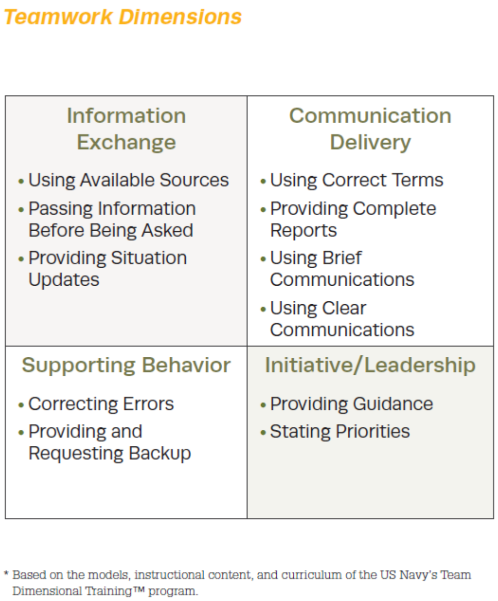
Resilience/Performance Enhancement (RPE) 21

Team Development (TD) 23

Integrated After Action Review (IAAR) 24

3

Figure 3. SOvM Human Performance Enhancement Quick Reference Guide (Adapted from Wolf & Johnston, 2016).



Teamwork Dimensions

Information Exchange	Communication Delivery
• Using Available Sources • Passing Information Before Being Asked • Providing Situation Updates	• Using Correct Terms • Providing Complete Reports • Using Brief Communications • Using Clear Communications
Supporting Behavior	**Initiative/Leadership**
• Correcting Errors • Providing and Requesting Backup	• Providing Guidance • Stating Priorities

* Based on the models, instructional content, and curriculum of the US Navy's Team Dimensional Training™ program.

Figure 4. Example of teamwork learning objectives (Adapted from Wolf & Johnston, 2016).

Establish Tactical Timeline

Following the learning objectives overview, an outline of tactical events was presented on the large screen display (Guideline 2). For example, Figure 5 depicts the flow of events for the live training scenario "M2:

Panther;" with the top half of the figure (Part 1 of 2) showing events in the first half of the scenario, and the bottom half of the figure (Part 2 of 2) showing the sequence of events in the second half of the scenario (Townsend et al., 2018). The location of the scenario was an outdoor, urban training site comprised of a set of buildings configured as a small village. The squad's mission was to conduct a zone reconnaissance in order to conduct a Key Leader Engagement (KLE); exploit intelligence; confirm location of a suspected arms cache; and exploit the site, if able. The facilitator quickly reviewed with the squad the sequence of events in Figure 5 to establish "ground truth" and a timeline for what happened, pointing out where the team's skill areas were expected to be exercised the most, and setting the stage to facilitate the team's sense-making activities (Guideline 5).

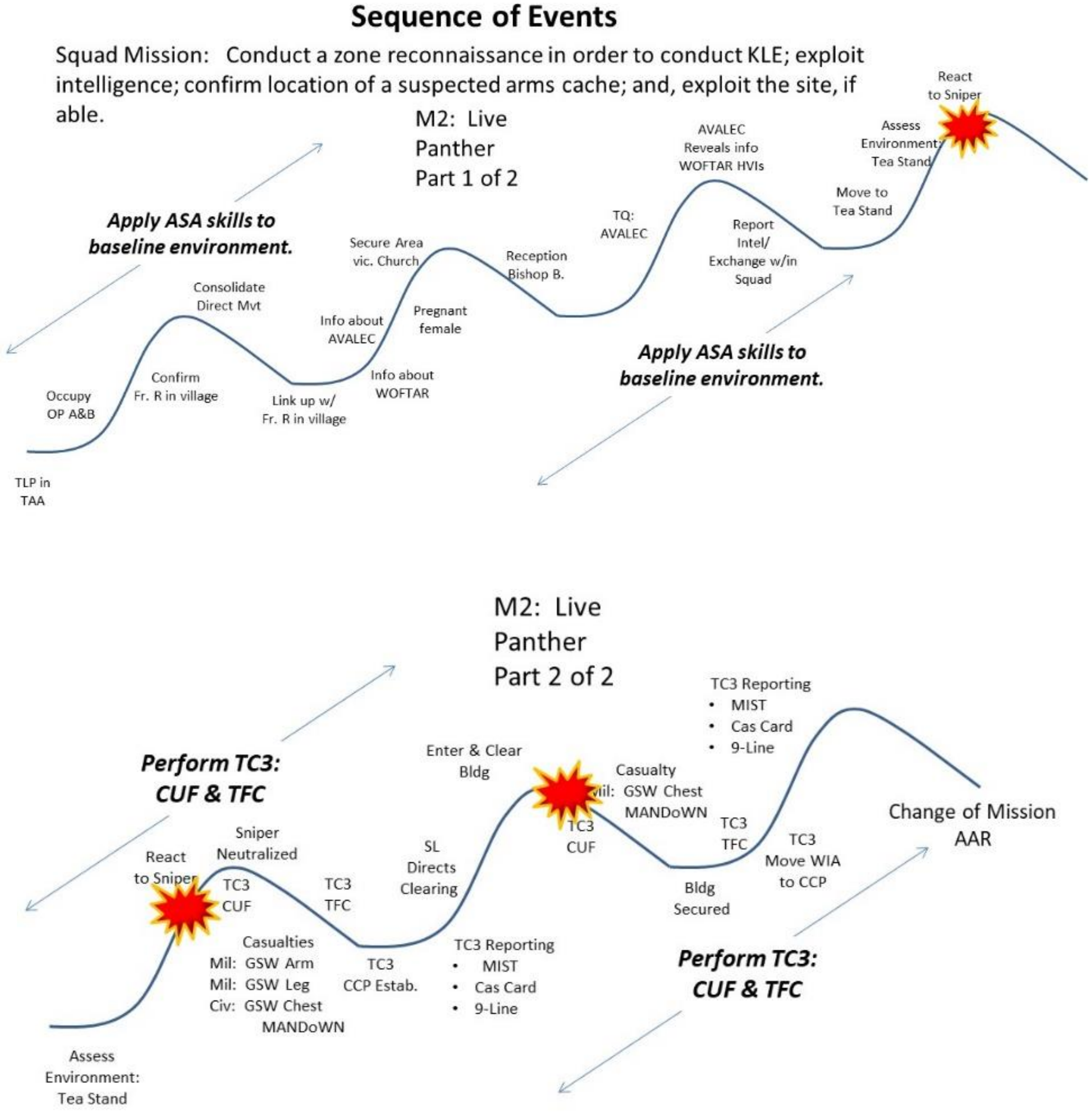


Figure 5. Example of tactical scenario events (Adapted from Townsend et al., 2018).

For example, scenario M2 begins with the squad establishing their location in the Tactical Location Point (TLP) and then moving to occupy the Observation Post (OP) with fire teams A and B. The third event is for the squad to confirm that the village elder (Father Romanov) is in the village. The fourth event is the squad moving to consolidate together and direct their movement toward Father Romanov, and the fifth event is for the Squad Leader to link up with Father Romanov in the village for the first KLE. Most of the first part of the scenario has the squad focusing on using their ASA skills. The red star blast in the upper right of Figure 5 depicts a casualty event in which the squad has to react to a sniper attack and use TC3 skills on several military and civilian casualties who received gunshot wounds (GSW). The second half of the scenario indicates greater emphasis on TC3 skills.

Conduct Force of Four Discussions

The tactical timeline such as shown in Figure 5 was then used to frame the Force of Four discussions. In addition, a picture depicting a birds-eye view of village buildings with identifying markers and color-coded roads (e.g., Route Green, Route Black, and Army Service Road Golden) was projected next to it on the large screen display (Figure 6). The squads had used Figure 6 throughout training as an important reference document for planning their movements throughout the village, and so it was also used in the AAR to establish a frame of reference to establish where each Soldier had been located during the scenario, and to conduct sense-making discussions (Guideline 5). The instructors then used their written notes to guide the Force of Four discussions. Each instructor presented expected learning objectives and asked questions requiring squad members to monitor and reflect on their own and their squad's performance using the Force of Four framework (Guideline 5). Squad members were asked to identify scenario events where they struggled and excelled; agree on what went wrong and right; propose a workable solution (i.e., identify correct procedures); and discuss real world outcomes and consequences, good or bad (Guidelines 4 and 5). As a visual support during the virtual training AAR, the simulation controller showed the simulation replay of squad member location and flow through the scenario in birds-eye view mode, and also first person mode (Guidelines 2 and 4). As a visual support during the live exercise AAR, the facilitator had a video manager present a video/audio snippet of the squad members during critical events in the outdoor facility (Guidelines 2 and 4).

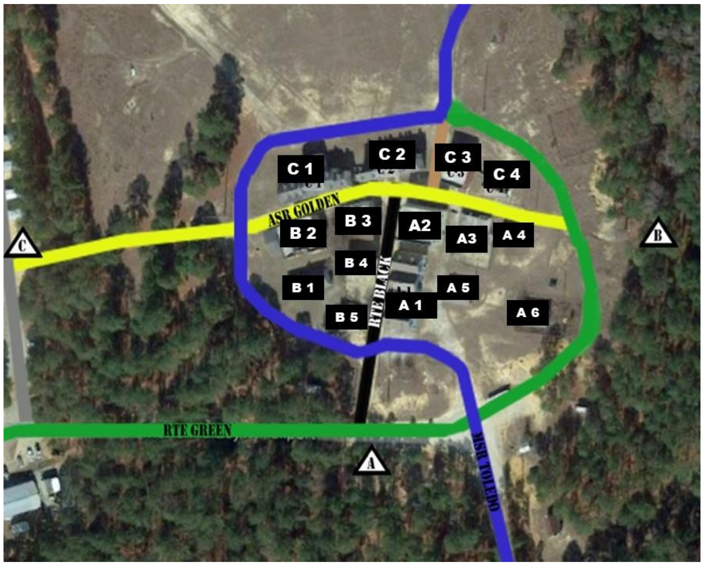


Figure 6. Birds-eye view of village buildings with identifying markers, and roads with color-codes.

Set Goals

When the AAR was complete, the facilitator and instructors filled in a blank PowerPoint slide on the large screen display based on the squad's consensus in setting new goals for behavioral improvements and behaviors they wanted to sustain, and if there was a new scenario, they integrated the goals and sustains into their next mission's planning brief (Guidelines 3 and 4).

Implications for Adaptive Tutoring and Training

The SOvM AAR visualization tools developed to support the team learning guidelines have implications for intelligent tutoring technologies. To integrate the five guidelines into training events, at least initially, two user capabilities would be needed. One capability would support facilitators, OCs, and team leaders during the phases of training management before, during, and after a training scenario run. While a scenario is being developed tools could be available to plan and author AAR templates (such as performance objectives in Figure 4, the event-based tool in Figure 5, and the Birds-Eye view in Figure 6) that would be populated following the scenario with a record of the critical training event (e.g., video, audio, and virtual replay), whether objectives were met, and other observations made by OCs during the exercise.

The second user capability would directly support the team being trained, collecting their experiences for the AAR and their sense-making activities. An approach would be needed for translating this information into a simplified representation that would be seen by the team and used in AAR. For example, the event timeline in Figure 5 could be populated to include evaluations from the facilitators and links to the training event record. The force-of-four process in Figure 2 could be embedded or induced to pop-up in the AAR event timeline to aid prompting the team to discuss performance using the force-of-four format during each key event. Discussions could be supplemented with the birds-eye view visualization (Figure 6) of the squad moving through the scenario to enable them to observe and hear what was said during the events.

In the future, aspects of the AAR should be automated. For example, critical events should be identified and noted during the plan and prepare stage of training when the scenario is first authored and then classifiers could be developed based on prior data to evaluate team performance at these critical events. Then, these evaluations should be automatically integrated into the timeline visualization that teams would step through during the AAR. Of course, this system would depend on the ability to automatically and reliably collect data from the training event which is more difficult in non-virtual scenarios. It would also be important to maintain flexibility in the system so that human facilitators could add new critical events and their own evaluations rather than completely relying on automation.

In the specific case of the SOvM scenario, elements of it could be represented in the Generalized Intelligent Framework for Tutoring (GIFT) (Sottilare, Brawner, Sinatra, & Johnson, 2017), and some of the existing tools could be expanded to assist in automating the described AAR. GIFT has been used in both computer-simulation based environments and as an element of psychomotor tasks (Goldberg, Amburn, Ragusa & Chen, 2018). In the psychomotor implementations, they have been tied to a computer system which provided automated feedback based on the performance of the learner and also available on a mobile device which provided direct feedback to the learner during a land navigation task.

GIFT has a fairly new interface and tool which is currently named GameMaster (Goldberg & Hoffman, 2019). In the current implementation the GameMaster interface was designed with OCs in mind. The interface can be displayed on a computer, or in most cases, a tablet for ease of use in a field environment. The concepts from the GIFT course are available in real time in GameMaster and it provides a way to see the automated assessments that are occurring and also override them if the OC chooses. If a GIFT assessment is not automatic, and requires human intervention, the OC will use this interface to do so. It also allows for human prompted injects of actions in a virtual environment, and for ratings to be provided by the OC during an interaction (Goldberg & Hoffman, 2019). The OC can also use this interface to make notes during performance, and the session can then be played back after the fact. While the playback of the scenario session can serve as an unstructured AAR, it is also a tool that can be used to support the AAR and allow for context to be provided for the actions that took place. Further, the GameMaster interfaces with the ARES sandtable (Garneau, Boyce, Shorter, Vey & Amburn, 2018), which allows for an additional way of visualizing a scenario.

As noted above, we next describe how GIFT and the GameMaster interface currently address the team learning guidelines and provide recommendations for necessary changes to improve capability in the future.

Guideline 1: Review learning objectives

The GameMaster interface reads in the content of the GIFT Domain Knowledge File (DKF) and the associated names that have concepts within it. Therefore, in real time (and in playback), the instructor/OC has access to how the learners are performing on each of the identified learning objectives and assessments. However, in order for this information to be useful, careful attention will need to be paid to the names that are given to each concept during the DKF authoring process.

Guideline 2: Use an event-based approach to scenario review

The GameMaster has the ability to play back only the scenario, but the events within it could be discussed and reviewed by the instructor after the fact. In the future it should automatically provide bookmarks or tags to important events that happen in a scenario.

Guidelines 3 and 4: Conduct behaviorally-based discussions focused on learning objectives; Facilitate participation, reflection, sharing multiple viewpoints, supporting challenges to views, revising views, and gaining consensus (team self-correction)

By engaging in playback of the performance during the scenario, discussions can be elicited by the OC about actions taken in the scenario and what was done right and what was done wrong. In the current implementation of GameMaster, and the associated demonstrations the playback consists of a map view of the movements that were made in a training environment (VBS3) and a playback of the ratings on concepts throughout the interaction. To support these guidelines the GameMaster should be integrated with the training simulation to enable showing the learner view as they engaged in the environment, and should be synchronized with the performance assessments.

Guideline 5: Facilitate sense-making through fundamental, inter-, and intra-team learning behaviors

The GIFT course and accompanying DKF should be written to highlight the fundamental, inter-, and intra-team behaviors so that it can assist in reviewing the successful and unsuccessful behaviors. This further relies on the course author and the way that the assessments are structured during the authoring process. Tags for types of team behaviors could be added to the concepts in the DKF to provide a separation to filter the different types of assessments that would support a new overall performance AAR interface for the GameMaster.

Without any major updates to GameMaster, the SOvM simulation scenarios could be implemented within GIFT. A DKF would need to be created that includes the information about the concepts that are being assessed both automatically and by the OC during the interaction in VBS3. A similar DKF could be used to have GIFT available for assessments during the live version of the SOvM event. However, in the case of the live event the ratings would primarily rely on human inputs. Additionally, the live assessments could also be supported by Global Positioning System (GPS) data for tracking movements and distance. With the current GameMaster interface a playback of the concepts and assessments could be shown after the fact in order for the instructor or OC to determine what the Squad did properly or improperly during the interaction. In the current form this information could be used as a tool to provide information that would allow a more traditional AAR to be generated.

To add to the current GameMaster tool, to support AARs, the consolidated information about the concepts and ratings throughout the experience could be filtered into additional configurations including a potential tactical timeline. An overall summary screen which includes the concepts, event timestamps, and assessments would be helpful. Additionally, the ability to click on a concept and see a playback on a map of the actions that were taken in VBS3 could be beneficial. The creation of an additional interface that reads in the information from GameMaster, and is reconfigurable based on the information that the instructor would like to see visualized could be useful.

In summary, in this chapter we discussed five guidelines for conducting an effective military AAR using visualization techniques that enhance team learning and significantly improve team performance. We discussed how GIFT and the GIFT GameMaster have the current capability to support the guidelines with visualizations, and the future potential to incorporate enhanced data visualization tools especially when combined with intelligent tutoring.

References

Allen, J. A., Reiter-Palmon, R., Crowe, J., & Scott, C. (2018). Debriefs: Teams learning from doing in context. *American Psychologist*, *73*(4), 504.

Garneau, C. J., Boyce, M. W., Shorter, P. L., Vey, N. L., & Amburn, C. R. (2018). *The augmented reality sandtable (ARES) research strategy* (No. ARL-TN-0875). US Army Research Laboratory Aberdeen Proving Ground United States.

Goldberg, B., Amburn, C., Ragusa, C., & Chen, D. W. (2018). Modeling expert behavior in support of an adaptive psychomotor training environment: A marksmanship use case. *International Journal of Artificial Intelligence in Education*, *28*(2), 194-224.

Goldberg, B., & Hoffman, M. (2019, May). Intelligent exercise control: Integrating GIFT and battle space visualization. Presentation at the *7th Annual Generalized Intelligent Framework Tutoring Symposium (GIFTSym7*), Orlando, FL.

Goodwin, G.F., Blacksmith, N., & Coats, M.R., (2018). The science of teams in the military: Contributions from over 60 years of research. *American Psychologist*, *73*(4), 322-333. Acquired January 8, 2020 from https://www.apa.org/pubs/journals/releases/amp-amp0000259.pdf.

Headquarters, Department of the Army (31 July 2019). *Training (Army Doctrine Publication 7-0*). Washington, DC: Headquarters, Department of the Army.

Johnston, J., Gamble, P., Patton, D., Fitzhugh, S., Townsend, L., Milham, L., Riddle, D., Phillips, H., Smith, K., Ross, W., Butler, P., Evan, M., & Wolf, R. (16 December, 2016). *Squad Overmatch for Tactical Combat Casualty Care: Phase II Initial Findings Report*. Orlando, FL: Program Executive Office Simulation, Training and Instrumentation.

Johnston, J. H., Phillips, H. L., Milham, L. M., Riddle, D. L., Townsend, L. N., DeCostanza, A. H., ... & Fitzhugh, S. M. (2019). A Team Training Field Research Study: Extending a Theory of Team Development. *Frontiers in Psychology*, 10, 1480.

Lacerenza, C. N., Marlow, S. L., Tannenbaum, S. I., & Salas, E. (2018). Team development interventions: Evidence-based approaches for improving teamwork. *American Psychologist*, 73(4), 517.

Morrison, J. E., & Meliza, L. L. (1999). *Foundations of the after action review process (Special Report 42*). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Salas, E., Reyes, D. L., & McDaniel, S. H. (2018). The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist*, *73*(4), 593.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT). *GIFTtutoring.org*.

Townsend, L., Johnston, J., Ross, W. A., Milham, L., Riddle, D., & Phillips, H. (2018). An integrated after action review (AAR) approach: Conducting AARs for scenario-based training across multiple and distinct skill areas. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics* (Vol. 10906, pp. 230–240). Cham, Switzerland: Springer International Publishing.

U.S. Army Common Synthetic Environment (CSE) Statement of Need (14 March 2019). Acquired January 8, 2020 from https://trainingaccelerator.org/wp-content/uploads/2019/03/CSE-RFS_Attachment-1-SoN_14Mar.pdf.

Wiese, C. W., & Burke, C. S. (2019). Understanding Team Learning Dynamics over Time. *Frontiers in Psychology*, *10*, 1417.

Wolf, R., & Johnston, J. (2016). *Squad Overmatch, Human Dimension, Human Performance Enhancement: Quick Reference Guide*. Orlando, FL: Program Executive Office for Simulation, Training, and Instrumentation.

Chapter 11– VISUALIZING THE MOMENTARY NEURODYNAMICS OF TEAM UNCERTAINTY

Ron Stevens[1, 2], Ryan Mullins[3], Xiangen Hu[4], Diego Zapata-Rivera[5], and Trysha Galloway[2]
UCLA School of Medicine[1], The Learning Chameleon, Inc.[2], Aptima, Inc.[3], University of Memphis [4],
Educational Testing Service[5]

Introduction

Every human decision whether conscious or not, has its origins in uncertainty. Quantitatively, uncertainty has many definitions; as a property of information (Aggarwal, 2009; Chatfield, 1995; Thomson, Hetzler, MacEachren, Gahegan, & Pavel, 2005), as a property of cognitive processes (Hodgkinson, Bown, Maule, Glaister, & Pearman, 1999; Kennedy, 2011), or as a property of the electrochemical interactions of neurons (Stevens, Galloway, Halpin & Willemsen-Dunlap, 2016). Operationally, uncertainty inhibits a decision-maker's ability to distinguish between and select options.

There is a rich history of using visualization techniques to represent uncertainty in various fields. These include literature reviews on visualization techniques for uncertainty (Brodlie, Osorio, & Lopes, 2012; Demmans Epp & Bull, 2015), frameworks for understanding and evaluating the use of these visualization techniques on affecting human behavior (Hullman, Qiao, Correll, Kale, & Kay, 2019), and work on depicting and supporting teachers' understanding measuring error (Zapata-Rivera, Zwick, & Vezzu, 2016; Zwick, Zapata-Rivera, & Hegarty, 2014).

This work has sought to depict the characteristics of uncertainty such that decision-makers are able to effectively and efficiently compare alternatives. Some examples include the cone of uncertainty (Cox, House, & Lindell, 2013), quantile dot plots (Kay, Kola, Hullman, & Munson, 2016), and bivariate classification (Howard & MacEachren, 1996). These visualization techniques have mostly been static and the temporal resolution of the uncertainty was seldom seen as being mission critical. The emerging use of heart rate and neurophysiologic sensors during surgery points to the need for more dynamic representations of uncertainty to match the speeds of real-time environments (Zenati et al., 2018; Stevens & Galloway, 2019).

Functionally, uncertainty arises when outcomes do not match our expectations. At the millisecond level this occurs all the time as we make small adjustments to our fingers when typing or to our eyes as we scan text, but most of the time these small adjustments go unnoticed. More complex decisions like those of a surgeon require more cognitive activity than rapid implicit responses.

As the hierarchy of uncertainty is ascended from the neuronal level, the multiple dimensions of its effects begin to be seen. When it becomes observable and can be deliberately acted on, it is termed hesitation; hesitation is thinking about what you do with uncertainty. While the hesitation to shoot of law enforcement agents, i.e. cautious hesitation, is perhaps the most visible example of controlled uncertainty, hesitation can be experienced anytime there is uncertainty.

While uncertainty might be observed as a pause, it is not always clear what this pause might mean. For instance, hesitation in surgical residents might manifest during operations as a pause while they try to search for more information, or as they try to disguise their uncertainty and hesitation in fear of looking incompetent (Ott, Schwartz, Goldsmith, Bordage & Lingard, 2018). Expert surgeons experiencing uncertainty might deliberately pause while they 'slow the pace down' as part of reflective awareness (Hale, Terrien, Quirk, Sullivan & Cahan, 2018), while novices might freeze or lose interest. In the extreme, chronic uncertainty may evolve into the disease entity anxiety (Grupe & Nitsche, 2013).

These data point to the hierarchical structure of the brain, decision making, and uncertainty. If uncertainty cannot be resolved at a particular (temporal and spatial) level(s) of the brain, the prediction errors drive the adjustment of the mental models or sensory inputs to higher levels by drawing on more complex representations of environmental states and richer cognitive models (Kiebel, Daunizeau & Friston, 2008). This allows access to, and accumulation of, a broader range of evidence over extended temporal scales to find the best explanation for the evidence. Engaging each additional level, however incurs a 'cost of control' and is equivalent to inference under a more complex model or a model that includes more variables (Zenon, Solopchuk & Pezzzulo, 2020).

Despite its ubiquity and our ability to define it, uncertainty has been difficult to track during real-world teamwork. This is due, in part, to uncertainty being based on each person's perceptions where sensory information is organized, identified and interpreted to represent and understand the environment (Roth, 2009; Hong, 2010); what might be highly uncertain for one individual may be unimportant or not even perceived by another. Perceptions are intangible and personal but the neurodynamics of uncertainty that is raised by them will involve some form of persistent mental states where evidence is weighed and mental models refined; these persistent states are signs of uncertainty (Friston, 2010).

We have been refining EEG-based analytics for the spatial and temporal characterization of neurodynamic organizations that parallel periods of team uncertainty (Stevens & Galloway, 2017; Stevens, Galloway, Halpin & Willemsen-Dunlap, 2016). Of the myriad of time frames between the implicit firing of neurons and periods of anxiety we have chosen to develop solutions for quantitatively measuring uncertainty on a second-by-second basis using 60s moving average windows. This time frame is greater than the 100-300 ms neurodynamics associated with everyday processes like conversations and non-verbal interactions (Sanger, Muller & Lindenberger, 2012; Stephens, Silbert & Hasson, 2010; Caetano, Jousmaki & Hari, 2007), and is optimized to identify periods of persistent uncertainty that could be mitigated if detected early.

These persistent neurodynamic states, which are close proxies for situational and verbal uncertainty (Stevens & Galloway, 2017; 2019), are present in team members as well as teams (Stevens, Galloway & Willemsen-Dunlap, 2018). They occur during submarine navigation and healthcare training and are also present during live patient surgery (Stevens, Galloway & Willemsen-Dunlap, 2019). They can be quantitatively reported as brain-wide averages and used for comparing across individuals, teams, or training protocols, and systematically mapped to different brain regions and frequencies to identify cognitive elements involved in resolving the uncertainty.

In this chapter we provide a hierarchical framework for visualizing neurodynamic uncertainty to determine what might make sense to instructors and teams in terms of actionable understanding of the frequency, magnitude and duration of uncertainty, and what this might mean for training and theory.

Methods

Task

Submarine Piloting and Navigation (SPAN) simulations are part of the standard training for Junior Officers in the Submarine Officer Advanced Candidacy course at the US Navy Submarine School, and for experienced teams prior to deployment. For this study we chose a high stakes training task performed by an experienced submarine navigation team practicing a distance port entry prior to deployment. The particular task involved establishing and maintaining the ship's position while in close proximity to other ships and the land. A three-minute repeating procedure, described below, culminates with the triangulation of visual markers, water depth, and radar and Global Positioning System (GPS) readings.

SPAN sessions contained three training segments: Briefing; Scenario; and Debriefing. The Briefing is where the team reviewed the environmental conditions and other ships in the area, and statically established the submarine's position. The Scenario is the training part of the navigation simulation where events included: encounters with approaching ships, the need to avoid shoals, changing weather conditions and instrument failure. One team process in the Scenario required updating the ship's position every three minutes. In this process called 'Rounds', three navigation landmarks are chosen and their visual or electronic bearings from the boat are measured and plotted on a chart. The regular 'Rounds' sequence usually begins with a '1 min to Next Round' announcement, followed by a 'Mark the Round' call 60s later. The 'Mark Round' segment is where individual taskwork like reading the fathometer or radar becomes teamwork as the information is shared. The observations are made, verified with the estimated position and depth of the water, and then the call to 'End Round' is made. The Debriefing was an after-action review where all team members participated in critical performance discussions.

The experienced team in this study was practicing piloting a submarine into port prior to re-deployment. The four crew members fitted with EEG headsets were the Navigator (NV); the Officer on Deck (OD); the Contact Manager (CM) who kept track of other ship traffic and the Quartermaster (QM) who maintained the ship's positon (Other people were "satellite" team members but were not directly involved in the team processes analyzed here).

Subjects

Informed consent protocols were approved by the Biomedical IRB, San Diego, CA (Protocol EEG01), and the US Navy Medical Review Board for the collection of EEG from the submarine navigation team. All participating subjects consented to participate with written approval, and to allow their images and speech to be available for additional analysis. To maintain confidentially, each subject was assigned a unique number known only to the investigators of the study, and subject identities were not shared. This design complies with DHHS: protected human subject 45 CFR 46; FDA: informed consent 21 CFR 50.

Modeling Neurodynamic Organizations and Information

The data acquisition began shortly after the EEG sensors were adjusted for good contact (< 10 Ω). Team member data streams were synchronized with electronic markers inserted into the EEG data streams. The recorded EEG data was processed using either the Matlab®-based FieldTrip® toolbox (Oostenveld, Fries, Maris & Schoffelen, 2011), as detailed previously (Stevens, Galloway, Halpin & Willemsen-Dunlap, 2016; Stevens & Galloway, 2017) or NeuroPype software. Commercial EEG headsets with both dry and wet electrodes have been used from multiple vendors, and with the number of sensors ranging from five to nineteen. A greater number of sensors allows more detailed analysis of the spatial locations of the sources of uncertainty in the brain.

Our modeling goal was to develop systems that would allow neurodynamic measures to be made for each team member at a 1Hz resolution, and provide quantitative comparisons across sensor sites (i.e. the occipital lobe vs the motor cortex) and across the 1-40 Hz frequency spectrum (from delta to gamma bands) within each sensor's data stream. The data for each team member were combined to provide the quantitative contributions of each team member's dynamics to the team dynamics. Finally the system should operate with teams with two to six team members. The resulting modeling scheme produces a large number of parallel neurodynamic data streams (*NDS*) (i.e. 1 team x 3 team members x 19 sensors x 40 frequency bands x other behavioral and physiologic variables) all of which can be quantitatively related to each other.

Neurodynamic organizations were detected and modeled by symbolically transforming the physical units of each team member's EEG power into bits of information (Figure 1) (Stevens & Galloway, 2015; 2016; 2017). The modeling began by categorizing the EEG levels each second into high, medium and low power levels and assigning them the symbols 3, 1, and -1; these could be any symbol, and there could be any number of categories, but these allow easy visualization. Thus, the entire performance of any team member can be described by several thousand parallel data streams of -1, 1, or 3's (Figure 1B).

The data from each team member (shown for a dyad in Figure 1A) can be combined to create a composite symbol that represents the state of the team. With 2 persons and 3 states, the number of possible symbols is 9, and the team symbolic history consists of thousands of parallel streams of the symbols 1 to 9. There is little information in the 1s expression of a symbol, but when the entropy of symbol expression is modeled over windows of 60s, long-term organizational trends begin to develop. This provides the measure of uncertainty that becomes bounded at the upper level by the maximum entropy for the number of symbols and 0 which is the entropy for a single symbol.

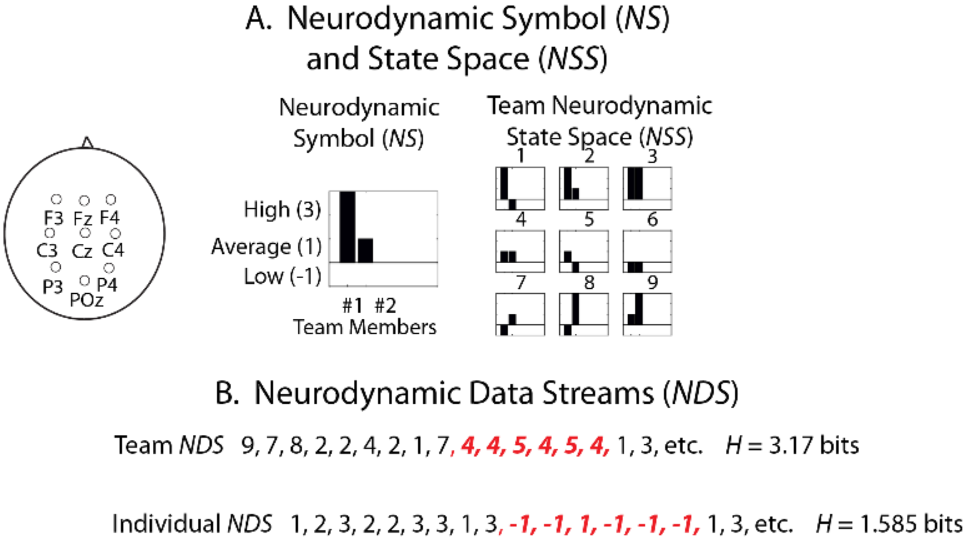

A. Neurodynamic Symbol (*NS*) and State Space (*NSS*)
Neurodynamic Symbol (*NS*)
Team Neurodynamic State Space (*NSS*)
F3 Fz F4
C3 Cz C4
P3 P4
POz
High (3)
Average (1)
Low (-1)
#1 #2
Team Members
1
2
3
4
5
6
7
8
9
B. Neurodynamic Data Streams (*NDS*)
Team *NDS* 9, 7, 8, 2, 2, 4, 2, 1, 7, 4, 4, 5, 4, 5, 4, 1, 3, etc. *H* = 3.17 bits
Individual *NDS* 1, 2, 3, 2, 2, 3, 3, 1, 3, -1, -1, 1, -1, -1, -1, 1, 3, etc. *H* = 1.585 bits

Figure 1. Team and individual neurodynamic modeling of a dyad. A) A sample Neurodynamic Symbol (*NS)* showing a 1s period where the EEG power was high for team member 1, and average for team member 2. Shown next to this symbol is the nine-symbol Neurodynamic State Space (*NSS)* for two persons with three EEG power levels. B) Neurodynamic Data Streams (*NDS*) are symbol sequences that span the performance. For a dyad they are the symbols in Figure 1A. For team members they are the -1, 1 and 3 values used symbolically. Note that the symbol expression for both team and individual *NDS* is not stochastic but is punctuated by periods of symbol repeats of various lengths which represent neurodynamic organization in response to task events.

Results

Neurodynamic Uncertainty Occurs in Temporal Chunks

The hierarchical dissection of the neural dynamics of a dyad into increasingly lower (i.e. faster) temporal and spatial scales is shown in Figure 2. This progresses from a two-person team, to the team members, to the scalp locations and frequency bands of each person, to the EEG micro-volts of power in each frequency band. This hierarchal modeling parallels the way information flows in the brain as it represents the complexity in the environment (Kiebel, Daunizeau & Friston, 2008; Flack, 2017).

The dyad was performing the Map Navigation Task where one person termed the Giver (G) guides a second person, the Follower (F) in drawing a line through landmarks on a computer map using only speech (Stevens & Galloway, 2015; 2016; 2017). During this Map Task performance the team was exchanging information and F was drawing paths around landmarks with the mouse. At one point F experienced drawing difficulties and began clicking rapidly with the mouse (Figure 2B) as he became frustrated; he also uttered comments like 'it's so hard', and 'ooooooo' as the *NI* levels rose. The *NI* patterns first increased in the parietal region and then transited to the pre-motor / motor region (C3 and then C4 sensors). Spectrum-wide *NI* deconstruction showed the maximum increase in the 18 and 22 Hz EEG bands (Figure 2C), with peak levels at18 Hz that were enriched fifteen times from the average of the Follower's *NI* (Fig. 2A). The 0.46 bits of information at the peak was ~30% of the theoretical maximum level for 3 symbols.

Lastly, a moving average of the numerically-calculated EEG power was calculated for this 18 Hz segment. The bar graph in the lower part of Figure 2D was below the performance average value of 1.0 bits with a mean of .26 bits. These results indicate that the elevated *NI* associated with this period of uncertainty was a reflection of a persistent deactivation of Mu rhythms that occurs when a person visualizes or performs a movement (Caetano, Jousmaki & Hari, 2007; Tognoli & Kelso, 2015).

In summary:

- The neurodynamics of the team can be quantitatively dissected into those of each team member and shared information.
- Neurodynamic organization is not continuous but occurs in chunks associated with periods of uncertainty or stress.
- Increased neurodynamic organization can result from either suppressed or elevated EEG power.
- The brain regions showing increased *NI* can be linked to plausible causes for the uncertainty, motor control for this example.

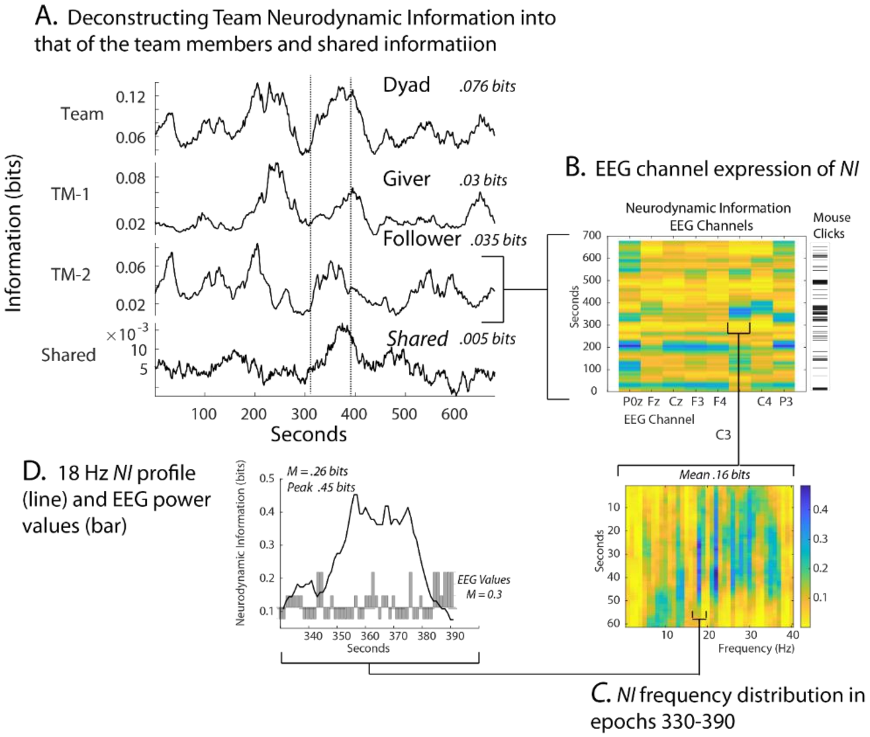


Figure 2. Quantitative Decoding of Team Neurodynamics. (A) The information in the NDS of the dyad, and each member, is displayed for a MT performance. (B) The NI dynamics at each sensor of the Follower; the bar to the right shows F's mouse clicks. (C) The elevated NI at the C3 sensor is expanded for one segment (epoch 330-390) and displayed across the 1-40 Hz EEG frequency spectrum. (D) A profile plot of NI in the 18 Hz frequency band is overlaid with a bar plot of the -1, 1, and 3 EEG power values.

Uncertainty during Submarine Piloting and Navigation

There were three goals for a similar dissection of the performance of a submarine navigation team. First, by using a high fidelity complex task, could we expect to obtain the same discrete dynamics that were observed in Figure 2 performed by Advanced Placement students? Second, were the peaks of *NI* associated with periods of possible uncertainty? Three, how far can the temporal scale of neurodynamics be reduced before the hierarchical decomposition of neurodynaic organization no longer relates to behavior? That is, is there a point on the micro-macro scale of teamwork dynamics where bits of information no longer give a complete description of the neural dynamics? The submarine navigation task represents a 'best' case of high fidelity data that could be obtained in complex situations in the sense that this was an expert team that was engaged in a meaningful and consequential task. The Briefing section at the beginning and Debriefing section as the end of the performance are not shown to highlight the countdown sequences through the last minute of five Mark Round events (Figure 3A).

The dynamics of the neurodynamic information are shown each second for the Navigator, Quartermaster and Contact Manager in Figure 3 B-D. Discrete peaks were seen for the three team members in the proximity of the 1 minute countdown segment of Rounds. The expression of the *NI* peaks of the three team members were coordinated, but not synchronous, with *NI* correlations of 0.36 for the Nav and the QM, .3 between the NAV and CM, and .34 between the QM and CC, all with p values < 0.01.

The discrete nature of the peaks during the Rounds sequences made it possible to estimate the magnitude and duration of the periods of neurodynamic organization associated with the Rounds sequence. These estimates were made using the findpeaks.m function in *Matlab* and the measures were made at half peak prominence. The magnitude was 0.23 ± .05 bits while the duration was 26.3 ± 14s.

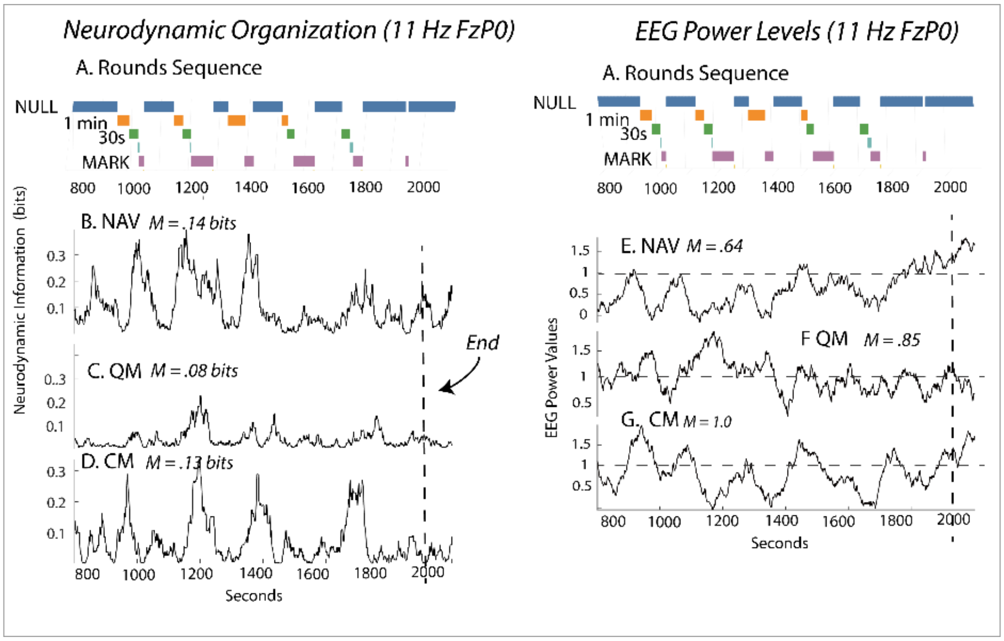


Figure 3. Visualizing neurodynamic organizations (left) and EEG power values (right) for a submarine navigation team during a port entry simulation. A) The countdowns for the last minute of the five Rounds sequences are plotted vs. time. B. C, D) These figures plot the temporal neurodynamic information (NI) values for the Navigator (NAV) Quartermaster (QM) and the Contact Manager (CM). E, F, G). These figures plot a centered 60s moving average window of the EEG power values of the 11 Hz frequency. All values are from the FzP0 montage of each person.

The data in Figure 2 suggested that while F was struggling with the mouse there was suppression, not activation of the mu EEG rhythm. To study similar suppressive and activation effects on the alpha band components during Rounds we also determined the fluctuating dynamics of EEG power levels (EEG-PV) for the three team members. The *EEG-PV* were modeled by using the -1, 1, and 3 symbols numerically and then aligning with the *NI* using a 60 s moving average (Figures 3 E, F, & G). The resulting peaks fluctuated slightly below the mean value of 1, and illustrated sustained periods of both suppression and activation. Visually the peaks were less discrete, and aligned poorly with the peaks of *NI.* The correlations between *NI* and EEG-PV averaged -0.06 for the three team members.

These results point to a fundamental difference between *NI* and *EEG-PV*: Periods of both EEG activation and suppression will result in an increase in organization, i.e. *NI*. To the extent that *NI* correlates with military and healthcare expertise while *EEG-PV* does not (Stevens, Galloway, Lamb, Steed, & Lamb, 2017), this suggests that the *EEG-PV – NI* junction is an important point where the analysis of the micro and macrodynamics of teams diverges. The implication for neurodynamic data selection when contemplating training Artificial Intelligence (AI) systems is that observationally relevant comparisons between teams or team members would be better made using measures of organization, than by using EEG power.

Undoubtedly there are myriads of neurophysiologic processes occurring during the minutes leading up until the next Rounds sequence which reflect the execution of episodic routines for which the team members were trained; like reading charts, instruments, making calculations, etc. The lack of these dynamics in Figure 3 most likely reflects: 1) the rapid nature of low-level dynamics (Tognoli & Kelso, 2015), 2) the frequency bands of the Rounds neurodynamics which is restricted to 10-11 Hz and mainly present in the FzP0 channel, and 3) the lack of unexpected uncertainty as they are likely executed almost automatically as episodes of action-control sequences (Cooper & Shallice, 2000; Schneider & Logan, 2015), and 4) the experience of the team as described previously.

Discussion

The 'principle of progress' is a valued tenant of surgical training where the idea of 'always moving forward' in emphasized during surgery. While this tenant is referenced most often with regard to healthcare, the same idea is expressed during navigation as 'looking down the road.' High stakes teams instinctively know when progress is being delayed, and that pauses or hesitation are early indicators of potential disruption.

The features of uncertainty that surround the principle of progress are its frequency, magnitude, and duration. The frequency of encountering uncertain events is a factor of experience (Kennedy, Regehr, Baker & Lingard, 2005; Stevens, Galloway, Lamb, Steed, Lamb, 2017; Stevens, Galloway & Willemsen-Dunlap, 2016). The magnitude of a period of uncertainty reflects the costs of cognition and how much the search for the information required for its resolution depends on higher levels of cognition (Alexandre, Oleg & Giovanni, 2020). In our experience the third dynamical component, duration, is the black box. How long a period of uncertainty will last is the most difficult to predict, yet, this is the measure most likely to be important for training purposes. Peak estimations like those calculated for the submarine team members in Figure 3 and for members of healthcare teams including a live patient surgery (Stevens, Galloway & Willemsen-Dunlap, 2019) are showing that periods of (expected) uncertainty are generally resolved in less than 1 min. The duration of unexpected uncertainty like that shown in Figure 2 may not follow that trend, and unchecked may lead to a freeze (Ott et al., 2018).

To address the duration problem we recently trained neural networks to recognize the neural correlates of uncertainty during teamwork (Stevens & Galloway, 2019). Although the training set for these AI models was derived from Map Task performances like those in Figures 2 and 3, the profiles of uncertainty learned have made accurate classifications of periods of uncertainty associated with a near collision of a submarine with another ship during a submarine navigation simulation (Stevens & Galloway, 2019). This domain-neutral capability points to human uncertainty having recognizable features and possibly structures that team members might revisit, i.e. people might have learned ways of working themselves out of uncertain situations.

There are three immediate applications for these technologies. While all three could benefit from traditional event timelines, none of them requires them. In fact, these technologies may allow the creation of a new form of event maps based on a person's / teams cognitive reactions.

- The first is a tool to validate and improve Intelligent Tutoring Systems (ITSs). A student training with an ITS should not experience uncertainty, or at least not more than the designer wished the student to experience. To the extent that the frequency, magnitude or duration of uncertainty exceeds these limits, either for the whole performance or specific events, it might suggest modifications to the simulation or feedback.

- The second is a support for after-action reviews. The immediate availability of maps like those in Figure 3 could direct the instructor, student, or other team members to these periods for discussion.

- The third is as a real-time predictive tool. We recently developed transition models from 155,000 AI state transitions that provided estimates of up to 20s in the future of the probability of the uncertainty increasing, decreasing, or remaining the same. This 20s predictive window when used with real-time EEG systems would enable ITSs to consider a future range of options, based on the predicted ability of the student to resolve the immediate difficulty.

These multi-domain tools will be relevant to work on assessing and supporting collaborative-problem solving in the military, healthcare, K-12 environments and the workforce (Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2016; Zhu & Andrews-Todd, 2019). The scale afforded by the AI tools would allow the assessment of collaborative problem-solving in large-scale assessments such as the OECD Program for International Student Assessment (PISA) test (Graesser et al., 2019). It is likely that new visualizations would be needed to encompass this scale.

Recommendations and Future Research

The demonstration of neurodynamic signatures of the complex Rounds activity during submarine navigation, while seemingly a simple accomplishment, has implications for the design of the next generation of simulations, for training, and for extending our understanding of teamwork. It shows that highly complex brain dynamics linked to periods of uncertainty, can be decomposed into discrete peaks whose frequency, magnitude and duration can be estimated and linked to neural oscillations and brain regions. It now becomes more likely that similar neural signatures, linked to uncertainty, can be identified for other extended tasks teams are asked to perform, like ventilating a patient during surgery for example (Stevens, Galloway & Willemsen-Dunlap, 2019).

These properties provide a language that AI agents can understand and can be trained to recognize, providing automated understandings about when individuals are becoming uncertain, how long the uncertainty will last, and what was it about the triggering event that caused the uncertainty. Real-time reporting of this information in ITSs like the team training Surveillance Scenario Tutor (Gilbert et al., 2018), might provide a useful cognitive perspective of new scenarios during development, as well as neurodynamic profiles of uncertainty during training.

References

Aggarwal, C. C. (2009). Trio a system for data uncertainty and lineage. In *Managing and Mining Uncertain Data* (pp. 1-35). Springer, Boston, MA.

Alexandre, Z., Oleg, S., & Giovanni, P. (2020). An information-theoretic perspective on the costs of cognition. bioRxiv

Brodlie, K., Osorio, R.A., & Lopes, A., (2012). A review of uncertainty in data visualization. In *Expanding the Frontiers Visual Analytics and Visualization*, J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong, Eds. London, U.K. : Springer. 81–109.

Caetano, G., Jousmaki, V., & Hari, R. (2007). Actor's and observers primary motor cortices stabilize similarly after seen or heard motor actions. *Proc. Natl. Acad. Sci. USA, 104,* 9058–9062.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *158*(3), 419-444.

Cooper R, & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cogn Neuropsychology 17*:297–338.

Cox, J., House, D., & Lindell, M. (2013). Visualizing uncertainty in predicted hurricane tracks. *International Journal for Uncertainty Quantification*, *3*(2).

Demmans Epp, C., & Bull, S. (2015). Uncertainty Representation in Visualizations of Learning Analytics for Learners: Current Approaches and Opportunities. *in IEEE Transactions on Learning Technologies*, 8 (3). 242-260.

Flack, J. (2017). Life's information hierarchy. In S. I. Walker, P. C.W. Davies, & G. F. R. Ellis (Eds.), From matter to life: Information and causality (pp. 283–302). Cambridge: Cambridge University Press.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci. 11*, 127-138.

Gilbert, S. Slavina, A., Dorneich, M., Sinatra, A., Bonner, D., Johnston, J., Holub, J., MacAllister, A., & Winer, E. (2018). Creating a team tutor using GIFT. *International Journal of Artificial Intelligence in Education 28*, 286-313.

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, 19(2), 59–92.

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience, 14*, 488-501.

Hale, J. F., Terrien, J. M., Quirk, M., Sullivan, K. & Cahan, M. (2018). The impact of deliberate reflection with WISE-MD™ modules on surgical clerkship students' critical thinking: a prospective, randomized controlled pilot study. *Adv. Med. Educ Pract*, *9*, 757- 766.

Hodgkinson, G. P., Bown, N. J., Maule, A. J., Glaister, K. W., & Pearman, A. D. (1999). Breaking the frame: An analysis of strategic cognition and decision making under uncertainty. *Strategic management journal*, *20*(10), 977-985.

Hong, S. L., (2010). The entropy conservation principle: Application in ergonomics and human factors. *Nonlinear Dynamics, Psychology and Life Sciences, 14*, 291-313.

Howard, D., & MacEachren, A. M. (1996). Interface design for geographic visualization: Tools for representing reliability. *Cartography and Geographic Information Systems*, *23*(2), 59-77.

Hullman, J., Qiao, X., Correll, M., Kale, A., & Kay, M. (2019). In pursuit of error: A survey of uncertainty visualization evaluation. *In IEEE transactions on visualization and computer graphics*, *25*(1). 903-913.

Kay, M., Kola, T., Hullman, J. R., & Munson, S. A. (2016, May). When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5092-5103).

Kennedy, T.J., Regehr, G., Baker, G.R., & Lingard, L.A. (2005). Progressive independence in clinical training: a tradition worth defending? *Acad. Med.*, *80*(10), 106-111.

Kieblel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarch of time-scales in the brain. *PLOS Computational Biology, 4, 11*,e1000209.

Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, D. (2016). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds). Handbook of Research on Computational Tools for Real-World Skill Development. Hershey, PA: *IGI-Global.* 344-359.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data, *Computational Intelligence & Neuroscience, 2011*, Article ID 156869, 9 pages.

Ott, M., Schwartz, A., Goldsmith, M., Bordage, G., & Lingard, L. (2018). Resident hesitation in the operating room: does uncertainty equal incompetence? *Surgical Training, Medical Education, 52,* 851:860.

Roth, R. E. (2009). A qualitative approach to understanding the role of geographic information uncertainty during decision making. *Cartography and Geographic Information Science*, *36*(4), 315-330.

Sänger, J., Müller, V., & Lindenberger, U. (2012). Intra- and interbrain synchronization and network properties when playing guitar in duets. *Frontiers in Human Neuroscience, 6*(312).

Schneider, D.W. & Logan, G. D. (2006). Hierarchical control of cognitive processes: Switching tasks in sequences. *J Exp Psychol Gen, 135*, 623-640.

Stephens, G., Silbert, L., & Hasson, U. (2010). Speaker- listener neural coupling underlies successful communication. Proceedings from: *National Academy of Science, USA* Retrieved from www.pnas.org/cgi/doi/ 10.1073/pnas. 1008662107.

Stevens, R. H., & Galloway, T. (2015). Modeling the neurodynamic organizations and interactions of teams. *Social Neuroscience*, *11*, 123-139.

Stevens, R. H., & Galloway, T. (2016). Tracing neurodynamic information flows during teamwork. *Nonlinear Dynamics, Psychology and Life Sciences, 20* (2), 271-292.

Stevens, R., & Galloway, T. (2017). Are neurodynamic organizations a fundamental property of teamwork? *Frontiers in Psychology*, *May, 2017*.

Stevens, R. H., Galloway, T., Halpin, D., & Willemsen-Dunlap, A. (2016). Healthcare teams neurodyamically reorganize when resolving uncertainty. *Entropy, 18,* 427.

Stevens, R., Galloway, T., Lamb, J., Steed, R., & Lamb, C. (2017). Linking team neurodynamic organizations with observational ratings of team performance. In *Innovative Assessment of Collaboration* (pp. 315-330). A. A. Von Davier, P. C. Kyllonen, & M. Zhu eds. Springer International Publishing, Cham, Switzerland.

Stevens, R., Galloway, T. L., & Willemsen-Dunlap, A. (2019). Advancing our understandings of healthcare team dynamics from the simulation room to the operating room: A neurodynamic perspective. *Frontiers in Psychology*, *10*, 1660.

Stevens, R., & Galloway, T. L. (2019). Teaching machines to recognize neurodynamic correlates of team and team member uncertainty. *Journal of Cognitive Engineering and Decision Making, 13*, 310-327.

Thomson, J., Hetzler, E., MacEachren, A., Gahegan, M., & Pavel, M. (2005, March). A typology for visualizing uncertainty. In *Visualization and Data Analysis 2005* (Vol. 5669, pp. 146-157). International Society for Optics and Photonics.

Tognoli, E., & Kelso, J. A. (2015). The coordination dynamics of social neuromarkers. arXiv preprint arXiv:1310.7275.

Zapata-Rivera, D., Zwick, R., & Vezzu, M., (2016). Exploring the Effectiveness of a Measurement Error Tutorial in Helping Teachers Understand Score Report Results. *Educational Assessment, 21* (3), 215-229,

Zenati, M. A., Leissner, K. B., Zorca, S., Kennedy-Metz, L., Yule, S. J., & Dias, R. D. (2018). First reported use of team cognitive workload for root cause analysis in cardiac surgery. *Semin Thoracic Surg*. Doi: https://doi.org/10.1053/j.semtcvs.2018.12.00.

Zénon A, Solopchuk O, Pezzulo G. An information-theoretic perspective on the costs of cognition. *Neuropsychologia*. 2018; 123: pp.5-18. doi:10.1016/j.neuropsychologia.2018.09.013

Zhu, M., & Andrews-Todd, J. (2019). Understanding the Connections of Collaborative Problem-Solving Skills in a Simulation-based Task through Network Analysis. *In the Proceedings of the International Conference on Computer Supported Collaborative Learning (CSCL 2019)*. Lyon, France

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.

CHAPTER 12 – PEDAGOGICAL USE SCENARIOS FOR DATA VISUALIZATIONS: PREPARE, CONDUCT AND EVALUATE

Morten Misfeldt[1], Zachari Swiecki[2], Diego Zapata-Rivera[3], and Xiangen Hu[4]
University of Copenhagen[1], University of Wisconsin Madison[2], Educational Testing Service[3],
University of Memphis[4]

Introduction: Visualizing Data for Pedagogical Situations

Data visualization for pedagogical purposes is a topic of increasing relevance and interest due to the many ways in which the educational sector is supported by new digital infrastructures. As Intelligent Tutoring Systems (ITSs) and more traditional educational systems move closer to one another, we anticipate an increased blending of face-to-face, mediated, and automated teaching. In this chapter, we address two key problems in relation to this blending: (a) the difficulties teachers have using data to meaningfully engage with students working in digital environments and (b) how the larger infrastructure of standards for teacher work and student learning impact pedagogical interactions.

We focus on the situation of teachers and adopt a dialectical interaction between them and resources: teachers' work may modify or influence the resource used, but, conversely, the resource also influences the teachers' work. The documentational genesis thereby foregrounds the interactions between teachers and resources, with a particular focus on how both teachers and resources are transformed by their interactions with one another (Misfeldt, Tamborg, Dreyøe, & Allsopp, 2019). In particular, we focus on data visualizations as one such resource, where data can consist of cognitive signs of learning, interaction footprints, information about learning standards, and so on.

Using this approach, we consider pedagogical data visualization as a bi-directional interaction with the teaching and learning processes: the visualization is affected by teachers and by student learning in ways that go beyond the intentions of the designers of the visualization. The visualization can also affect teaching and learning situations in unforeseen ways. Visualizing data hence has pedagogical and cognitive consequences (Zapata-Rivera, Graesser, Kay, Hu, & Ososky, 2020, this volume).

To illustrate these consequences and their relevance for technology-enhanced learning and ITS in particular, we describe two examples of the interaction between teachers and pedagogical visualizations, one in the context of online educational simulations, the other in the context of the Danish mathematics curriculum. We begin with a discussion of teacher action in pedagogical situations before turning to the examples.

Prepare, Conduct and Evaluate: A Framework for Understanding Pedagogical Situations

Following Sherin and Drake (2009), we consider pedagogical situations in terms of three phases: before, during, and after instruction. Each of these phases have particular processes associated with them. For example, before instruction, teachers may read their curriculum in order to plan their instructional strategies. Here, curriculum refers broadly to the resources or materials teachers use to guide their instruction, which traditionally has included learning standards, texts, exercises and so on, but have increasingly been scaffolded by technologies like automated tutors and immersive digital environments. During instruction, teachers enact this planned curriculum. But as Remillard (2005) argues, enactment is not simply delivery of pedagogical content, but an interaction between the content and the teacher in which the teacher (re)designs the content in the moment. Finally, after instruction is complete, teachers judge the enacted curriculum either from the perspective of the students or the teacher in order to adapt it for future use.

These theories of curriculum use highlight that teachers *prepare* their teaching, *conduct* their teaching, and *evaluate* their teaching. While prepare and conduct align naturally with before and during instruction, as Sherin and Drake (2009) argue, the process of evaluation can occur before, during, or after instruction. For example, teachers may evaluate their own understanding of the curriculum prior to teaching, they may evaluate student understanding during the instruction, and evaluate student understanding and their own teaching after instruction. While we acknowledge this complexity, to simplify our analysis, we associate evaluation with teacher action that occurs after the instruction, and assume that prepare and conduct may entail teacher judgments regarding their own understanding prior to instruction and judgments of student understanding or their own performance during instruction.

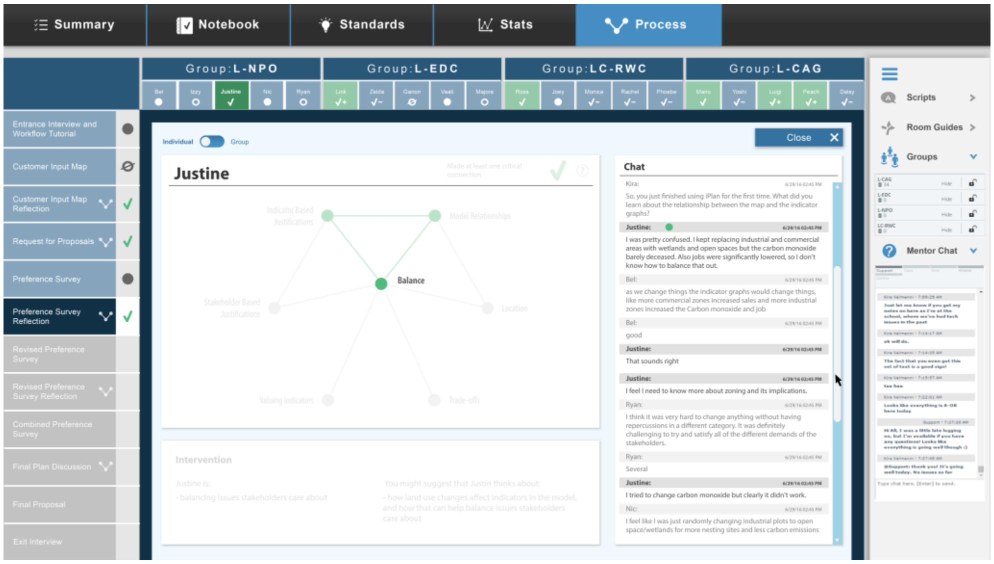
In what follows, we use prepare, conduct, and evaluate (PCE) as a framework for investigating how teachers interact with two pedagogical visualizations, the consequences of these interaction, and the implications they have for technology-enhanced learning.

Example 1: The Process Tab

The first visualization was designed to support teachers' use of *virtual internships* in their classrooms. Virtual internships are online simulations of professional practices in which students work to solve complex and ill-defined problems (Shaffer, 2006). For example, in the virtual internship *Land Science* students work in teams to develop a re-zoning plan for a city. During the internship, teams conduct research, model the effects of land-use changes on socioeconomic and environmental indicators, and create a proposal to argue for their plan.

Teams communicate in virtual internships using a chat tool built into the online platform. At several points during the internship, teams are asked to reflect on a substantive task they have just completed. These reflective discussions have several key topics associated with them that high quality discussions are supposed to cover. The discussions are co-managed by automated rules and human mentors (distinct from the teacher) that direct the flow of the discussion. Teachers in this context mainly fulfill a facilitator role, monitoring discussion quality and intervening only if necessary. However, given that there are typically five teams of five to six students each having discussions simultaneously, it can be difficult for teachers to monitor their quality and intervene in real-time.

To address this issue, we developed the *Process Tab*, an interface that represents the content and quality of team discussions in virtual internships to teachers in real-time. To assess discussion quality, the interface uses regular expression matching to identify key topics in the chat and Epistemic Network Analysis (ENA) (Shaffer, Collier, & Ruis, 2016) to identify the connections (i.e., co-occurrences) between topics that are indicative of high quality discussions. The Process Tab is part of a larger suite of tools for virtual internships that support various administrative and pedagogical tasks of the teacher.



Intervention

Justine is:
- balancing issues stakeholders care about

You might suggest that Justin thinks about:
- how land use changes affect indicators in the model, and how that can help balance issues stakeholders care about

Chat

Kira: 6/29/16 02:45 PM
So, you just finished using iPlan for the first time. What did you learn about the relationship between the map and the indicator graphs?

Justine: 6/29/16 02:45 PM
I was pretty confused. I kept replacing industrial and commercial areas with wetlands and open spaces but the carbon monoxide barely deceased. Also jobs were significantly lowered, so I don't know how to balance that out.

Bel: 6/29/16 02:45 PM
as we change things the indicator graphs would change things, like more commercial zones increased sales and more industrial zones increased the Carbon monoxide and job

Bel: 6/29/16 02:45 PM
good

Justine: 6/29/16 02:45 PM
That sounds right

Justine: 6/29/16 02:45 PM
I feel I need to know more about zoning and its implications.

Ryan: 6/29/16 02:45 PM
I think it was very hard to change anything without having repercussions in a different category. It was definitely challenging to try and satisfy all of the different demands of the stakeholders.

Ryan: 6/29/16 02:45 PM
Several

Justine: 6/29/16 02:45 PM
I tried to change carbon monoxide but clearly it didn't work.

Nic: 6/29/16 02:45 PM
I feel like I was just randomly changing industrial plots to open space/wetlands for more nesting sites and less carbon emissions

Kira Valmanni · 7:09:29 AM
Just let me know if you get my notes on here as I'm at the school, where we've had tech issues in the past

Kira Valmanni · 7:14:17 AM
ok will do.

Kira Valmanni · 7:14:35 AM
The fact that you even got this set of text is a good sign!

Kira Valmanni · 7:15:57 AM
tee hee

Kira Valmanni · 7:22:01 AM
Looks like everything is A-OK here today

Support · 7:27:28 AM
Hi All, I was a little late logging on, but I'm available if you have any questions! Looks like everything is going well though :)

Kira Valmanni · 7:27:49 AM
@Support: thank you! It's going well today. No issues so far

Figure 1. Network view of the Process Tab

The Process Tab (Figure 1) includes (a) a high-level summary of the quality of each student's contributions to the discussions throughout the internship, (b) simplified network models representing the connections between key topics teams or individuals made during their discussions, (c) a complete record of the chats sent in each discussion, and (d) suggested teacher interventions based on what was said in the discussions. Green connections in the network models indicate those connections associated with high quality discussions. Teachers can interact with the networks to see which chats contributed to the connections. They can also scroll through the chat record to see how and when the connections occurred over time.

These features of the interface were designed in collaboration with teachers and were intended to help them conduct their teaching in the moment. Specifically, we hypothesized that the high level summaries and network models would help them to monitor discussion quality and that the intervention and scrolling features would scaffold their ability to provide feedback in the moment. However, after interviewing teachers who used the Process Tab as part of their curriculum, we found that their use of the tool was constrained by their in-person responsibilities as facilitators of the class. In particular, teachers said that they were able to use the Process Tab to point their attention to particular individuals/teams and provide feedback, but their attention was more often directed to classroom management tasks, such as answering questions, troubleshooting, and maintaining student engagement that prevented them from using the interface as much as they had hoped (Herder et al., 2018). These findings align with research in the area of Open Learner Models (OLM) that suggests the need to take into account teachers' responsibilities and operational constraints when designing OLM (Bull & Kay, 2016; Zapata-Rivera, Hansen, Shute, Underwood, & Bauer, 2007).

Example 2: The Goal Arrow

The second example relates to learning administration software for Danish mathematics teachers. The Danish curriculum consists of competencies described as pairs of knowledge and skills that the students should be able to do after the teaching has ended. The curriculum for Mathematics consists of four overall areas (1) Mathematical Competencies, (2) Numbers and Algebra, (3) Geometry and Measures, and (4) Statistics and Probability. The many learning objectives together with an increased focus on documental work of teachers has led to the development of several new platforms that support documentation, preparation and the actual conduction of teaching (see Misfeldt et al., 2019 for further description).

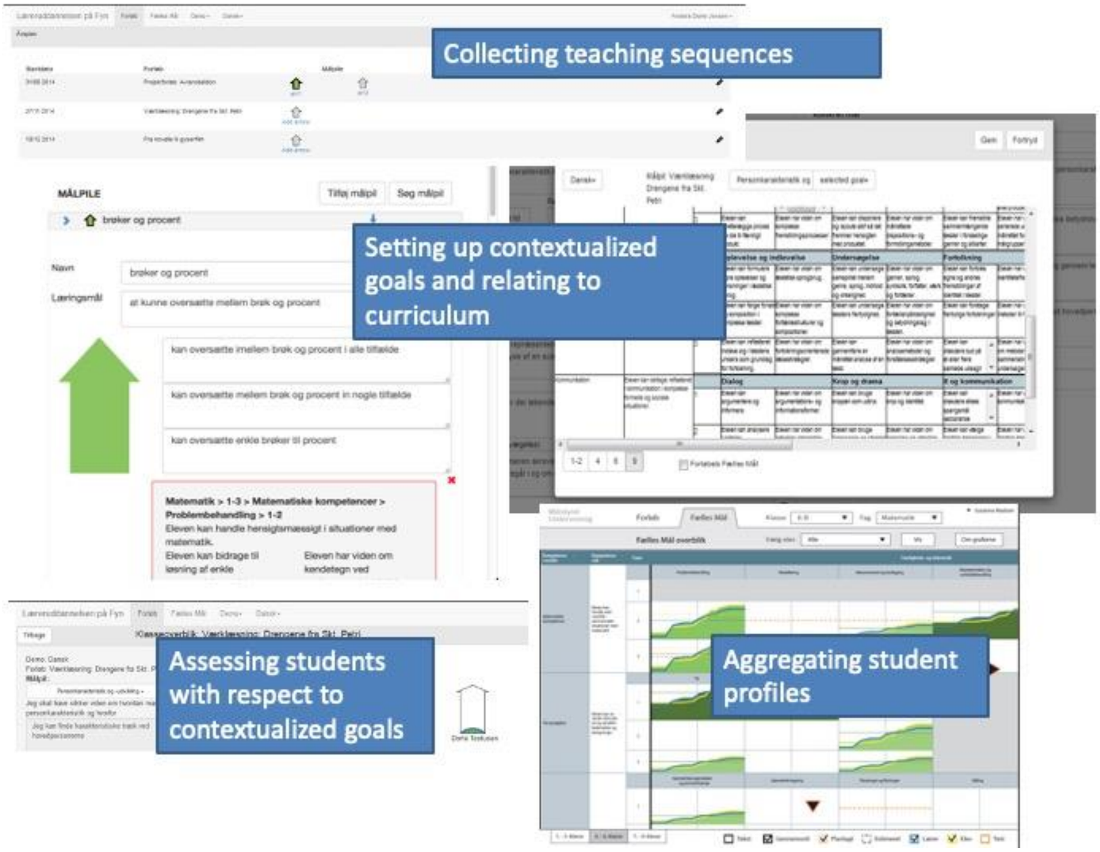

Collecting teaching sequences
Setting up contextualized goals and relating to curriculum
Assessing students with respect to contextualized goals
Aggregating student profiles
MÅLPILE
Tilføj målpil
Søg målpil
brøker og procent
Navn
Læringsmål
at kunne oversætte mellem brøk og procent
kan oversætte imellem brøk og procent i alle tilfælde
kan oversætte mellem brøk og procent in nogle tilfælde
kan oversætte enkle brøker til procent
Matematik > 1-3 > Matematiske kompetencer > Problembehandling > 1-2
Eleven kan handle hensigtsmæssigt i situationer med matematik.
Eleven kan bidrage til løsning af enkle
Eleven har viden om kendetegn ved
Gem
Fortryd
Fælles Mål overblik

Figure 2. The workflow in the Goal Arrow

We have conducted a design-based research project involving the development of a digital tool that distinguishes the National standards from the day to day objectives in the classroom. The *Goal Arrow* is a tool that allows teachers to express their own learning goals for their students and evaluate their progress in relation to those goals. One of the promises of the tool is to move the assessment practice away from single situations (e.g. tests) and towards a more integrated, ubiquitous and ongoing part of the teaching practice.

The Goal Arrow helps teachers to scaffold their process of describing lesson plans, expressing associated situated learning goals, and relating these to the National Curriculum. Each learning goal is specified into three objectives, which can be identified in students' actions or products. The descriptions, goals and objectives can be used when the teacher communicates with the students about the goals of the course. Each goal is also related to the national curriculum by clicking on the various elements in the matrix representation of the curriculum.

The workflow of teaching with the Goal Arrow begins with setting up learning objectives, or Goal Arrows (bottom left of Figure 2), and relating them to the Danish learning standards in a pop-up window (top left of Figure 2). Each learning objective can then be assessed on the level of individual pupils (top right of Figure 2). This input is used to create an overview of the extent to which national learning standards are met on both an individual and a class-based level (lower right of Figure 2).

During teaching, the Goal Arrow supports teachers' ability to log student progression with respect to the situated learning goals. This allows teachers to collect data in class about individual student performance and understanding related to the specific objectives of the lesson plan. In addition, it allows teachers to conduct teaching and assess student performance with strong focus on the learning goals for the day and without having to consider the entire system of curriculum standards. Furthermore, the system contains functions that allows students to assess their own performance based on the situated learning goals.

The relation between the situated learning goals and the national curriculum standards are supported by the Goal Arrow system. This means that the system proposes a student profile related to the national standards based on the data that teachers and students collect. Hence, the tool allows teachers to evaluate student progress both in the classroom – where the system provides an easy way to collect data about each specific learning activity – and after the class – where the teacher can access and manipulate an overview of the students' performance in relation to the national standards.

Discussion

The two examples above highlight the ways in which technology in general, and visualizations in particular, can support teachers in preparing, conducting, and evaluating their teaching. In particular, the main use-case of the Process Tab is to help teachers conduct their teaching in the moment via a representation of the complex and collaborative processes of their students to guide formative feedback and interventions. While this visualization was designed with real-time use in mind, it may also be used to help prepare teaching by reviewing connections that students may have missed in order to guide future interventions, and to evaluate students after the activity for assessment purposes.

The main use-case of the Goal Arrow is to help teachers prepare their teaching via a representation that connects their goals for student learning with curricular standards, prior to instruction. In addition, the goal arrow representation may (1) help teachers to conduct their teaching in the moment by allowing them to collect data on students (in relation to goals and standards) as they progress through the curriculum; and (2) help them to evaluate student progress after each class.

The above examples show that the way in which data visualizations support teacher work is a complex topic with important consequences for pedagogy. In the case of the Process Tab, the visualization allows teachers to engage directly with a representation of student thinking during instruction. While this is potentially powerful due to the scaffolds it provides for assessment and intervention, we found that it highlighted a tension between how teachers wanted to conduct their teaching and how they actually conducted it. The teachers in the study wanted to use the visualizations to conduct their teaching, but competing demands on their attention limited their ability to take advantage of the affordances of the tool.

In the case of the Goal Arrow, we found similar issues with teachers finding the necessary time to use the tool in class. This is especially problematic since the functionality of the system depends on teachers entering data into the system during class. The approach taken in the Goal Arrow of aligning teacher work with the national curriculum standards was a double-edged sword in supporting teachers' engagement with the system. The relevance of the system is supported by the close relation to the standards, and the system provides organized information about students' progression with respect to the standards. But simultaneously, this legislative relation makes the status of the tool unclear for teachers (Misfeldt et al., 2019), which challenges the functionality of the tool.

These finding suggests that future work with these pedagogical visualizations should either include different mechanisms for attracting teacher attention (e.g., alerting them only when action is needed, rather than requiring them to monitor) or focus more on the prepare and evaluate phases of teaching. Given the difficulties previously found for teachers using the tools in real-time, they may be better suited to preparation and evaluation tasks that take place outside of class time. Work in the areas of designing and evaluating OLM and reporting systems for teachers can inform this work (Zapata-Rivera & Greer, 2004; Zapata-Rivera & Katz, 2014).

Recommendations and Future Research on ITSs

Generalized Intelligent Framework for Tutoring (GIFT) managed ITSs may not involve teachers in a traditional sense, but often have authors and controller observers (facilitators) whose roles overlap with teachers. As such, visualizations incorporated into these ITSs could help such authors and facilitators to prepare, conduct, and evaluate their instruction, designs, and interventions with students. For example, visualizations that link goals and curriculum may help authors prepare their content using GIFT. In addition, visualizations that track participant progress or activity could help facilitators monitor participants' activity during the ITSs, scaffold interventions if necessary, and evaluate their progress after the activity.

Our prior work shows two examples of how this might be done, but suggests that it is difficult to develop visualizations that are easy to use in the moment to conduct teaching. In turn, this suggests that focusing on developing visualizations that support preparation and evaluation may be most beneficial at least in early stages of development. Important future work will be to investigate how to develop visualizations that help educators to conduct their teaching given that they have competing demands on their attention. This will be particularly challenging for team tutoring situations, because educators may have to focus on both the team level and the individual level. Previous work has shown that teachers would like control over a system that can perform most tasks autonomously (Zapata-Rivera et al., 2007). That way, teachers can benefit from having the system handle common cases while alerting them of cases that need additional attention. This approach takes into account their time constraints while highlighting the importance of designing visualizations to support teachers' use of data in the classroom, and is a promising direction for future work bridging ITSs and classroom-based teaching.

As a final recommendation, we would like to propose that the PCE division itself is implemented as a framework for discussing ITS solutions. Based on the research and experience with implementation solutions, we recommend that PCE is used as a part of the quality assurance guidelines for content creators, learning scientists, and teachers, when assessing the practical usability and value of solutions.

Conclusions

The findings presented in this chapter suggest that our existing knowledge of teacher actions in technology-enhanced systems lacks depth and is difficult to model. We hope that a simple division into prepare, conduct, and evaluate can support discussion and exploration of teacher work with technology, both to the benefit of the teacher profession, but also to support the development of ITSs that allows for relevant and timely inputs from teachers and other educators.

Acknowledgements

This work was funded in part by the National Science Foundation (DRL-1661036, DRL-1713110), the U.S. Army Research Laboratory (W911NF-18-2-0039), the Wisconsin Alumni Research Foundation, the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison, the

Danish Ministry of Education and Municipalities Denmark. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

References

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.

Herder, T., Swiecki, Z., Fougt, S. S., Tamborg, A. L., Allsopp, B. B., Shaffer, D. A., & Misfeldt, M. (2018). Supporting teachers intervention in students virtual collaboration using a network based model. *I LAK '18 Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (s. 21-25). Association for Computing Machinery. ACM International Conference Proceeding Series (ICPS)

Misfeldt, M., Tamborg, A. L., Dreyøe, J., & Allsopp, B. B. (2019). Tools, rules and teachers: The relationship be tween curriculum standards and resource systems when teaching mathematics. International Journal of Educational Research 94, 122-133. https://doi.org/10.1016/j.ijer.2018.12.001

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of educational research*, *75*(2), 211-246.

Shaffer, D. W. (2006). *How Computer Games Help Children Learn*. New York, NY: Palgrave.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9–45.

Sherin, M. G., & Drake, C. (2009). Curriculum strategy framework: Investigating patterns in teachers' use of a re form-based elementary mathematics curriculum. Journal of Curriculum Studies 41(4), 467-500.

Swiecki, Z., Misfeldt, M., Stoddard, J., & Shaffer, D. A. (2017). Dependency-Centered Design as an Approach to Pedagogical Authoring. In Y. Baek (Ed.), Game-Based Learning: Theory Strategies and Performance Outcomes (1 ed., Vol. 1, pp. 167-188). Hauppauge NY: Nova Science Publishers. Education in a Competitive and Globalizing World.

Zapata-Rivera, D., Graesser, A., Kay, J. Hu, X., & Ososky, S. (2020) Visualization Implications for the Validity of ITS. This volume.

Zapata-Rivera, D., & Greer, J. E. (2004). Interacting with inspectable Bayesian student models. *International Jour nal of Artificial Intelligence in Education*, 14(2), 127-163.

Zapata-Rivera, D., & Katz, R. I. (2014). Keeping your audience in mind: Applying audience analysis to the design of score reports. *Assessment in Education: Principles, Policy & Practice*, 21, 442–463. (2014).

Zapata-Rivera, D., Hansen, E. G., Shute, V. J., Underwood, J. S., & Bauer, M. I. (2007). Evidence-based approach to interacting with open student models. *International Journal of Artificial Intelligence in Education.* 17(3), 273-303.

Chapter 13 – Using Digital Twins to Visualize System Performance during Adaptive Instruction of Maintenance Tasks

Robert A. Sottilare[1], Jeremiah Folsom-Kovarik[1], Jody L. Cockroft[2], and Andrew J. Hampton[2]
Soar Technology, Inc.[1], University of Memphis[2]

Introduction

The US military needs technical personnel capable of diagnosing and resolving operational system issues as part of their maintenance duties. The acquisition of these skills requires a long process and fully qualified individuals are rare commodities. The fielding of new equipment is particularly challenging because opportunities to gain maintenance experience prior to equipment deployment are limited. This complication does nothing to mitigate the demands of the operational environment for maintenance personnel to perform their duties proficiently and efficiently. Military service members typically attend school and receive most of their rate-specific training prior to deployment. This intensive training can last up to two years. By the time trained personnel reach their duty assignments, their skills have decayed, operational systems have been updated, and the training content used to instruct them may be outdated.

One approach to improve the skills of qualified technical personnel is through an apprenticeship model where qualified personnel mentor the next generation of maintainers both prior to their access to the physical system (in the schoolhouse) and then continuing their coaching while deployed with the physical system (at sea). The apprenticeship approach has proven to be a highly effective method for military training domains, but it requires large time investments to deliver one-to-one tutoring (Sottilare & Proctor, 2012).

To further enhance each maintainer's exposure to expert support, accelerate their learning, minimize their skill decay, improve their transfer of skills from the schoolhouse to the operational platform, and support on-the-job training once deployed, we recommend the use of digital twin technology. The digital twin will form the basis of an adaptive instructional tool that will augment the current apprenticeship model from the schoolhouse to the initial deployment and beyond.

A digital twin is a virtual model of a system or other physical entity where data are transmitted seamlessly between the physical entity and virtual model. The entity and model then simultaneously co-exist as identical entities through real-time updates. A key feature for developing a self-improving digital twin is to make the digital twin part of the Internet of things (IoT). The twin (virtual system model) may then be visualized on any device with Internet access or, with cached data, in an offline mode anytime and anywhere (Datta, 2016). This is its primary advantage over other standalone systems or wholly simulated systems.

Three uses of digital twins are product design, planning for manufacturing and production, and analysis of system performance. Digital twin technology can support the latter use by capturing, analyzing, and acting on operational system data to create realistic system performance models and maintenance training scenarios. These models and scenarios can, in turn, support anytime, anywhere virtual training prior to deployment and in structured on-the-job training (OJT). This chapter addresses the advantages of using digital twin technology to visualize system performance for adaptive maintenance training, and in the next section we define visualization and discuss common types of visualization and their applicability to digital twins.

Visualization in Digital Twins

As we discuss visualization for digital twins, we will refer to the form that each visualization takes with respect to an example digital twin of an automobile. According to the Oxford English Online Dictionary (2020), *visualization* is defined as "the representation of an object, situation, or set of information as a chart or other image". Tory and Moller (2004, p. 1) note that visualization is historically in two primary categories: "scientific visualization which involves scientific data with an inherent physical component" and "information visualization which involves abstract, nonspatial data".

Nagel (2006, p. 2) defines *scientific visualization* as "accurate visualizations of the real world" while *information visualization* represents "abstract, non-physically based data to amplify cognition". In describing visualization processes for digital twins, we are primarily representing real world processes and measures so this is primarily scientific visualization. However, short term events like a tachometer reading of revolutions per minute (RPM) might reflect a machine learning algorithm based on system performance over a long period of time. This type of visualization is more abstract and informational since it is not necessarily a real-time view of RPM at a precise moment in time. The goal of visualization is to refine raw data into graphical formats to allow easier understanding of complex relationships within the data. Toward this purpose, there are several common types of visualization.

Linear visualization is usually a list of items organized by a single feature (e.g., ranking, numerical or alphabetical order). Gauges in an automobile provide linear visualization of the numerical representation of engine speed (RPMs) and vehicle speed in miles per hour (MPH).

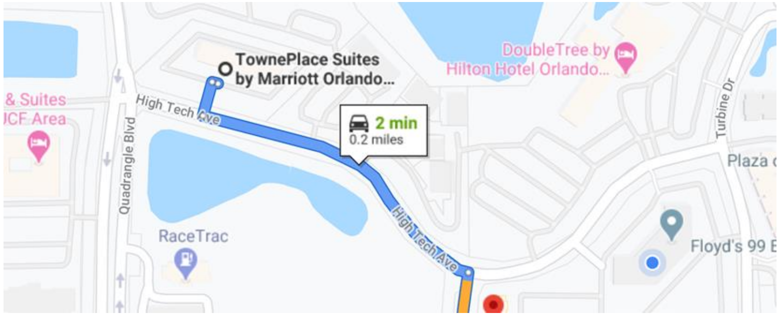


Figure 1. Planar representation of GPS location

Planar visualization is a graphical representation of data defined in two-dimensional grids (e.g., areas, floorplans, and geographic maps). Figure 1 illustrates the location of a vehicle in a map view as measured by a Global Positioning System (GPS). While this is a two dimensional visualization, GPS coordinates represent a grid on the surface of the sphere of the earth and are actually three dimensional in that they represent a position in two dimensions and an elevation coordinate. For example, the GPS coordinates for the center of Orlando, Florida are 28.5383° N, 81.3792° W, but there is a difference in positioning for someone at ground level and someone in a multi-story building.

Volumetric visualization is a graphical representation of data defined in three-dimensions (e.g., real world spaces, and spheres based on 3 variables). Returning to our example of a digital twin of an automobile, sensor data might be used as a three-dimensional heat map to represent thermal flow within an engine. In this case, the areas of higher temperature visualize the effects of friction or thermal flow from the internal combustion process.

Multidimensional visualizations include the representation of two or more variables (e.g., three dimensional frames that change over time and represent four dimensions). Referring back to our example of a digital twin of an automobile, thermal efficiency and heat rejection in various engine components may be shown graphically to increase with shaft power output (Al-Shemmeri, 2011). Optimization problems may also be represented by multidimensional visualizations. For example, an electric automobile design might tradeoff passenger seating, weight, torque and motor speed (RPMs).

Temporal visualization represents data as a set of dependent variables visualized with respect to time (e.g., daily closing stockprice over the course of one year). For our automobile example, the digital twin might track variables like engine temperature over time for four different lubricants (Xin, 2013).

Tree visualization is a representation of hierarchical data where a treemap shows relationships in a series of nested rectangles of sizes proportional to the corresponding data value. While tree visualizations or treemaps are important methods of representing the relationships between data (e.g. ontologies), they are not applicable to the visualization of measures in systems and their digital twins. *Network visualization* (also called graph visualization or link analysis) is a representation of very large datasets as connected entities of links and nodes. Networks visualize connections between nodes or highlight paths (e.g., paths of learning activities). While repair processes may be represented in network visualizations, they are generally not useful for digital twins.

In the next section, we discuss the elements of various physical systems and how they might be visualized in a digital twin for maintenance. We explore visualization of digital twins in the context of automotive maintenance.

Visualizing Physical Systems

Because a digital twin is a virtual model of a physical system, a key objective is to visually represent elements of the system. Mechanical, electrical, fluid, and thermal systems and components are likely elements of digital twins. Both physical components, associated attributes and processes must be visualized to allow authors to reconfigure the digital twin to support maintenance training scenarios. Though the measures of the physical system may be translated as discrete measures, many instrumented systems involve continuous measures that vary and should be visualized as continuous plots in order to be diagnosed properly. A digital twin for adaptive maintenance training should also include peripheral tools. For example, a multi-meter should also be modeled as part of a digital twin to measure electrical system resistances, currents, and voltages.

Mechanical Systems

Mechanical properties like elasticity, resistance, and inertia must be visualized to represent the attributes of physical system objects (e.g., springs and masses). Inputs to the system (e.g., forces) and outputs (e.g., displacement) must also be represented to accurately model the physical system. Mechanical systems may be used to represent system functions, wear, and failure in the context of automobile maintenance.

Electrical Systems

Similarly, the basic building blocks of electrical systems are resistance, inductance and capacitance. Resistance is the opposition that an object or material has in impeding the flow (of electricity, fluid, gas, or heat). Inductance is the tendency of an electrical conductor to oppose a change in the electric current flowing through it. Capacitance is the ratio of the change in an electric charge in a system to the corresponding change in its electric potential.

Fluid Systems

The primary measures in fluid systems are volume/rate of flow and pressure differences. Fluid systems are either hydraulic or pneumatic. Hydraulic systems are concerned with the flow of liquid through valves, connectors, and pipes. Hydraulic system properties include resistance, capacitance, and inertance (a measure of the pressure difference in a fluid). Pneumatic systems are concerned with the flow of compressed gas through valves, connectors, and pipes, and have the same properties of resistance, capacitance, and inertance.

Thermal Systems

The primary measures for thermal systems are resistance and capacitance. Heat flows between two points if there is a temperature difference between them. Thermal resistance is a reciprocal of thermal conductance and a measurement of a temperature difference by which an object or material resists heat flow. Thermal capacitance is the thermal energy storage capacity of an object or material.

Visualization Types vs. Physical Systems

If we examine mechanical, electrical, fluid and thermal systems they generally have analogous physical attributes. For example, in a mechanical system, the rate of displacement in a particular direction is velocity. While in an electrical system, the flow of charge over time is current. Using a hybrid (gasoline-electric) automobile as an example system, Table 1 provides examples of various physical phenomena that might be visualized in support of maintenance training.

Table 1. Visualization Types vs. Physical Systems

		Physical Systems			
		Mechanical	**Electrical**	**Fluid**	**Thermal**
Visualization Types	**Measures (Linear, Planar Volumetric)**	Displacement, Rotation, Mass	Resistance, Inductance, Capacitance	Pressure	Temperature
	Temporal	Velocity	Current	Hydraulic and Pneumatic Flow	Heat Flow
	Tree	Ontology: relationship of parts (e.g., carburetor is part of an engine; carburetor mixes air and fuel and involves fluid processes)			

Each of these building blocks can receive data from a digital twin to update its performance model to support maintenance training for a specific individual system. Much in the way mission rehearsal is training for a specific set of tasks related to a specific scenario (set of events and conditions), digital twins can support offline training specific to the performance characteristics of an individual system, but there are some challenges to enabling training outside of the operational system.

Challenges in Using a Digital Twin for Adaptive Maintenance Training

In this section, we explore four challenges to developing digital twins for effective maintenance training along with associated risks, potential approaches, and expected outcomes.

Designing an effective and secure digital twin model

The primary challenge to designing an effective and secure digital twin is the availability and secure transfer of system data to the digital twin on a regular basis. These data ensure that the performance model stays current. The acquisition of system data on a frequent basis is possible as part of the IoT. One approach is to design the digital twin as a discrete event process in which system data are used to develop various models of physical processes (e.g., mechanical, electrical, fluid, and thermal). Encryption and secure cloud storage of system data offer the opportunity to pull needed data from geographically disparate systems. Machine learning (ML) techniques can then be used to construct various system state models.

Discrete event simulations based on operational system data might be used to support maintenance training for a variety of systems. This would allow the training system developer to inject what-if scenarios including catastrophic failures that would not be possible to train on the operational system. This approach is expected to deliver a low-cost, modular design with reusable, interoperable virtual components that can be arranged to represent various systems. This in turn should result in the ability to support anytime, anywhere maintenance training or system diagnosis of distant physical systems, as well as the ability to closely mirror real systems. In this way, digital twins can support near transfer of training to operations.

Designing an easy-to-use scenario authoring tool

The primary challenge described here is the ability to accurately represent system processes to author maintenance scenarios in the digital twin. Using a finite set of virtual building blocks (VBBs) as part of a discrete event simulation process is one approach to representing physical processes in the digital twin. An easy to use drag and drop dashboard concept allows subject matter experts (SMEs) to configure physical systems using reconfigurable VBBs. SMEs can represent different maintenance scenarios by injecting faults or by changing system parameters through a visualization of the system in a dashboard format. The ability to reuse VBBs to support visualization of physical processes occurring in operational systems will provide both realism and the opportunity for reduced development costs, but how should these VBBs be designed?

Designing a system of interoperable components

If our maintenance training system relies on configurable VBBs, we need to easily represent interoperable physical system elements and associated physical processes (e.g., hydraulic resistance, capacitance). Our goal is to enable the digital twin architecture to integrate new systems, create and adapt scenarios, and extend the system without major software changes. One approach is to design a standard protocol for data exchange between the physical system, the digital twin, and digital system elements represented by VBBs in a dashboard format. By developing a gateway specification to exchange standardized messages between the system elements, VBBs, and across systems, authors (e.g., instructors, designers, SMEs) will be able to reuse common elements and build new components that are compatible with the existing digital twin.

Steps toward a robust self-improving instructional model

Our final challenge is to incorporate a self-improving instructional model—a mechanism which learns and becomes more effective with increased exposure to learners and new scenarios. Intelligent agents are well suited to taking in information from their environments and making decisions with a goal to improve instruction. The ability to model the relationships between learning objectives and measures, various learner states, and the conditions of the system represented by the system performance model are critical to enhancing instructional policies over time (Figure 2).

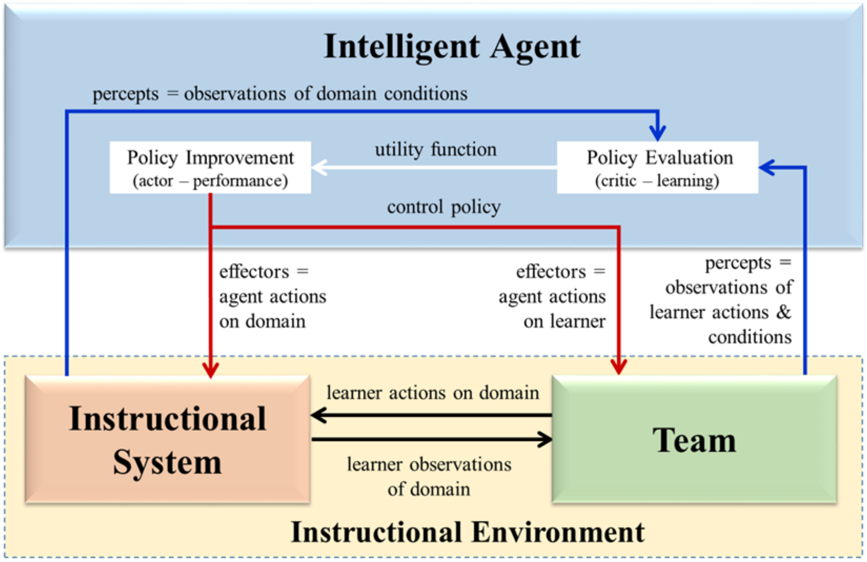


Figure 2. Intelligent Agent interaction within Adaptive Instructional Systems for Teams

Such a self-improving instructional model may be accomplished via an agent-based system that updates policies based on contextual decisions (e.g., content selection, feedback). In this approach, the self-improving module tracks a range of variables related to the learner (e.g., overall mastery, knowledge component scores, psychological characteristics), the content (e.g., difficulty, novelty, mathematical versus conceptual), and subsequent decisions (e.g., recommended next problems and exercises), combining relevant dimensions to create complex contextual decisions. Comparison of contextual decisions to their effectiveness can then influence instructional policy and future instructional decisions. The expected results of this approach include the ability to:

- understand the relationship between learner actions, tutor decisions, and system conditions to gauge instructional effectiveness
- understand the changing competency model of the learner(s)
- automatically adapt policy for an ever-improving instructional model
- sequence and present training in a manner that supports far transfer of training

Recommended Next Steps

If the challenges noted herein can be overcome, then digital twins can be used not just for maintenance training, but also to support the design of systems and planned product improvement of deployed systems. Preventative and conditional maintenance schedules might also be structured based on predictive models grounded in digital twin technology. Recommended next steps for using digital twins in military and commercial systems are:

- the standardization of VBBs to support interoperability and reuse
- the development of VBBs to represent the most common physical system elements and processes
- the development of secure methods to move data from the operational system to the digital twin in real-time

Digital twin technology might also be applied beyond military and commercial systems. Future digital twins might include digital health twins for patients to support remote health screenings and train physicians on virtual patients. Physiological data is already being remotely collected from patients via smartphones (Haberman et al., 2015). Digital twins might also be applied to the monitoring of complex physical phenomena where accurate predictive models based on real-time data are needed. The ability to understand the dynamic systems behind earthquakes, hurricanes and tsunamis could benefit from digital twins.

References

Al-Shemmeri, T. T. (2011). Thermodynamics, performance analysis and computational modelling of small and micro combined heat and power (CHP) systems. In Small and Micro Combined Heat and Power (CHP) Systems (pp. 42-69). Woodhead Publishing.

Datta, S. P. A. (2016). Emergence of digital twins. arXiv preprint arXiv:1610.06467.

Haberman, Z. C., Jahn, R. T., Bose, R., Tun, H., Shinbane, J. S., Doshi, R. N., ... & Saxon, L. A. (2015). Wireless smartphone ECG enables large-scale screening in diverse populations. *Journal of cardiovascular electrophysiology*, *26*(5), 520-526.

Nagel, H. R. (2006). Scientific visualization versus information visualization. In Workshop on state-of-the-art in scientific and parallel computing, Sweden (pp. 8-9).

Oxford English Online Dictionary. "Visualization defined". Accessed March 2020.

Sottilare, R. A., & Proctor, M. (2012). Passively classifying student mood and performance within intelligent tutors. Journal of Educational Technology & Society, 15(2), 101-114.

Tory, M., & Moller, T. (2004, October). Rethinking visualization: A high-level taxonomy. In IEEE Symposium on Information Visualization (pp. 151-158). IEEE.

Xin, Q. (2013). Diesel aftertreatment integration and matching. *Diesel Engine System Design*, 503-525.

Chapter 14 – Visualizing Intelligent Tutoring System Data in Real time

Xiangen Hu[1, 2], Vasile Rus[1], Jody L. Cockroft[1], and Liang Zhang[1]
University of Memphis[1], Central China Normal University[2]

Introduction

We often consider Data Visualization in the context of data analysis and reports. Data visualization can be intuitively understood as the process to represent data in pictorial or graphical format. The process starts from collected data and ends with human understandable pictorial or graphical displays. Based on Friendly (2008), data visualization has been part of human civilization almost as early as the 14th century. Some of the earlier efforts of data visualization made significant contributions to modern society and technology, such as the invention of maps (Friendly, 2008).

Only recently, data visualization has been closely associated with scientific research. There are some basic rules in statistics that require different types of graphics for different data. For example, the graphic representation for interval scaled data would be different from that of ratio scaled data (Dixon & Massey, 1951). Some researchers in social sciences and psychology have been very particular when analyzing and representing numerical data. For example, the analysis of data types and their meanings are the basis of a well-developed theoretical framework called foundation of measurement theories (Luce, Suppes, & Krantz, 2007) and meaningfulness (Falmagne & Narens, 1983; Narens, 2012).

With the so-called big-data revolution (Cuzzocrea, Song, & Davis, 2011; Kitchin, 2014) computer scientists have extended the use of data visualization to a new research field called visual data mining (Keim, 2002). In the digital era, Data Visualization can be, and is becoming the cornerstone of new knowledge discovery and decision making, not just for traditional descriptive statistics, but for virtually any domain (Ward, Grinstein, & Keim, 2015). For example, a timeline for historical events in modern history with dates and population data could be visually displayed to help readers to see trends of potential issues. Some reasons for increasingly embracing data visualization other than historical ones are: 1) the ease of cognition by graphical representations within the human perceptual system; 2) the effectiveness and efficiency of the visual process in knowledge discovery; and 3) the great potentiality of visual analysis in exploring strategies for decision-making in one virtual digital space, especially from a modern perspective (Friendly, 2008; Ward et al., 2015).

In this chapter, we explore Data Visualization from the point of view of Friendly (2008) where he argued that data in Data Visualization can be any type of data (not just digital, for example it could be analogue (such as display of maps, qualitative diagram, emoticons in social media)). We consider the process of data visualization simply to be a mapping that transforms information between two different systems where the target system happens to be human cognition. This process will help humans (the target system) to better and more effectively comprehend information from the source. To illustrate, we consider the process of making a map. The map producers first measure distances and geographic information of the physical world and then transform this information pictorially on paper for humans to better understand the geographical reality of the map. It is worth mentioning that modern dynamic and interactive Data Visualization have extended the static map to complex (virtual environment) interactions between the virtual and the real world (from traditional one-way to modern two-way).

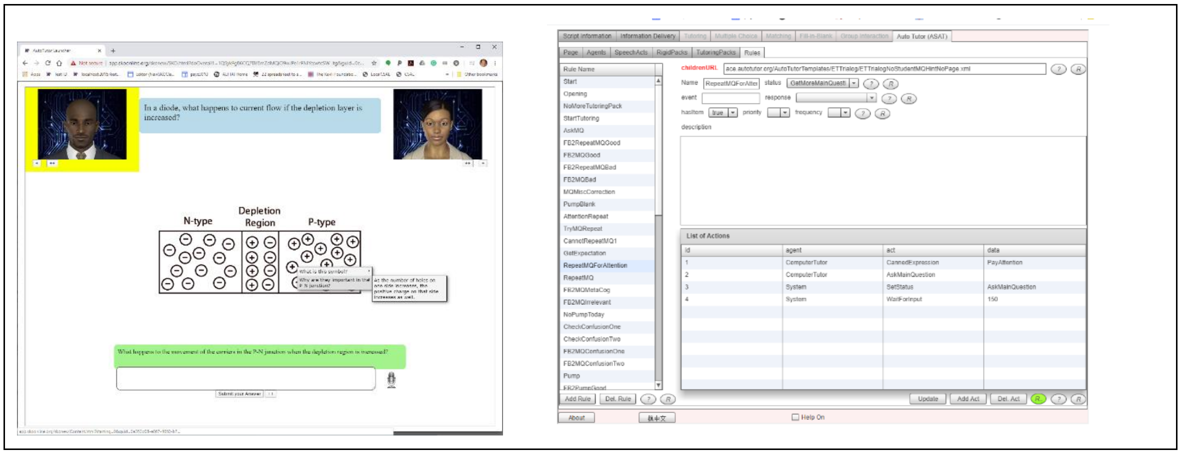
Another nontrivial example of data visualization would be human facial display of emotions. When there is any discomfort in the human body, such as a headache, there is a (visible) facial display correlated to the severity of the discomfort. There has been extensive research on how humans display emotions as a function of mental status (Ekman, 1984; Ekman & Davidson, 1994; Ekman & Friesen, 1971). The human emotion display has been used widely in social media such as emoticons (Walther & D'Addario, 2001). Another relevant example is the technique used in cognitive neurosciences where Data Visualization has made it possible for human researchers (and AI computer cognition) to visually/pictorially examine the mental process in the human brain (Raichle, 2003). We argue that when we consider Data Visualization in the context of an Intelligent Tutoring System (ITS), which is a system mimicking human tutors, we need to consider how one would visualize data for an ITS like a human visualization of emotions.

We plan to explore the Data Visualization of ITSs from this new perspective by considering ITSs in a symmetric framework proposed earlier by the authors (Hu, Cai, & Graesser, 2019). It was proposed that ITSs be considered a symmetric framework where human learners and dynamic learning resources such as ITS applications are the two major components of this framework. From this perspective, we argue that one of the tasks is to apply Data Visualization processes in ITSs to pictorially communicate how ITS applications work internally, like brain imaging in cognitive neurosciences.

Two Examples

We first provide two example ITSs and demonstrate the potentials for Data Visualization as a process that communicates pictorially the underlying operations of an ITS. The first example is the most updated implementation of AutoTutor (Nye, Graesser, & Hu, 2014). The second example is the current baseline of GIFT (Generalized Intelligent Framework for Tutoring), in its desktop version (not the cloud version) (*Overview - GIFT - GIFT Portal*, n.d.).

AutoTutor internal operations



In a diode, what happens to current flow if the depletion layer is increased?

What is this symbol?

Why are they important in the P-N junction?

As the number of holes on one side increases, the positive charge on that side increases as well.

What happens to the movement of the carriers in the P-N junction when the depletion region is increased?

Rule Name
Start
Opening
NoMoreTutoringPack
StartTutoring
AskMQ
FB2RepeatMQGood
FB2MQGood
FB2RepeatMQBad
FB2MQBad
MQMiscCorrection
PumpBlank
AttentionRepeat
TryMQRepeat
CannotRepeatMQ1
GetExpectation
RepeatMQForAttention
RepeatMQ
FB2MQMetaCog
FB2MQIrrelevant
NoPumpToday
CheckConfusionOne
CheckConfusionTwo
FB2MQConfusionOne
FB2MQConfusionTwo
Pump
FB2PumpGood

childrenURL ace.autotutor.org/AutoTutorTemplates/ETTrialog/ETTrialogNoStudentMQHintNoPage.xml

Name RepeatMQForAtter status GetMoreMainQuesti

event response

hasItem true priority frequency

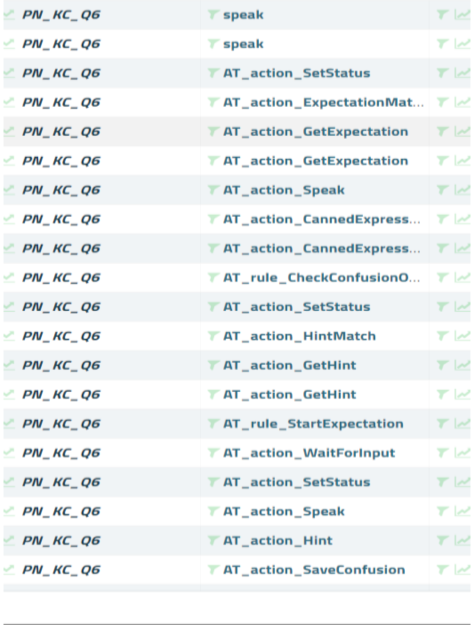
description

List of Actions

id	agent	act	data
1	ComputerTutor	CannedExpression	PayAttention
2	ComputerTutor	AskMainQuestion	
3	System	SetStatus	AskMainQuestion
4	System	WaitForInput	150

Figure 1. Interface of AutoTutor: Learners interact with the Avatars by answering their questions in natural language. Learners interact with AutoTutor in multiple different forms. The AutoTutor Conversation Engine (ACE) will apply one or multiple rules (the list on the left) to process the input and decide the next action to take.

AutoTutor has shown learning gains of about 0.8 sigma over reading a book on a topic for the same amount of time (Graesser et al., 2003). AutoTutor poses problems to learners and manages the resulting dialogues using expectation-misconception tailored (EMT) dialogue, a dialogue frame, and conversational templates. EMT uses pumps, hints, and prompts to get a student to correctly fill in missing words, phrases, and sentences while anticipating expectations and misconceptions (Graesser et al., 2004). Conversations between learners and AutoTutor applications are made possible by the AutoTutor Conversation Engine (ACE) (Cai, Hu, & Graesser, 2019). ACE operates on AutoTutor Scripts, which organizes main questions, pumps, hints, prompts, and elaborations in the form of xml, that are necessary for EMT dialogues.



PN_KC_Q6	speak
PN_KC_Q6	speak
PN_KC_Q6	AT_action_SetStatus
PN_KC_Q6	AT_action_ExpectationMat...
PN_KC_Q6	AT_action_GetExpectation
PN_KC_Q6	AT_action_GetExpectation
PN_KC_Q6	AT_action_Speak
PN_KC_Q6	AT_action_CannedExpress...
PN_KC_Q6	AT_action_CannedExpress...
PN_KC_Q6	AT_rule_CheckConfusionO...
PN_KC_Q6	AT_action_SetStatus
PN_KC_Q6	AT_action_HintMatch
PN_KC_Q6	AT_action_GetHint
PN_KC_Q6	AT_action_GetHint
PN_KC_Q6	AT_rule_StartExpectation
PN_KC_Q6	AT_action_WaitForInput
PN_KC_Q6	AT_action_SetStatus
PN_KC_Q6	AT_action_Speak
PN_KC_Q6	AT_action_Hint
PN_KC_Q6	AT_action_SaveConfusion

Figure 2. Example xAPI Statements in the Learning record store (LRS): Several xAPI statements are sent to the LRS each time when a learner contributes an answer. The statements are records of AutoTutor Conversation Engine (ACE) actions processing the natural language input.

With a given AutoTutor Script, an ACE enabled interface (called Sharable Knowledge Object (SKO)) (Nye et al., 2014) accepts and sends learners natural language input to ACE. ACE computes the similarity between student's input and stored EMT elements (Expectations, misconceptions, pump completions, hint completions, etc. (see Graesser et al., 2005)). Figure 1 shows a sample interface of a SKO, where ACE is operating and manages the interaction. Each time when a learner answers a question (a main question, a hint, or a prompt), ACE will provide feedback based on a set of rules (See Figures 1 and 2). When any of the rules is selected, a sequence of actions will be performed. For example, if *RepeatMQForAttention* is selected, then there are four actions performed in the following order: ComputerTutor will 1) select a canned Expression (*CannedExpression*), 2) Ask main question (*AskMainQuestion*), 3) Set the systems status that main question is asked again, 4) Wait for response. While the observed actions are limited, in this case, only four, there are complicated computations and decisions in ACE. Typically, it takes possibly hundreds of computations to decide what to do next. If we capture all the computations and microscopic levels of decision (see Figure 2 for a sample list of lower level activities recorded in the LRS).

Based on the example of AutoTutor described, we can imagine that an appropriate visualization of data (underlying microscopic operations/decisions) would be pictorially displayed on the "face" of the AutoTutor avatar. For example, the AutoTutor avatar could display a sad face when there is a lower level confusion, instead of simply asking the main question again. Technologically, it is relatively simple to implement. We may just assign different facial displays to the avatar as a function of the evaluation of learners' input. However, we are not sure if this will have a positive or negative influence on learning. As soon as an ITS can display emotions based on the learner's answers, the learner's attitude towards the "machine" may change. So appropriately assigning facial displays to avatars (based on learner's input) could potentially impact learning.

GIFT Message Flow

GIFT prototype (https://cloud.gifttutoring.org/) can collect exhaustive internal event data with the ability to decide what data is used to build reports (see Figure 4). This data is very helpful to create graphic reports to describe the performance of learner and system post-hoc providing most of the needs for Data visualization.

One of the less familiar features of GIFT is available for the purpose we have alluded to earlier. This feature is only available when GIFT is running locally as a desktop application. When GIFT prototype runs in this mode, it provides some very informative data about GIFT, especially the underlying operations of GIFT when it interacts with learners. The GIFT Monitor Module (Figure 3) provides a dashboard that one can control each of GIFT modules. Like ACE, GIFT not only provides the real-time status of the learner, but also detailed real-time event data of the GIFT server. The event information reflects underlying modules communicating as a function of the students' learning status (see Figure 4).

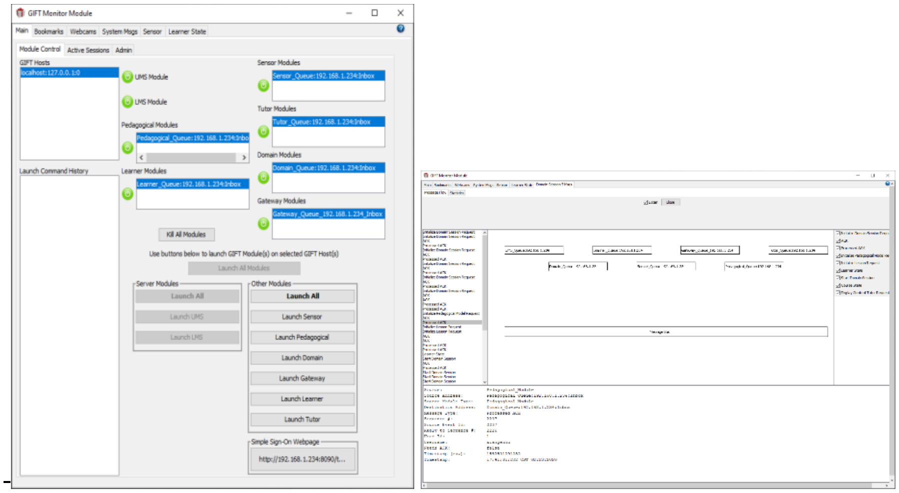


Figure 3. GIFT Monitor Module: GIFT Monitor Module available when running GIFT prototype on local windows systems. This is not yet available to users when running on the cloud version. Users can use this module to toggle different modules.

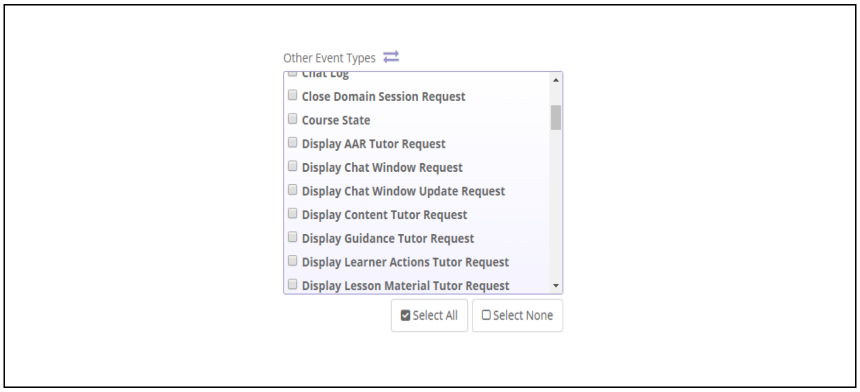
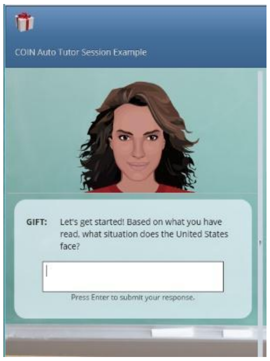

Other Event Types
Chat Log
☐ Close Domain Session Request
☐ Course State
☐ Display AAR Tutor Request
☐ Display Chat Window Request
☐ Display Chat Window Update Request
☐ Display Content Tutor Request
☐ Display Guidance Tutor Request
☐ Display Learner Actions Tutor Request
☐ Display Lesson Material Tutor Request
☑ Select All
☐ Select None

Figure 4. GIFT Monitor Module 2: The in the figure is also from the GIFT Monitor Module. It displays interaction among different components of the GIFT server when learners interact with the ITS.



GIFT: Let's get started! Based on what you have read, what situation does the United States face?

Press Enter to submit your response.

Figure 5. Avatar: The Avatar in the GIFT prototype interacts with learners using Natural Language. It is capable of displaying emotions.

Data Visualization for ITSs in real time

The examples in the previous section are typical examples of ITSs. Most ITS applications are implemented with data capturing capabilities that collect exhaustive internal data exchanges between modules. Very often, the collected data are used for post-hoc analysis for the improvements of the applications. We propose to use this data for real time feedback, like human emotional display. A similar approach has been rather frequently used in games and even in game-based learning applications (Johnson, 2007; Johnson & Wu, 2008), but has only been used in selected ITS applications (Johnson & Lester, 2018). Given that most ITS applications are capable of tracking internal activities, we believe these data can be used both for post-hoc analysis, as it has been used in ITS research, and as a basis for Data Visualization of an ITS in real-time in a form that is different from the current view. Instead, the Data Visualization we want to propose in this chapter can be thought of as a connection between the ITS application and human learner, such that the Data Visualization process involves visually displaying human-like, learning-relevant facial expressions in ITSs to communicate with human learners. In the two examples provided in the previous section, the data captured is the internal communication details among different modules of the ITS. It is time to propose this type of Data Visualization for ITSs due to the advanced technology both in simulating human emotions on computer screens (such as DAZ 3D (*DAZ 3D | 3D Models and 3D Software by Daz 3D*, n.d.), or realistic robots with embodied emotion display capabilities (*home - Hanson Robotics*, n.d.). We argue that we should consider a behavioral "symmetry" between human learners and ITSs. This new perspective of Data Visualization suggests that although Human learners and ITS applications have different underlying mechanisms, they behaviorally are the same.

With the discussions above, we consider Data Visualization for ITSs in real time the process to create learning-relevant facial displays meant to mimic human tutors, such as AutoTutor, or some GIFT powered ITS applications that have human avatars. A few fundamental questions need to be considered:

1. Psychologically how different is it when a learner interacts with a real human tutor versus an animated avatar?

2. Even if it is technologically possible to simulate all human emotions in a non-human avatar that is indistinguishable from real humans, what are the necessary emotional displays of the tutors that are relevant to and which promote learning?

3. ITSs such as AutoTutor applications and GIFT enabled applications can record exhaustive and detailed interactions between modules of ITSs (learner, domain, pedagogy, and interface); what are the minimum requirements for a list of interactive activities among the modules that can be used for a Data Visualization process?

There were relevant studies for the first and second questions. Graesser (2011) reported that students react differently with different types of tutors (rude, polite AutoTutor). He reported that the rude tutor is very engaging for some students, whereas other students would prefer to interact with a polite, supportive tutor. It was observed that the impact on learning is dependent on the phase of tutoring and the student's level of mastery. It is necessary to consider the learning-relevant emotions that are appropriate and promote learning as a function of a) phase of tutoring, b) student's knowledge levels, and c) the student's personality. Although the third question is technological in nature, it nevertheless requires answers from researchers and technicians about data collection and use: What can we collect with a given technology, what do we want to use to answer research and technological questions, and what do we need to create effective and efficient real world applications.

We consider the process of Data Visualization simply to be a mapping that transforms information between two different systems where the target system happens to be human cognition. While most of the Data Visualization for learning systems use students' data as the source of the process, we consider the sources as exclusively the internal intra-module communication data of the ITS. In addition, we consider the target systems of the process exclusively the learners while the outcome of other Data Visualization processes may also be consumed by non-students such as researchers, teachers, and administrators. We further consider that this information transforms within the symmetric framework of ITSs (Hu et al., 2019). In this framework, we specifically consider the ITS as the counterpart of human learners to present themselves in the form of human avatars. We specifically consider that when an ITS interacts with human learners that mimic human tutors, the result of the Data Visualization should resemble what a human tutor does: interact with learners with emotions. These emotional displays are a function of a student's performance.

Recommendations and Future Research

We consider one of the Data Visualizations processes for GIFT powered applications as the pictorial/graphical display of interaction data among GIFT modules (Figure 4). Specifically, we explore the possibility that these pictorial/graphical displays be transformed into human emotion displays of the avatar in real time. To make this possible, we will need to answer some theoretical questions raised in the previous section. In addition, we will need to solve some technical questions such as how to log system behavior data and what data standard is appropriate for the special data visualization needs. Initial proof of concept implementation from systems such as AutoTutor and the GIFT prototype made it possible for researchers to examine detailed systems behavior of ITSs. Although they are only available for system administrators of the systems, they nevertheless offer great potential for a new Data Visualization process for ITSs. We recommend that GIFT makes such capability a default feature for the course authors and learning scientists (in the cloud version) in addition to system administrators. Specifically, we suggest making the existing report feature of GIFT (Figure 4) as a real time update to the course creators. This will make it possible for the course creators to selectively transform relevant system messages as an emotional display on the face of the avatar. For example, when the system detects that the learner is making more careless errors, the avatar shows concern and helps the learner to pay more attention to details. With the availability of cameras and other biometric devices on personal digital devices, we recommend the GIFT prototype enable the make sensor module enabled as an advanced feature to further help researchers to explore the proposed Data Visualization processes in this chapter.

Conclusions

While conventional Data Visualizations processes help human consumers to comprehend information pictorially/graphically, we argue that Data Visualization processes implemented in ITSs should consider human learners as the consumer of the outcome. We demonstrated the feasibility of such implementation by examining existing prototypes of ITS in which the exhaustive system information is already tracked and stored. We pointed out that similar approaches have already been implemented in interactive games and game-based learning environments, yet it needs to be systematically studied before implementing in ITS where human learners are directly impacted. We have proposed a few theoretical and technological questions in this chapter as well as a few specific recommendations for future versions of GIFT prototypes.

References

Cai, Z., Hu, X., & Graesser, A. C. (2019). Authoring Conversational Intelligent Tutoring Systems. *Adaptive Instructional Systems*, 593–603.

Cuzzocrea, A., Song, I.-Y., & Davis, K. C. (2011). Analytics over large-scale multidimensional data: the big data revolution! *Proceedings of the ACM 14th International Workshop on Data Warehousing and OLAP*, 101–104.

DAZ 3D | 3D Models and 3D Software by Daz 3D. (n.d.). Retrieved February 27, 2020, from https://www.daz3d.com/

Dixon, W. J., & Massey, F. J., Jr. (1951). *Introduction to statistical analysis*. *370*. https://psycnet.apa.org/fulltext/1951-07185-000.pdf

Ekman, P. (1984). Expression and the nature of emotion. *Approaches to Emotion*, *3*(19), 344.

Ekman, P. E., & Davidson, R. J. (1994). The nature of emotion: Fundamental questions. *Series in Affective Science.*, *496*. https://psycnet.apa.org/fulltext/1995-97541-000.pdf

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129.

Falmagne, J. C., & Narens, L. (1983). Scales and Meaningfulness of Quantitative Laws. *Synthese*, *55*(3), 287–325.

Friendly, M. (2008). A Brief History of Data Visualization. In C.-H. Chen, W. Härdle, & A. Unwin (Eds.), *Handbook of Data Visualization* (pp. 15–56). Springer Berlin Heidelberg.

Graesser, A. C. (2011). Learning, thinking, and emoting with discourse technologies. *The American Psychologist*, *66*(8), 746–757.

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, *48*(4), 612–618.

Graesser, A. C., Jackson, G. T., Matthews, E. C., Mitchell, H. H., Olney, A., Ventura, M., Chipman, P., Franceschetti, D., Hu, X., Louwerse, M. M., & Others. (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*. https://cloudfront.escholarship.org/dist/prd/content/qt6mj3q2v1/qt6mj3q2v1.pdf

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: a tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *36*(2), 180–192.

home - Hanson Robotics. (n.d.). Hanson Robotics. Retrieved February 27, 2020, from https://www.hansonrobotics.com/

Hu, X., Cai, Z., & Graesser, A. C. (2019). GIFT as a Framework for Self-Improvable Digital Resources in SIAIS. *Of the 7th Annual GIFT …*. https://books.google.com/books?hl=en&lr=&id=3MvKDwAAQBAJ&oi=fnd&pg=PA49&dq=GIFT+Framework+Self+Improvable+Digital+Resources+SIAIS&ots=yBd9x5Q1ll&sig=SnzhN4OkxnK-mjIVTjzY8SmMU3ns

Johnson, W. L. (2007). Serious use of a serious game for language learning. *Frontiers in Artificial Intelligence and Applications*. https://books.google.com/books?hl=en&lr=&id=GEK93NUHdXYC&oi=fnd&pg=PA67&dq=johnson+lewis+Game+Iraq&ots=Rsv8jm8CcZ&sig=A4JVvQ0xmbkeEoAVJqEWVRhcWDo

Johnson, W. L., & Lester, J. C. (2018). Pedagogical Agents: Back to the Future. *AI Magazine*. http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=07384602&AN=130572466&h=Rd%2Fp0g3GUYc7VtaYT2v6%2BoFdrV4j%2BIitfd9uvKnvFs4%2F%2FkxWRjxfU8n%2B3ff1GCL1vrilmcLdq9YuP7xLRjOerQ%3D%3D&crl=c

Johnson, W. L., & Wu, S. (2008). Assessing Aptitude for Learning with a Serious Game for Foreign Language and Culture. *Intelligent Tutoring Systems*, 520–529.

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, *8*(1), 1–8.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE.

Luce, R. D., Suppes, P., & Krantz, D. H. (2007). *Foundations of Measurement: Representation, axiomatization, and invariance*. Courier Corporation.

Narens, L. (2012). *Theories of meaningfulness*. https://content.taylorfrancis.com/books/download?dac=C2009-0-26376-5&isbn=9781135640736&format=googlePreviewPdf

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, *24*(4), 427–469.

Overview - GIFT - GIFT Portal. (n.d.). Retrieved May 13, 2020, from https://www.gifttutoring.org/projects/gift/wiki/Overview

Raichle, M. E. (2003). Functional brain imaging and human brain function. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *23*(10), 3959–3962.

Walther, J. B., & D'Addario, K. P. (2001). The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication. *Social Science Computer Review*, *19*(3), 324–347.

Ward, M. O., Grinstein, G., & Keim, D. (2015). *Interactive data visualization: foundations, techniques, and applications*. https://content.taylorfrancis.com/books/download?dac=C2015-0-61440-7&isbn=9781482257380&format=googlePreviewPdf

Chapter 15 - Open Learner Model Visualizations for Contexts where Learners Think Fast or Slow

Judy Kay[1], Vasile Rus[2], Diego Zapata-Rivera[3], and Paula Durlach[4]

University of Sydney[1], University of Memphis[2], Educational Testing Service[3], U.S. Army Combat Capabilities Development Command (DEVCOM) – Soldier Center – Simulation and Training Technology Center[4]

Introduction

This chapter presents a new foundation for thinking about the design of Open Learner Models (OLMs) so that they can support learners in a broad range of important metacognitive processes. With the growing use of visualizations of learning data, there is an urgent need for new guidelines to inform such design of OLMs. It is particularly relevant for the Generalized Intelligent Framework for Tutoring (GIFT). GIFT aims to provide a platform for authoring personalized teaching systems, where self-regulated learning is aided by "open learner models that allow learners to decide what to learn next and to inspect their progress in mastering the subject matter and their measured psychological attributes" (Sottilare, Baker, Graesser, & Lester, 2018, p.142).

Many learning systems have been built with diverse forms of learner modeling and OLMs. In Intelligent Tutoring Systems (ITSs), learner models drive the personalization that is a defining characteristic of an ITS. Designing a learner model has the following stages:

- *Determine **what** to model*. Essentially, this involves an abstract modeling phase, to determine the questions that the learner model should answer, such as: Does the learner know how to read a simple loop in C++? When an OLM is a core purpose of creating the learner model, the decision about what to model should be driven by careful decisions about what aspects the learner needs to be able to see in their OLM.
- *Define the learner model **ontology***. This is the learner model "Components" that the designer of the model considered important to represent. These are often in a hierarchy and may include the learner's knowledge, preferences, goals and other attributes. For example, one Knowledge Component may represent the learner's knowledge of reading simple C++ loops. Here, too, when the ITS designer plans to provide an OLM for the learner, this ontology must be designed so that it maps to the information to be made available at the OLM interface.
- *Design the **representation** for the learner model components.* This is the way that the internal software operates. This, in turn, determines the nature of the answers that a learner model can provide about the learner. For example, the value of a Knowledge Component may be reported as a number (say, 0 to 100), or qualitative ratings (such as novice, intermediate, expert) or those from an educational taxonomy such as Bloom or a Neo-Piagetian level (Gluga, Kay, Lister, Simon, & Kleitman, 2013). These decisions also need to be made with the OLM in mind, so that the values of the learner model components can be presented in a meaningful way to the learner.
- *Design the modeling **processes** that provide **evidence** about each component*. This covers all aspects that drive the modeling. For example, each problem-solving step that the learner takes in an ITS could contribute evidence about the Knowledge Components that represent the knowledge needed for that step.

An OLM is an interface onto any part of such a learner model. There has been a large and diverse range of OLMs, such as those reviewed in (Bull & Kay, 2007), as well as similar work from the Learning Analytics research community where it is described as learner-facing analytics (Bodily, Kay, Aleven, Jivet, & Davis, 2018). While an OLM does not need to be a visualization, much of the literature has presented them in visual forms, and many use the conventional bar chart visualization. There has been huge diversity in these visualizations. For example, one delightful OLM within a children's mathematics learning environment (Lee & Bull, 2008) appeared as a charming sketch of a tree; each correct answer made it look healthier, greener and with more leaves. Incorrect answers made it progressively less healthy. But many OLMs appear as much simpler bar charts. A glance across the OLM literature gives a bewildering array of possible OLM forms.

This makes it timely to establish principles that can be used to design OLMs. This chapter aims to do this by drawing on the dual processing model in psychology. There has been wide acceptance of the importance of the dual-processing model for decision making (Tversky & Kahneman, 1974). We are certainly aware that there has been considerable debate about the nature of that model (Evans & Stanovich, 2013). However, we take inspiration from the core notion that thinking is fast or slow (Kahneman, 2011) or on a continuum of faster or slower. This model distinguishes decision processes that need to be fast, efficient, and may be automated, as opposed to those that involve slow, deliberate thought (Kahneman, 2011).

Kahneman's body of work earned him a Nobel Prize in Economics. In our learning context, his core insights are useful for designers of OLMs. This is because the design of the OLM visualization should take account of the nature of decisions the learner needs to make, aided by the OLM visualization. The dual processing model can help designers consider the time that the learner has available to make the decision. By definition, fast decisions need to be made quickly. There are many such fast learning decisions that can be made within a single session in a learning interface. Taking a very simple example, an arithmetic drill application could have a skill-o-meter showing four bar charts, one for each skill in addition, subtraction, multiplication and division. On completing each small task, the learner may be able to see the change in the relevant skill bar in an animation, with the bar growing if the last task was done correctly or shrinking if it was incorrect. Each skill bar could have a target skill level; when this is met the bar could change color. Examples of fast decisions are:

- Did I just make progress? (This can be fast if the learner can easily see an animated growth in one of the skill bars.)
- Do I need to do another addition-problem to reach the addition-target skill level? (Fast to judge when the skill bar changes to a new distinctive color.)

Both these examples involve pre-attentive stimuli (movement of the animation and distinctive color change). These are just the classes of interface elements that have earned dashboards the description of delivering information the learner can understand *at-a-glance*. By contrast, the design constraints are quite different for cases where the learner needs to take the time to carefully think about the information presented (e.g., "I struggle with recognizing loops in C++ when they're too large").

The next section of this chapter identifies key metacognitive processes that an OLM should support within an ITS. It analyses these in terms of whether they involve fast or slow thinking. The following two sections propose design guidelines for each of these. We conclude with a discussion of ways that our analysis can inform future work on designing and evaluating OLMs.

Metacognitive Roles for Open Learner Model Visualizations

Metacognition describes the range of knowledge, skills and processes that a learner uses to reason about their own learning—it describes the higher order knowledge about how to self-regulate learning (Flavell 1979; Schraw, 1998; Panadero, 2017). Education theory has long highlighted the importance of metacognition in successful learning (Flavell, 1979; Schoenfeld, 1987). For example, the influential 1995 meta-analysis of metacognition and feedback (Butler & Winne, 1995) begins with a clear statement that successful learners are able to self-regulate their learning and then provides an analysis of the way that feedback can support a range of important metacognitive processes. This means that OLMs have the *potential* to deliver feedback to learners as a means of support for their metacognitive processes (Zapata-Rivera & Greer 2003; Bull & Kay, 2013; Reimann & Bannert, 2017). A recent review (Hooshyar, Kori, Pedaste, & Bardone, 2019) highlighted the range of ways that OLMs have been shown to support self-regulation and metacognitive processes. Broadly, these relate to higher levels of reflection, self-monitoring progress, and more accurate self-judgements that resulted in students spending more time on tasks, doing better in selecting problems to do, and learning gains.

Theories and models for self-regulated learning highlight that some of these metacognitive processes need to be "automated" as learners unconsciously apply their own metacognitive strategies (Panadero, 2017). But we have not seen previous work that analyzes the notions of fast and slow thinking within metacognitive processes. We consider this as a critical factor in informing the design of OLM visualizations as it determines key features of the OLM interface, such as its visual elements and complexity.

Table 1 shows a selection of key metacognitive skills. For each of these, we have categorized them in terms of whether they involve thinking fast or slow. For example, the first row relates to self-monitoring progress. We will discuss this particular metacognitive process in some detail. This is both to draw on the long history of work on OLMs in this role and because it illustrates the character of both fast and slow thinking involved. We will then discuss the other metacognitive activities and skills in the table, and how they involve just fast, just slow, or both forms of thinking. This provides the foundations for the later discussions of the implications for designing visualizations for each.

Table 1. Mapping key metacognitive skills to the form of thinking, fast or slow, that OLM visualizations need to support.

Metacognitive role for OLMs	**Fast**	**Slow**
Self-monitoring progress	yes	yes
Deciding what learning activity to do next	yes	yes
Monitoring activities and attitudes of others in collaborative or competitive learning activity	yes	yes
Self-reflection, reflection on collaborative or competitive learning activity	no	yes
Planning	no	yes

Self-monitoring of progress means that a student can answer questions such as:

1. Am I making progress?
2. Have I met my current goal?

Such self-monitoring is already widely supported in the diverse forms of feedback provided by learning applications. One ubiquitous and simple example is the Multiple Choice Question (MCQ). For example, these may be embedded in Massive Open Online Course (MOOC) lectures or online teaching systems. In such cases, they commonly appear at the end of a short period of direct instruction. The MCQ can drive several metacognitive processes. The simple act of stopping to attempt an MCQ gives the learner an opportunity to pause and reflect on whether they have learnt enough to tackle the question. If they proceed to do a question, they learn whether the system assesses their answer as correct, and that contributes to self-monitoring. If the system rates their answer as incorrect, there is also a chance to self-monitor, perhaps to plan a remediation strategy, such as replaying a video.

There is a huge body of research on how to provide effective task level feedback, for example (Shute, 2008). An OLM visualization can provide a quite different level of support for the learner to assess their progress and to judge whether they are meeting their goals. We illustrate this with an example from the ELM-ART ITS. It has a special place in OLM research as it is part of one of the earliest systems (Brusilovsky, Schwarz, Weber, 1996) and it is still in operation today. Figure 1 shows two recent screenshots of the upper left of its main screen, where it delivers personalized instruction on the programming language, Lisp. The left screenshot shows the OLM when the learner first starts the system. Lesson 1 is green to indicate the learner should do it. The other lessons are red indicating the learner does not know enough to tackle them. After the student has done some of the content in Lesson 1, the OLM changes to reflect the progress as in the screenshot at the right. This now reveals Learning Components within Lesson 1.

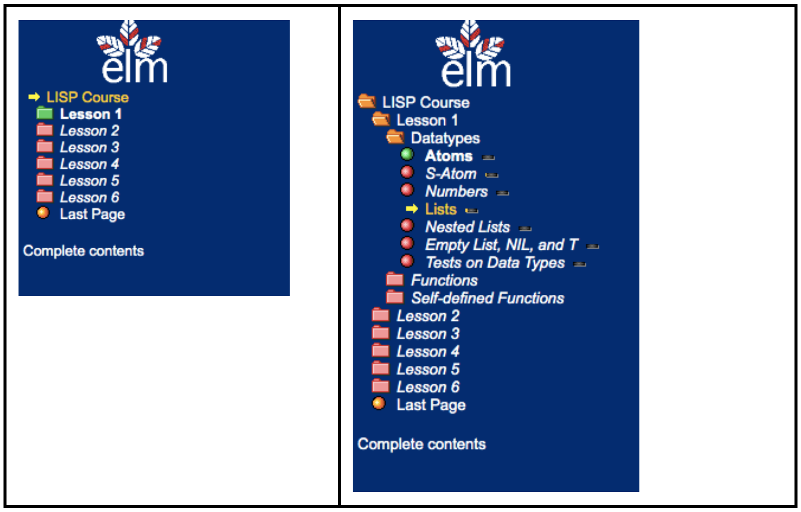


Figure 1. ELM-ART open learner model that shows navigation recommendations and progress. Reprinted with permission. Screenshot from *http://art2.ph-freiburg.de/Lisp-Course*

This OLM is always visible at the ELM-ART teaching interface. As a student successfully completes each learning activity, this may cause the OLM to change—the learner can see that they are making progress. This is an example of fast, at-a-glance self-monitoring of progress. It also provides fast self-monitoring of progress against a goal, such as wanting to complete enough of Lesson 1 to make Lesson 2 turn green.

Other parts of this OLM support slow scrutiny of the learner model. For example, a student would need to take a break from their learning to read the names of each of the components visible in the right screenshot. This, too, can be part of a slow form of self-monitoring progress, now against the detailed concepts that need to be mastered.

Figure 2 shows another example, the current, early version of the OLM in the CSEdPad (CS Education Pad; a National Science Foundation project led by Vasile Rus and Peter Brusilovsky) which aims to teach code comprehension processes, a critical skill for both learners and professional programmers. This OLM has one cell for each major topic in CS1 and CS2 (intro to programming courses). This is shaded to indicate the level of mastery of each Knowledge Component in the learner model.

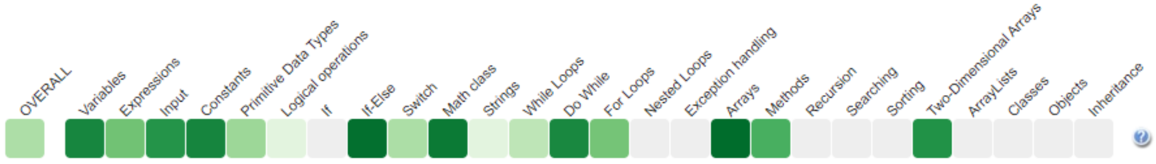


Figure 2. OLM in the CSEdPad (CS Education Pad)

At one level, this is a quite simple visualization of learning progress, where a learner can readily see that they are doing well on some of the topics and less well in others. Although this view of the OLM is simple and intuitive, it would be difficult for a learner to quickly see changes in their progress after each learning activity—a successfully completed task may cause a few of the cells to become a little darker. So, this form of OLM, with many elements and depicting several learning levels is well-designed for slow thinking. Faster use may be possible if a learner uses this OLM in parallel with learning activities labelled just as in this OLM. For example, suppose their slow use involved scanning from the left to find the first topic that had no progress. Based on this strategy, they would select “If” as a learning target. But then, when they successfully completed a learning activity on this topic, they may be able to quickly monitor that “If” cell.

It should be noted that students can hover over and click each square/topic to get a more detailed analysis of their performance on each topic. Such an action would indicate a more deliberate, slow thinker as they are taking a deliberate step to better understand their performance with respect to one particular topic. A social aspect is also being used in the CSEdPad project, i.e., performance of peers is summarized for the learners for comparison, serving motivational purposes.

A very similar OLM by Guerra-Hollstein, Barria-Pineda, Schunn, Bull, and Brusilovsky (2017) provided two more parallel heatmap bands, one showing the average performance of a peer group and another showing the student compared with the group. These may provide a student with a comparative answer to the second question above, because the student can use this to determine if they are ahead of the class. This work builds on the earlier exploration of social student models (Brusilovsky et al., 2015).

Table 1 lists several other metacognitive processes that can play an important role in learning. Like self-monitoring, the decision about what task to do next can involve either fast or slow thinking, where each case puts different design requirements on the OLM visualization. The ELM-ART interface illustrates a way to help the learner decide what to do next. Suppose the learner has just come to ELM-ART for the first time. A fast decision on the next task to do can be made by using the heuristic of clicking the first green topic, Atoms in the right screenshot in Figure 1. A learner might use slower thinking if they consider that they already know many Lisp topics that they learnt outside ELM-ART. So the color coding is for a model based on the mastery that the student has demonstrated within this system. This student cannot use it directly and instead needs to look at topics, then consider which they believe that they know and then select one they want to learn next.

A well-designed OLM could support fast decisions during a learning activity. When the learner is deeply engaged in a task, particularly in a deep flow state (Csikszentmihalyi, 1991), they are unlikely to pay any attention to the OLM at all. If, however, they are less engaged or there is a break in the flow, they may make many micro-decisions about whether to continue working on that task or not. Consider the case of a learner who decided to ignore the navigation advice of the ELM-ART OLM and they are working on a topic coded red. This coding reflects the learner model's belief that the student is not ready for that topic; for example, they may not have demonstrated mastery of the prerequisite knowledge. In that case, the OLM sits in the periphery of the learner's vision as an on-going reminder that the topic is marked red. If the learner finds the task difficult, for example submitting answers that are graded as incorrect, a glance at the OLM could help them decide to reconsider continuing with that topic. Once that fast decision has been made, there would follow a decision about a more appropriate task to tackle. This may involve considering the topics in the OLM carefully, expanding lessons and topics to identify a better task to do next.

The third row is another form of monitoring, now involving other learners. OLMs could support important fast thinking monitoring in synchronous collaboration, especially where the learner could be alerted to important changes in the learner models for other students. Slow thinking could deal with more complex reasoning needed to track longer term changes as well as larger groups of students.

The last two rows can also be supported by OLMs but these activities of reflection and planning call for slow thinking. In these cases, there is potential for far more complex visualizations, for example, covering long time periods or multiple data sources. Both reflection and planning are cognitively demanding. In the case of reflection, students need to make sense of data to discover insights about themselves. Reflection has been argued to be particularly important for effective learning (Schön, 1938) although the term covers diverse processes, with different interface designs being appropriate for each.

Design Guidelines for OLMs that Support Fast Thinking

The design of an OLM visualization should draw on the body of research on formative feedback about the student's progress at the task level (Shute, 2008; Hattie, 2012). For example, Shute (2008) provides a rich set of guidelines for providing task-level feedback. These include recommendations on their design and timing as well as the benefits of personalization, particularly for low versus high performing learners. There are also guidelines to avoid potential negative effects, such as interrupting the learner's concentration or giving vague normative scores compared with peers.

The design of the visualization for fast thinking metacognitive processes needs to ensure that there is little cognitive demand and that the learner can see the salient information at-a-glance. This calls for:

- Simple interfaces;
- Few visual elements;
- Visual features that are readily perceived (Bertin, 1983).

The best example of this is a simple and moving visual element. This is readily noticed even in the peripheral vision. This means that a learner who is concentrating on a learning activity will still notice the information.

Fast thinking also includes cases where the learner wants to quickly check their progress, perhaps to gain reassurance they are on the right track and progressing. In that case, the interface should enable the learner to easily see the key information for fast decisions such as those in the first three rows of Table 1.

We draw on work that linked the dual-processing model for decision making with design of visualizations (Padilla, Creem-Regehr, Hegarty, & Stefanucci, 2018). After a comprehensive review, they conclude with the following recommendations for design of visualizations for fast thinking:

- Identify what information is critical;
- Use a visual encoding technique to direct attention to that information;
- Design visualizations to match the learner's mental schema and the current task demands;
- Minimize the calculations and transformations needed to make sense of the information;
- Take account of individual levels of graphic literacy and numeracy (Lallé & Conati, 2019).

Gamification and persuasive interfaces also make use of fast, unconscious thinking. There is a substantial body of research in gamification in educational contexts (Dicheva, Dichev, Agre, & Angelova, 2015) and it is widely used in commercial applications. It has many dimensions and some of these overlap with the goals of OLMs. Similarly, there has been considerable work on persuasive interfaces (Fogg, 2002) that make use of diverse and often subtle ways to influence users. Both of these have been so widely used that people are familiar with the elements and would understand them in an OLM. This is a rich area to explore.

Design Guidelines for OLMs that Support Slow Thinking

There are two main approaches to support users in slow thinking. One is to create far more complex OLMs that reveal things that we never expected to see (Tukey, 1977). The other approach is to create OLMs that slow the user down. For example, four COVID-19 simulations[1] were designed to help people understand how exponential growth can cause a virus to spread quickly. Each entice the viewer to pause, watch the way each plays out, and have time to really consider the differential impact. This latter approach is likely to be important for complex visualizations. It may also help reduce illusions that come from the fast thinking mode, such as the illusion of mastery.

There are many ways to help the learner cope with complex OLMs that need thinking slowly. A *scaffolding* interface may play a valuable role to help the student consider and see the relevant information in the OLM (Tang & Kay, 2018). There are many possible other approaches that slow the learner down and encourage them to reflect. For example, the student could be asked to self-assess their knowledge. If the learner believes they know something but the learner model indicates they do not know it, the OLM interface could invite the student to negotiate with it to convince it (Bull, Ginon, Boscolo, & Johnson, 2016). Different guidance mechanisms can be used to support student interaction with OLMs (Zapata-Rivera & Greer, 2003). In the user studies of the scaffolding in Tang and Kay (2018), participants called for personalized recommendations to match their individual goals and knowledge. This is a place for Artificial Intelligence (AI) driven recommendations, and these may also advise learners to shift between fast and slow modes of interaction with their OLM.

Similarly, self-assessment can be combined with an OLM as in a math equation solving tutor (Long & Aleven, 2017). In that work, the learner was asked to self-assess before they saw the learner model. This meant that they needed to pause and reflect to decide the self-assessment. Having done that, they could

[1] Why outbreaks like coronavirus spread exponentially and "how to flatten the curve", by Harry Stevens, https://www.washingtonpost.com/graphics/2020/world/corona-simulator/ visited March 14, 2020

consider the new information in the OLM. This approach provided learning benefits over an OLM without this approach to slowing down the learner before they saw the OLM.

Our final example of a way to slow down learners as they use an OLM, is the work of Al-Shanfari, Demmans Epp, Baber, and Nazir (2020). They created a system where the student self-assesses both their knowledge and their confidence in that assessment. Only after this step were they shown the OLM, with the opportunity to see how well this matches their self-assessment. The introduction of certainty offers ways to interpret the data that comes from this process. For example, suppose two learners, A and B, assess the knowledge level as high and their confidence in that assessment also as high. If A's learner model indicates low knowledge, this indicates A is unaware of their lack of knowledge. This may be due to a flawed fast thinking mode of student A, giving them an illusion of mastery (Graesser, D'Mello, & Person, 2009). If B's OLM indicates their knowledge is actually high, it confirms the self-assessment.

Discussion of Implications for GIFT and Conclusions

For a framework like GIFT, it is important to create tools that enable an ITS author to select the right form of visualization for the metacognitive processes intended. We recommend that the system provide an interface for the author of a new teaching system to combine the overall authoring with the design of the OLM. They can start the design of an OLM by selecting the metacognitive process they want the OLM to support. This could be based on just a subset of those listed in Table 1. This involves a complex design process where the teacher needs to decide how the visualization and learning data available in the OLM should fit into the overall learning experience. The interface can help the teacher with this design process. For example, it could provide a checklist of issues to consider, highlighting the options available in terms of:

- the metacognitive process that the OLM is particularly designed for;
- whether the OLM has been designed for fast or slow thinking; and
- the nature of the learning activities that the student is engaged in and how well that matches the OLM.

Then, for OLMs that are intended to support thinking fast, the authoring interface should guide the choice of which elements to display, following the principles above. This covers the classic dashboard intended for at-a-glance use. It would be useful to design even these OLMs so that the system can readily track the learner's use of them. For example, at a very modest time cost for the learner, the OLM could remain static until the learner clicks it. Only then would it show the changes since the last click. Even in an OLM with the complexity of that in Figure 2, this strategy may well enable the learner to quickly see changes. It would have the side-benefit that the system could capture data about the use of the OLM to build a model of the learner's use of it. This could provide a valuable way to model the learner's metacognitive activity.

The case of OLMs for thinking slowly offers many more possibilities. For GIFT, the simplest and still useful case is to support authoring of OLMs that are more complex. This could be combined with one of the simplest ways to help a learner slow down productively; this is a combination of self-assessment widgets with the OLM (Long & Aleven, 2017; Al-Shanfari et al., 2020). Simple scaffolding questions are another promising approach, as they are both simple and generic, with the author needing to identify the questions for the learner to consider to help make sense of a complex visualization (Tang & Kay, 2018). The whole reflective process is itself a potential role for tutoring. This can operate at two levels. One of these is the domain of the core teaching system. In Tang and Kay (2018), where the user was trying to understand their own personal data about physical activity, one question asked them to consider the level of activity on weekends versus weekdays. This is because public health research has established that people have very different physical activity patterns between these times and many people are less active on weekends without realizing that this is the case. So, asking a learner to consider this question will enable them to look at the OLM to discover the answer as it applies to them. There is certainly an opportunity to also provide information about public health research that underpins the question. The second level of tutoring to support reflection concerns learners' awareness of their own metacognitive skills. For example, this could include support for decisions about how often to take time for slow reflection on long-term data, what goals to set and what plans are needed to achieve them.

It would be valuable to guide the author of the tutoring system in determining whether the OLM should support fast or slow thinking at each point in the use of the ITS. This could be based on the broad classes of thinking described in Table 1. The guidance could help the teacher consider:

- Which metacognitive processes should the OLM support at this point?
- Does this call for fast or slow thinking at this point in the activity?
- Then the author could be offered a small set of OLMs which match these needs; for example, a progress monitoring, fast decision could be based on a histogram where changes in the OLM are animated so that they are readily noticed.

One important class of GIFT tutoring systems makes use of simulations, such as the first-person shooter game VBS3, used for training small army unit tactics. In that context, OLMs for thinking slowly can support student reflection based on performance on simulations, e.g., in an after-action review session. Just as in other systems, carefully designed OLMs for thinking fast might help learners as they use the simulation. For example, the OLM could help the learner to monitor progress. But there is a particular value for a slow thinking reflection phase after a simulation, especially if it is very engaging and immersive as in virtual reality. The OLM can provide a valuable consolidation of the activity, ensuring that the learner links the activities they did during the simulation with the key learning goals for which it was designed.

ITSs have evolved over several decades, and learner models with them. Though the original design intent of crafting individualized instruction remains, there is increasing focus on broader roles for ITSs. An important aspect of human tutoring is to build curiosity, build motivation, and to ask leading questions that stimulate thought. Even today, these aspects of technology-based tutoring have been relatively ignored compared to mastery-based interventions, and OLMs should be designed to help make up for this omission. Research demonstrates that deliberately involving learners in their own instruction provides substantial benefits that demand the adoption of OLMs. This means that a learning application authoring framework like GIFT should explicitly support authors in two quite different aspects of learner model design.

This is the first work that we are aware of that links the dual-processing model of cognition with the design of OLMs. There is considerable discussion of dashboards that refers to them as being used at-a-glance. This corresponds to designing for thinking fast because these should be simple enough to understand and make use of without effortful cognition. This is the case for on-going self-monitoring and for micro-decisions such as deciding the next task to do, by selecting the first green one in a sequence. However, we believe it is important to distinguish these from more sophisticated OLM visualizations for thinking slow. There are many potential design options for these, with scaffolding and interface elements designed to help the learner slow down for metacognitive processes such as reflection and planning.

This chapter has had a tight focus on exploring the OLM design implications of the dual processing model of thinking. However, we emphasize that any such principles are just the first stage of exploring the design space. We advocate that all interfaces be subject to rigorous evaluation in well-designed user studies. In line with this, OLMs (both for fast and slow thinking) should be evaluated with the intended users to make sure users correctly interpret and use the information provided (Zapata-Rivera, Graesser, Kay, Hu, & Ososky, 2020).

References

Al-Shanfari, L., Epp, C. D., Baber, C., & Nazir, M. (2020). Visualising alignment to support students' judgment of confidence in open learner models. *User Modeling and User-Adapted Interaction*, *30*(1), 159-194.

Bertin, J. (1983) Semiology of Graphics; Diagrams Networks Maps. No. 04; QA90, B7.

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018, March). 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).

Brusilovsky, P., Schwarz, E., & Weber, G. (1996, June). ELM-ART: An intelligent tutoring system on World Wide Web. In *International conference on intelligent tutoring systems* (pp. 261-269). Springer, Berlin, Heidelberg.

Brusilovsky, P, Somyürek, S, Guerra, J, Hosseini, R, Zadorozhny, V & Durlach, P J. (2015). Open social student modeling for personalized learning. *IEEE Transactions on Emerging Topics in Computing, 4*(3), 450-461.

Bull, S., Ginon, B., Boscolo, C., & Johnson, M. (2016, April). Introduction of learning visualisations and metacognitive support in a persuadable open learner model. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 30-39).

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. (2013). Open learner models as drivers for metacognitive processes. In *International handbook of metacognition and learning technologies* (pp. 349-365). Springer, New York, NY.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, *65*(3), 245-281.

Csikszentmihalyi, M. (1991). Flow: The psychology of optimal experience (Vol. 41)." HarperPerennial New York.

Dicheva, D, Dichev, C, Agre, G & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society*, *18*(3), 75-88.

Evans, J. S. BT, & Stanovich, KE (2013). *Dual-process theories of higher cognition: Advancing the debate. Perspectives on Psychological Science*, *8*(3), 223-241.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.

Fogg, B. J. (2002). Persuasive technology: using computers to change what we think and do. *Ubiquity*, *2002*(December), 2.

Gluga, R., Kay, J., Lister, R., Simon, & Kleitman, S. (2013). Mastering cognitive development theory in computer science education. *Computer Science Education*, *23*(1), 24-57.

Graesser, A. C, D'Mello, S & Person, N. (2009). Meta-Knowledge in Tutoring. In D J Hacker, J Dunlosky & A. C. Graesser (Eds.), Handbook of metacognition in education (pp. 361-412). Mahwah, NJ: Erlbaum.

Guerra-Hollstein, J., Barria-Pineda, J., Schunn, C. D., Bull, S., & Brusilovsky, P. (2017, July). Fine-grained open learner models: Complexity versus support. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization* (pp. 41-49).

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

Hooshyar, D., Kori, K., Pedaste, M., & Bardone, E. (2019). The potential of open learner models to promote active thinking by enhancing self-regulated learning in online higher education learning environments. *British Journal of Educational Technology*, *50*(5), 2365-2386.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Lallé, S., & Conati, C. (2019, March). The role of user differences in customization: a case study in personalization for infovis-based content. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 329-339).

Lee, S. J., & Bull, S. (2008). An open learner model to help parents help their children. *Technology Instruction Cognition and Learning*, *6*(1), 29-51.

Long, Y., & Aleven, V. (2017). Enhancing learning outcomes through self-regulated learning support with an Open Learner Model. *User Modeling and User-Adapted Interaction*, *27*(1), 55-88.

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: a cognitive framework across disciplines. *Cognitive research: principles and implications*, *3*(1), 29.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology*, *8*, 422.

Reimann, P., & Bannert, M. (2018). Self-regulation of learning and performance in computer-supported collaborative learning environments.

Schoenfeld, A. H. (1987). What's all the fuss about metacognition. *Cognitive science and mathematics education*, *189*, 215.

Schön, D. (1938). The reflective practitioner. *New York*, *1083*.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional science*, *26*(1-2), 113-125.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189.

Sottilare, R. A., Baker, R. S., Graesser, A. C., & Lester, J. C. (2018). Special Issue on the Generalized Intelligent Framework for Tutoring (GIFT): Creating a stable and flexible platform for Innovations in AIED research. *International Journal of Artificial Intelligence in Education*, *28*(2), 139-151.

Tang, L. M., & Kay, J. (2018, July). Scaffolding for an olm for long-term physical activity goals. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (pp. 147-156).

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131-160).

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

Zapata-Rivera, D., Graesser, A., Kay, J., Hu, X. & Ososky, S. (2020). Visualization Implications for the Validity of ITS. In Design Recommendations for Intelligent Tutoring Systems: Volume 8 – Data Visualization. US Army Combat Capabilities Development Command (DEVCOM) Soldier Center.

Zapata-Rivera, J. D., & Greer, J. (2003). Analyzing student reflection in the learning game. In S. Bull, P. Brna & V. Dimitrova (Eds.) Proceedings of Workshop on Learner Modelling for Reflection, Supplementary Proceedings of the 11th International Conference (Vol. 5, pp. 288-298).

Section III – Technical Applications and Data Visualization

Dr. Anne M. Sinatra, Ed.

Chapter 16 – Introduction to Technical Applications for Data Visualization

Anne M. Sinatra
U.S. Army Combat Capabilities Development Command (DEVCOM) – Soldier Center – Simulation and Training Technology Center (STTC)

Core Ideas

The chapters in this section cover a wide range of technical applications for data visualization. Some of the chapters are framed around direct suggestions and recommendations for the Generalized Intelligent Framework for Tutoring (GIFT), whereas others are focused on technology applications that exist and how their lessons learned can be utilized to improve GIFT. The wide range includes discussion of how to implement existing processes such as xAPI and GIFT, and even lessons learned from looking at the gaming industry.

The first set of chapters (Mullins, Kay, & Conati; Rus, Swiecki, Cockroft & Hu; Brawner & Ososky) are applied examples including lessons learned and suggestions, and the second set of chapters (Blake-Plock, Hoyt, Casey, & Zapata-Rivera; Brihoum, Heylum, Kalaf, Meyer, & Woodman; Sinatra, Ososky & Johnson; Goldberg, Hoffman, & Graesser) are specific recommendations within the context of the GIFT framework.

Individual Chapters

Mullins, Kay, and Conati provide an overview review of artificial intelligence (AI) and machine learning (ML) focused data visualization system approaches and implications. They highlight that some of the challenges of implementing AI/ML systems, which include that humans sometimes do not understand how the systems come to the conclusions that they do. The authors suggest that it is important for work to continue into the field of Explainable AI (XAI). In their chapter the authors review work that has been done in the field of XAI as it relates to training outcomes, as well as the implications for data visualizations and output. Finally, the authors provide recommendations including how the lessons learned from these systems can be applied to GIFT and other ITSs.

Rus, Swiecki, Cockroft, and Hu describe two specific data visualization techniques that can be implemented in intelligent tutoring systems to visualize dialogue interactions. These two visualization techniques are sequence logo analysis and heatmap visualizations. The authors discuss both techniques in detail through the lens of examples of specific implementations of the approaches in ITS research. They further discuss how these techniques could be applied in GIFT, and the design implications of doing so.

Brawner and Ososky highlight approaches for data visualization that have been used in the video game industry, and suggest ways that they can be applied to assist in creating data visualizations. A specific example that is provided is understanding the way that information is displayed in a game environment, and how similar techniques could be used in an intelligent tutoring system to convey meaning. They highlight the importance of understanding who the intended user of the visualization is, and planning accordingly to meet the relevant needs of the user. They also make the important point that games and simulations provide a large amount of data, and visualizations need to be designed specifically to narrow down the data to only what is relevant to the user. Further, the authors provide specific design recommendations and guiding questions to keep in mind while designing data visualizations for users.

Blake-Plock, Hoyt, Casey, and Zapata-Rivera discuss considerations and benefits of developing an xAPI (Experience API) ontology that can be used to implement a consistent xAPI Profile specification in GIFT. The approach discussed also has additional application for ITSs in general, as consistency between data in systems could lead to additional interoperability. The chapter discusses the Data Analytics and Visualization Environment (DAVE) framework, as well as the applications and lessons learned from the project. The authors highlight how the processes of the DAVE framework align with the capabilities of GIFT, and how they could potentially be integrated. The authors also discuss how the use of consistent xAPI Profiles can assist in producing data visualizations that are useful and informative.

Brihoum, Heylmun, Kalaf, Meyer, and Woodman report the process of implementing data analytics software with GIFT Cloud. They describe how an open source data analytics tool, Piwik, was used with GIFT, and the types of output that it provides. They discuss the types of aggregate data that was able to be collected with this tool, and how general patterns of user interactions were discovered by examining the data. They provide suggestions on understanding user behaviors as they interact with a system, and how the visualizations of this data can assist with making design decisions for systems.

Sinatra, Ososky, and Johnson discuss the differing needs in dashboards for potential user roles in GIFT. Generally GIFT has three types of users: Learners, Instructors, and Researchers. The interfaces and information that each of these groups need have some overlap, but also significant differences. The authors suggest identifying the needs of each user group, and conducting tasks analyses in order to develop dashboards. The chapter discusses a similar process that was used for a Navy training system which was not created with GIFT, but is highly relevant to the current topic. The outcomes and lessons learned from the previous training experiences help to support recommendations for the development of user role dashboards and data visualizations in GIFT.

Goldberg, Hoffman, and Graesser provide background on some of the driving requirements behind current GIFT development, and recent efforts to include human-in-the-loop functionality. Their chapter discusses the need and subsequent development of GIFT’s Game Master interface that allows a human observer to input information directly into GIFT. The chapter discusses the functionality of Game Master, how the user interacts with the system, and how data is input into GIFT. The chapter concludes with a description of a way forward and intended improvements that are expected to be made to the Game Master in the future.

Chapter 17 – Enabling understanding of AI-enabled Intelligent Tutoring Systems

Ryan Mullins[1] and Cristina Conati[2]
Aptima, Inc.[1], The University of British Columbia[2]

Background

The period between January 2015 and January 2020 has seen rapid, massive advancements in the fields of artificial intelligence (AI) and machine learning (ML). Some high-profile examples in the media include: *Project Maven*, the culmination of decades of research in computer vision, classification, and pattern recognition work, this AI/ML system is fielded across the Department of Defense (DoD) to aid in processing full-motion video and imagery data (Pellerin, 2017); *Cyber Reasoning Systems* (CRSs; Avgerinos et al., 2018; Nguyen-Tuong et al., 2018), these autonomous machines demonstrated that AI can find, prove, and mitigate vulnerabilities in software (Walker, 2015), with more recent efforts integrating CRSs with humans as teammates (Fraze, 2018); *AlphaGo*, the first AI to record a win over a human grand master in the game Go (Wang et al., 2016); and *GPT-2*, what many consider to be the first believable generative language model (Radford et al., 2019).

These examples highlight how AI/ML can be leveraged in what the DoD would call an *operational* setting; the use of technologies to directly achieve mission objectives. AI/ML has similarly revolutionized the *training* world. The *Force Fitness Initiative* within the United States Marine Corps is driven by AI-enabled technologies to predict injuries and dynamically adapt planned physical training activities (Clark et al., 2019). *Intelligent tutoring systems* (ITSs) have been enhanced with adaptive technologies to guide both content (Sottilare, Goldberg, Brawner, & Holden, 2012) and feedback (Carlin, Nucci, Oster, Kramer, & Brawner, 2018) selection for individual trainees.

AI capabilities will continue to evolve, and the scope of their influence will expand, in the coming years. Some argue that AI/ML capabilities, collectively, have achieved a level of maturity that justifies establishing an interdisciplinary field of *artificial science* to study and explain machine behavior (Rahwan et al., 2019). This will be necessary to understand how AI/ML influence the methods, measures, and metrics of success in training and tutoring. However, a problem exists with many current AI/ML methods: humans are frequently unable to understand why an AI/ML exhibited a behavior (Edwards & Veale, 2017). Explainable artificial intelligence (XAI) is a line of research that seeks to enable such an understanding. Efforts to create XAI have existed in the training domain for more than 15 years (Lane, Gore, Van Lent, Solomon, & Gomboc, 2005), but the pace of advancement was recently hastened through investment by the Defense Advanced Research Projects Agency (DARPA). Gunning (2017) outlines four explanatory modes for XAI–statements, visualizations, cases, and rejections that are enabled by architectures that include explainable models and explanation interfaces.

In this chapter, we explore how the visualization mode of XAI can and will impact the design and use of the Generalized Intelligent Framework for Tutoring (GIFT) ITS framework in the future. We assume that the AI is integrated into GIFT as part of the instructional design, where it can influence the selection of content and feedback during the tutoring process, as described in the examples above. This chapter summarizes the findings of a literature review of measures and metrics for assessing the explainability of AI, visualizations that can be used for explanatory purposes, methods for evaluating the effectiveness of XAI visualizations, and explanations to improve training and tutoring outcomes. We present a set of design recommendations that integrate these findings, and situate these recommendations in specific use cases found in the literature.

Methods

This chapter presents the results of an integrative and selective literature review that explores *measures and metrics* of explainability and *visualizations* enabling explainable interfaces. Data were collected from open-access publication outlets, including both peer-reviewed and non-peer-reviewed content as appropriate.

Results

Measures and Metrics of Explainability

As part of DARPA's XAI program, Mueller, Hoffman, Clancey, Emrey and Klein (2018) have prepared the definitive text defining explainability in XAI systems available at this time. Here, we situate their analysis in the context of ITSs, emphasizing key design considerations for visualizations enabling understanding of AI behaviors.

Trust

Hoffman, Mueller, Klein, and Litman (2018) describe trust as an active process under constant tension between positive and negative sentiments. Should trust reach a sufficiently negative sentiment, it may be impossible for the system to recover, which is to say impossible for the user to believe and act upon the output of the system (Dzindolet, Peterson, Pomranky, Pierce, and Beck, 2003; Madhavan & Wiegmann, 2007). In AI-enabled systems, maintaining trust requires interaction between the human and the AI components, with each providing and acting upon the feedback given to them. For the AI in particular, trust measurement instruments must account for undulations between positive and negative sentiments (Sarter, Woods, & Billings, 1997), requiring that the AI provide discrete, tangible, and predictable changes to the feedback it provides to allow human users to anticipate and become accustomed to variations in the AI's function. Hoffman, Mueller, Klein and Litman (2018) also note that the trust measurement literature relies heavily on scale-based instruments that variably emphasize *trust* (defined as believing that the AI is correct) and *reliance* (defined as following the AI's recommendation) as independent dimensions of a static quality of a human-AI relationship (Adams, Bruyn, & Houde, 2003; Johnson, 2007). Depending on the nature of the ITS, static measures of trust may or may not be appropriate. For example, from the perspective of a learner, adding a self-assessment that includes some assessment of the learner's trust in the ITS to the end of a lesson or series of lessons may be appropriate, and is likely to generate a useful response rate. On the contrary, it is unlikely that such a scenario would present itself when considering the trainers' perspective. Therefore, more frequent measurement is required, such as passive interaction monitoring that enables aggregation and inference to estimate trust.

Reasoning

Research shows that explanations most frequently cite causal reasoning about the mechanisms driving events (Ahn, Kalish, Medin, & Gelman, 1995; Klein, Rasmussen, Lin, Hoffman, & Case, 2014; Murphy & Medin, 1985; Pearl, 2014; Pearl & Mackenzie, 2018; among others). Abductive reasoning is also strongly linked to causal reasoning and explanations (Harman, 1965; Overton, 2013) as a process used to derive and examine causal mechanisms. Counterfactual reasoning is commonly used to compare typical and atypical events to understand and communicate differences and patterns therein, and/or differences and patterns in their causal mechanisms (Byrne, 1997; Woodward, 2003; among others). Analogy is often used to relate events, individually or in the aggregate, to each other or to other examples in context (Clement, 1988; Hoffman, 1995; Spiro, 1988). Causal, abductive, counterfactual, and analogical reasoning are all methods that allow the human to engage directly in the reasoning process by constructing, exploring, and refining models to explain events. Measurement of reasoning should focus on human engagement in one or more of these processes through interaction classification (Bruni, Lucia, Cummings, & Ford, 2019; Fouse, Mullins, Ganberg, & Weiss, 2017).

Visualization Methods

Much of the foundational research underlying the measures and metric of explainability rely on visualizations of models to enable reasoning, understanding, and comprehension. Here we discuss three types of visualizations—graphs, trajectories, and comparisons—that can enable human-AI interaction and collaboration and engender understanding about the decisions made or recommended by AI.

Graphs

Causal reasoning mechanisms are often depicted as graphical structures. Structural equation models (SEMs; Hox & Bechger, 1998) are presented as node-link diagrams, with labeled nodes and edges. SEMs have seen a recent resurgence in the AI field, as Halpern and Pearl (2001) and Pearl and Mackenzie (2018) cite their potential for encoding and representing the behavior of AI. Further, SEMs lend themselves well to computationally efficient representations, such as Microsoft Excel workbooks (Keith, 2014) or graph databases (Icoz, Sanalan, Ozdemir, Kaya, & Cakar, 2015). More recent work has built tools to enable standards-based visualization of SEMs at scale (Epskamp, 2015; Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012). Concept maps are another graphical representation used to represent complex relationships in content (Davies, 2011). Concept maps have been studied extensively in the literature (Moon, Hoffman, Novak, & Canas, 2011), including the skills associated with expert concept mappers working individually and in teams (Moon, Hoffman, Eskridge, & Coffey, 2011), and the potential for information loss and challenges of information recovery as a result of knowledge encoding (Hoffman, Feltovich, & Eccles, 2007). Nobre, Meyer, Streit, and Lex (2019) provide a review of additional methods for graph visualization, including node-link layouts beyond SEMs and concept maps, tabular layouts (such as adjacency matrixes, quilts, and bio fabrics), and implicit tree layouts. Graph visualizations provide a tangible representation particularly well suited to the exploration of relationships in data, but with two key limitations: (1) they have issues with readability when dealing with large volumes of data (Ghoniem, Fekete, & Castagliola, 2004), requiring the ability to collapse or expand subgraphs to improve navigation and sense-making; and (2) they require explicit and rigid semantics and syntactics in order to encode data and translate interactions into meaningful manipulations that can be understood by the AI (Marks, 1991).

Trajectories

Tracking learning progress is an essential function of an ITS that becomes even more critical when transitions are managed or influenced by AI. Trajectory visualizations are used to describe movement through a space over time. Often, the space being visualized is geographic, such as the movement of maritime objects in a port of call (Scheepens, Willems, van de Wetering, & van Wijk, 2011), plumes of gas following an explosion (Tominski, Schumann, Andrienko, & Andrienko, 2012), or individuals moving through a building (Meghdadi & Irani, 2013). Movement is often represented as a line with segments for each discrete time step captured in the data. Additional glyphs and/or dimensionality may be provided to indicate the directionality (as in Tominski et al., 2012) or changes in attribute value (Andrienko & Andrienko, 2008). Considering the ITS use case, the learner traverses a space defined by content and learning objectives, with discrete lessons bounding different regions in the space. Content and feedback selection are managed as two independent loops (VanLehn, 2016). The learner's location within this space can be represented as a $\langle content, feedback \rangle$ pair, and each interaction the learner has with the ITS (e.g., their lesson outcomes) influences the AI's recommendation for the next location that should be explored by the learner. It is also important to note that learning is imperfect and may require remediation over time (Carlin, Nucci, Oster, Kramer & Brawner, 2018). Visualizing learners' journeys through the learning space may help to reveal patterns and relationships between learners, content, and feedback received, such as individual differences or aggregate preferences.

Comparisons

As noted above, counterfactual and analogical reasoning are often used to compare discrete events to one another. Comparative visualization comprises a family of techniques that enable the direct comparison of data to assess similarities and differences. Examples of comparative visualizations include shared space (Zobel, 2010), side-by-side (also called A-B comparisons; Pagendarm & Post, 1995), and small multiples (MacEachren, Xiping, Hardisty, Guo, & Lengerich, 2003) techniques. Shared space visualizations use multiple instances of the same sign vehicle (MacEachren, 1995), such as lines, to explore relationships between one (Zobel, 2010) or many (Wegman, 1990) attribute relationships within and between events. This enables direct comparison of data in the same space, but can induce issues with readability such as occlusion. Side-by-side and small multiples take a different approach to comparison. The visual space allotted to the data is constant, but each datum is given its own independent space in which to be presented. This technique is commonly used with maps (Griffin, MacEachren, Hardisty, Steiner, & Li, 2006) and other visualization techniques where area is used, or required, to represent attribute values. These methods, and especially small multiples, require significant interaction to enable direct comparison between any two data elements, and they come at the cost of visual space within an interface.

Discussion

It is important, when considering the design of GIFT or any other ITS, to remember that the concepts presented above are meant to be used in concert with each other to provide for different aspects of explainability. Key design implications include:

- ***Classes of visualization can, and likely must, be combined***. No single visualization technique is a panacea for all human cognitive needs. The three classes presented in this review were chosen because they represent three complementary aspects of visualization: graphs excel at conceptual relations, trajectories at communicating processes and progress, and comparisons at enabling the inspection and understanding of similarities and differences. Each can be tailored and integrated to suit the needs of trainers or trainees.

- ***Visualizations should provide interactivity***. If explanation is an inherently exploratory process (Mueller et al., 2019), then the visualizations aiding in that explanation must provide the interactions necessary for exploration of meaning. The traditional mantra is "overview first, zoom and filter, then details-on-demand" (Shneiderman, 1996, p. 337), and GIFT should be no exception. Providing interactivity will engage users in the reasoning process (Pirolli & Card, 2005), engendering trust in the system and its recommendations.

- ***Interactions should be captured and analyzed***. Interactive visualizations provide the necessary framing to enable constant, passive measurement of key concepts of explainability, trust and reasoning. Engagement in the reasoning process, which is to say, interaction with and exploration of the visualizations provided, allow for the collection and measurement of key events, such as acceptance or rejection of a recommendation, or expansion of details for a specific sign-vehicle to explore the meaning in more detail. Inference over reasoning measures can enable estimation of trust. For example, automatic acceptance of all recommendations is indicative of overreliance on the AI. Automatic rejection of all recommendations is indicative of complete distrust in a non-resilient system (Woods, 2017). Ideally, captured interactions would be a mix of acceptance, rejection, and exploration, indicative of a functional human-AI partnership.

- ***Change should be constant***. A healthy human-AI relationship is one of constant, but measured, change (Schurr, Fouse, Freeman, & Serfaty, 2019). XAI is one example where measurable change can improve the qualities of human-AI collaboration, as well as learning and training outcomes.

Recommendations and Future Research

Case 1: Explainability for Content and Feedback Selection

As discussed above, GIFT has the ability to provide AI-based content (Sottilare et al., 2012) and feedback (Carlin et al., 2018) recommendations to trainers and trainees. Throughout a training experience, trainers and trainees may want explanations of why an AI recommended a learning trajectory. One visualization to support this would be based in a trajectory visualization that conveys the path that the trainee has taken through the learning space. Given the trainee's current state, the explanatory visualization would present its recommended next lesson and alternate lessons that it chose not to recommend. Exploration of the recommended and alternate lessons would describe the differences in content, learning objectives, past performance in similar lessons, and future needs in order to achieve the desired or required learning outcomes.

Case 2: Explainability Considerations for Teams

Expanding to XAI in GIFT to support team training requires only a simple modification of the visualization described in Case 1. A shared space representation would integrate individual learning trajectories from the team onto a single learning space to examine aggregate learning patterns and identify individual differences that could lead to increased understanding about individual trainees' needs and behaviors. The visualization would also enable comparison of differences about the AI's recommended trajectories between trainees, to assess its awareness of individual needs and performance characteristics in order to optimize learning outcomes for the team as a whole.

Conclusions

The literature review presented above describes two measurement dimensions—*trust* and *reasoning*—for explainable AI that can be applied to GIFT, and three classes of visualizations—*graphs*, *trajectories*, and *comparisons*—that can be used to present explanations of AI-derived recommendations to the user. The case studies provided illustrate how a limited set of cognitive needs of trainers and trainees working individually or in teams can be addressed by integrating XAI and data visualization into GIFT. Considerable further research is required to understand the specific user needs that should be addressed by XAI, the comprehension, acceptability, and utility of the visualizations used to communicate explanations, and the impact on learning outcomes.

References

Adams, B. D., Bruyn, L. E., & Houde, S. (2003). *Trust in Automated Systems, Literature Review*. Humansystems Incorporated.

Ahn, W. K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*(3), 299.

Andrienko, G., & Andrienko, N. (2008, October). Spatio-temporal aggregation for visual analysis of movements. In *2008 IEEE symposium on visual analytics science and technology* (pp. 51-58). IEEE.

Avgerinos, T., Brumley, D., Davis, J., Goulden, R., Nighswander, T., Rebert, A., & Williamson, N. (2018). The Mayhem cyber reasoning system. *IEEE Security & Privacy*, *16*(2), 52-60.

Bruni, S., Lucia, L., Cummings, P., & Ford, B. (2019). Comparing Modeling Techniques in their Ability to Predict Current and Likely Next Tasks for Cognitive Workers. In *Proceedings of the 2019 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation.*

Byrne, R. M. (1997). Cognitive processes in counterfactual thinking about what might have been.

Carlin, A. S., Nucci, C., Oster, E., Kramer, D., & Brawner, K. (2018, May). Data mining for adaptive instruction. In *Proceedings of the Thirty-First International Flairs Conference*.

Clark, T., Cassani, L., Lucia, L., Smith, A., Ganberg, G., Bharadwaj, V., … Jenkins, C. (2019). FitForce Planner: Data-driven Physical Training Programming for Reducing USMC Musculoskeletal Injuries. Presented at the *Military Health System Research Symposium*, Kissimmee, FL.

Clement, J. (1988). Observed methods for generating analogies in scientific problem solving. *Cognitive Science*, *12*(4), 563-586.

Davies, M. (2011). Concept mapping, mind mapping and argument mapping: what are the differences and do they matter?. *Higher education*, *62*(3), 279-301.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International journal of human-computer studies*, *58*(6), 697-718.

Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech. Rev.*, *16*, 18.

Epskamp, S. (2015). semPlot: Unified visualizations of structural equation models. *Structural Equation Modeling: a multidisciplinary journal*, *22*(3), 474-483.

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1-18.

Fouse, A., Mullins, R. S., Ganberg, G., & Weiss, C. (2017, July). The Evolution of User Experiences and Interfaces for Delivering Context-Aware Recommendations to Information Analysts. In *International Conference on Applied Human Factors and Ergonomics* (pp. 15-26). Springer, Cham.

Fraze, D. (2018). Computers and humans exploring software security. *Defense Advanced Research Projects Agency*.

Ghoniem, M., Fekete, J. D., & Castagliola, P. (2004, October). A comparison of the readability of graphs using node-link and matrix-based representations. In *IEEE Symposium on Information Visualization* (pp. 17-24). IEEE.

Griffin, A. L., MacEachren, A. M., Hardisty, F., Steiner, E., & Li, B. (2006). A comparison of animated maps with static small-multiple maps for visually identifying space-time clusters. *Annals of the Association of American Geographers*, *96*(4), 740-753.

Gunning, D. (2017). Explainable artificial intelligence. *Defense Advanced Research Projects Agency*.

Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British journal for the philosophy of science*, *56*(4), 843-887.

Harman, G. H. (1965). The inference to the best explanation. *The philosophical review*, *74*(1), 88-95.

Hoffman, R. R. (1995). Monster analogies. *AI magazine*, *16*(3), 11-11.

Hoffman, R. R., Feltovich, P. J., & Eccles, D. W. (2007, October). The cost of knowledge recovery: a challenge for the application of concept mapping. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 51, No. 4, pp. 328-331). Sage CA: Los Angeles, CA: SAGE Publications.

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable ai: Challenges and prospects. *arXiv preprint arXiv:1812.04608*.

Hox, J. J., & Bechger, T. M. (1998). An introduction to structural equation modeling. *Family Science Review*, *11*, 354-373.

Icoz, K., Sanalan, V. A., Ozdemir, E. B., Kaya, S., & Cakar, M. A. (2015). Using Students' Performance to Improve Ontologies for Intelligent E-Learning System. *Educational Sciences: Theory and Practice*, *15*(4), 1039-1049.

Johnson, D. S. (2007). Achieving customer value from electronic channels through identity commitment, calculative commitment, and trust in technology. *Journal of interactive marketing*, *21*(4), 2-22.

Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. Routledge.

Klein, G., Rasmussen, L., Lin, M. H., Hoffman, R. R., & Case, J. (2014). Influencing preferences for different types of causal explanation of complex events. *Human factors*, *56*(8), 1380-1400.

Lane, H. C., Core, M. G., Van Lent, M., Solomon, S., & Gomboc, D. (2005). *Explainable artificial intelligence for training and tutoring*. University of Southern California Institute for Creative Technologies. Marina Del Ray, CA.

MacEachren, A. M. (1995). *How maps work: representation, visualization, and design*. Guilford Press.

MacEachren, A., Xiping, D., Hardisty, F., Guo, D., & Lengerich, G. (2003, October). Exploring high-D spaces with multiform matrices and small multiples. In *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No. 03TH8714)* (pp. 31-38). IEEE.

Madhavan, P., & Wiegmann, D. A. (2007). Effects of information source, pedigree, and reliability on operator interaction with decision support systems. *Human Factors*, *49*(5), 773-785.

Marks, J. (1991). A formal specification scheme for network diagrams that facilitates automated design. *Journal of Visual Languages & Computing*, *2*(4), 395-414.

Meghdadi, A. H., & Irani, P. (2013). Interactive exploration of surveillance video through action shot summarization and trajectory visualization. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2119-2128.

Moon, B., Hoffman, R. R., Eskridge, T. C., & Coffey, J. W. (2011). Skills in applied concept mapping. *Applied concept mapping: Capturing, analyzing, and organizing knowledge*, 23-46.

Moon, B., Hoffman, R. R., Novak, J., & Canas, A. (Eds.). (2011). *Applied concept mapping: Capturing, analyzing, and organizing knowledge*. CRC Press.

Mueller, S. T., Hoffman, R. R., Clancey, W., Emrey, A., & Klein, G. (2019). Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI. *arXiv preprint arXiv:1902.01876*.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological review*, *92*(3), 289.

Nguyen-Tuong, A., Melski, D., Davidson, J. W., Co, M., Hawkins, W., Hiser, J. D., ... & Rizzi, E. (2018). Xandra: An autonomous cyber battle system for the Cyber Grand Challenge. *IEEE Security & Privacy*, *16*(2), 42-51.

Nobre, C., Meyer, M., Streit, M., & Lex, A. (2019, June). The state of the art in visualizing multivariate networks. In *Computer Graphics Forum* (Vol. 38, No. 3, pp. 807-832).

Overton, J. A. (2013). “Explain” in scientific discourse. *Synthese*, *190*(8), 1383-1405.

Pagendarm, H. G., & Post, F. H. (1995). *Comparative visualization: approaches and examples*. Delft: Delft University of Technology, Faculty of Technical Mathematics and Informatics.

Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic Books.

Pellerin, C. (2017). Project Maven to deploy computer algorithms to war zone by year’s end. *United States Department of Defense*, *21*.

Pirolli, P., & Card, S. (2005, May). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (Vol. 5, pp. 2-4).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, *1*(8), 9.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., ... & Jennings, N. R. (2019). Machine behaviour. *Nature*, *568*(7753), 477-486.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics*, *2*, 1926-1943.

Scheepens, R., Willems, N., van de Wetering, H., & van Wijk, J. J. (2011, March). Interactive visualization of multivariate trajectory data with density maps. In *2011 IEEE Pacific Visualization Symposium* (pp. 147-154). IEEE.

Schurr, N., Fouse, A., Freeman, J., & Serfaty, D. (2019, February). Crossing the Uncanny Valley of Human-System Teaming. In *International Conference on Intelligent Human Systems Integration* (pp. 712-718). Springer, Cham.

Shneiderman, B. (1996, September). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages* (pp. 336-343). IEEE.

Sottilare, R. A., Goldberg, B. S., Brawner, K. W., & Holden, H. K. (2012, December). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference* (pp. 1-13).

Spiro, R. J. (1988). Multiple analogies for complex concepts: Antidotes for analogy-induced misconception in advanced knowledge acquisition. *Center for the Study of Reading Technical Report; no. 439*.

Tominski, C., Schumann, H., Andrienko, G., & Andrienko, N. (2012). Stacking-based visualization of trajectory attribute data. *IEEE Transactions on visualization and Computer Graphics*, *18*(12), 2565-2574.

Walker, M. (2015). Machine vs. machine: lessons from the first year of cyber grand challenge.

Wang, F. Y., Zhang, J. J., Zheng, X., Wang, X., Yuan, Y., Dai, X., ... & Yang, L. (2016). Where does AlphaGo go: From church-turing thesis to AlphaGo thesis and beyond. *IEEE/CAA Journal of Automatica Sinica*, *3*(2), 113-120.

Wegman, E. J. (1990). Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, *85*(411), 664-675.

Woods, D. D. (2017). Essential characteristics of resilience. In *Resilience engineering* (pp. 21-34). CRC Press.

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford university press.

VanLehn, K. (2016). Regulative loops, step loops and task loops. *International Journal of Artificial Intelligence in Education*, *26*(1), 107-112.

Zobel, C. W. (2010, May). Comparative visualization of predicted disaster resilience. In *Proceedings of the 7th International ISCRAM Conference* (pp. 1-5). ISCRAM.

Chapter 18 – Visualization Of Sequential Interactions in Adaptive Instructional Systems

Vasile Rus[1], Zachari Swiecki[2], Jody L. Cockroft[1], and Xiangen Hu[1, 3]
University of Memphis[1], University of Wisconsin - Madison[2], Central China Normal University[3]

Introduction

This chapter explores several ways to visualize tutorial interaction between tutors (human or computer-based), and tutees. Such interactions can be viewed as mixed sequences of actions by the tutors and tutees. The actions can be raw, such as a tutor giving a hint (e.g., by displaying a message on screen or saying it) or the learner selecting an answer choice by clicking a button on the interface or the learner simply typing or speaking the answer. The actions can also be more abstract/latent, and derived from raw input. That is, the actions can be derived as derived variables from raw data or latent actions can be inferred through a more advanced process intended at generalizing the kind of observed sequential patterns across diverse but similarly evolving action sequences (Guo et al., 2018). Examples of derived actions are the speech acts such as asking a *Question* or making an *Assertion* or dialogue modes such as *Scaffolding* (Cade, Copeland, Person & D'Mello, 2008; Rus, Maharjan, & Banjade, 2015; Rus, Niraula, Maharjan, Banjade, 2015). Speech acts are derived from raw utterances that tutors and tutees articulate in tutorial dialogues. Dialogue modes correspond to sequences of speech acts that may correspond to a pedagogical strategy.

Specifically, this chapter focuses on dialogue interactions and two kinds of visualizations of tutorial dialogue interactions: (1) sequence logo analysis and (2) heatmaps illustrating differences between epistemic frame stationary distributions (to be explained shortly). The former provides in an intuitive, graphical form a summary of sequences of pedagogical strategies used by tutors thus allowing researchers to infer, for instance, the most likely sequence of strategies in successful tutorial sessions. A successful session could be defined as either a session in which sound pedagogical strategies have been applied and/or student learning gains have been observed.

The second type of visualizations, the heatmap visualizations, can help provide a better understanding of the role students play in multi-party conversations in virtual internship environments which are online education technologies that offer students professional internship experiences based on the epistemic frame theory. The epistemic frame theory claims that professionals develop epistemic frames, or a network of skills, knowledge, identity, values, and epistemology (SKIVE elements) that are unique to that profession (Chesler, Bagley, Breckenfeld, West, & Shaffer, 2010). For example, engineers share ways of understanding and doing (knowledge and skills), beliefs about which problems are worth investigating (values), characteristics that define them as members of the profession (identity), and ways of justifying decisions (epistemology). Virtual internships based on the epistemic frame theory need to be able to detect and track interns' SKIVE profile at each moment in order to trigger appropriate instructional strategies that will help the interns acquire the necessary SKIVE profile that would enable them to become successful professionals. The proposed visualizations can help researchers and developers monitor interns' SKIVE elements during their virtual internships offering an important feedback tool for virtual internship design which in turn should result in improved virtual internships for the learners.

Sequence Logo Analysis

A key question in learning sciences is discovering effective pedagogical strategies also called tutorial strategies when the focus is on one-on-one instruction. A data-driven approach to discovering such strategies is to record tutor-tutee interactions and then automatically discover patterns of tutor and tutee actions. Such an automatic, data-driven approach is particularly useful when a large volume of data is available which cannot possibly be analyzed by humans due to sheer size (Rus, Maharjan, & Banjade, 2015).

As already noted, our focus is on tutorial dialogues. To that end, we analyzed one-on-one chatroom like interactions between human tutors and students. Those interactions were collected from an online commercial tutoring service where students can get access to a human tutor for a limited amount of time for a fee. A large corpus of about 19,000 tutorial sessions between professional human tutors and actual college-level, adult students was collected via an online human tutoring service. Students taking two college-level developmental mathematics courses (pre-Algebra and Algebra) were offered these online human tutoring services at no cost. The same students had access to computer-based tutoring sessions through Adaptive Math Practice, a variant of Carnegie Learning' Cognitive Tutor. A subset of 500 tutorial sessions containing 31,299 utterances was randomly selected from this large corpus for annotation with the requirement that a quarter of these 500 sessions would be from students who enrolled in one of the Algebra courses (Math 208), another quarter from the other course (Math 209), and half of the sessions would involve students who attended both courses.

To map dialogue interactions to sequences of actions, the dialogue-based interactions between tutors and tutees are segmented into sequences of dialogue acts based on the language-as-action theory (Austin, 1962). Contiguous segments of dialogue acts in turn can be associated with general conversational segments (e.g., openings/closings) and task-related and pedagogical goals (e.g., scaffolding). Such patterns are called dialogue modes (Cade, Copeland, Person & D'Mello, 2008; Rus, Niraula, Maharjan, Banjade, 2015) and, when they can be linked to pedagogical goals, are regarded as tutorial strategies (Maharjan, Rus, & Gautam, & Rus, 2018). Once tutorial dialogues are mapped onto sequences of dialogue modes, they are automatically analyzed, and a summary visualization generated in order to allow a visual inspection and interpretation by experts. A detailed description of the dialogue modes is available in work by Morrison and colleagues (Morrison et al., 2014).

We use sequence logos as an efficient visualization tool for representing distribution of various observations over discrete time. For instance, they are used in bio-medical research for visually representing sequences of genes. In our work, we used sequence logos to investigate the profile of dialogue modes in temporal space. The sequence logo regards each dialogue session as a discrete sequence of dialogue modes and then determines the dominant mode at each discrete moment in the sequence. The dialogue mode at the top of a stack of modes at each discrete moment of the dialogue is the most frequent mode at that moment. Furthermore, the height of each letter in a stack represents the amount of information contained. The bigger the letter/mode at a particular discrete time, the more certain the dominance of the corresponding mode. For instance, at the discrete time 1 in the sequence logo shown in Figure 1 the dominant mode is Opening.

From the sequence logo, we can infer the most certain sequence of dialogue modes in a typical human tutoring session as the sequence of the most certain dialogue modes at each discrete moment. The dominant sequence of modes/logos up to average mode switch length is O, P, N, P, S, S, S, D, S, S, K, S, S, R, S, S, S, S, F, S and S[2] for the top 10% of sessions in terms of learning gains as illustrated in Figure 1 (see Maharjan, Rus, & Gautam, 2018). The average mode switch length is the average number of mode switches across all analyzed tutoring sessions.

[2] Assessment(A), Closing(C), Fading(F), ITSupport(I), Metacognition(M), MethodID(E), Modeling(D), Opening(O), ProblemID(P), ProcessNegotiation(N) RapportBuilding(B), RoadMap(R), Scaffolding(S), SenseMaking(K), SessionSummary(Y), Telling(T)

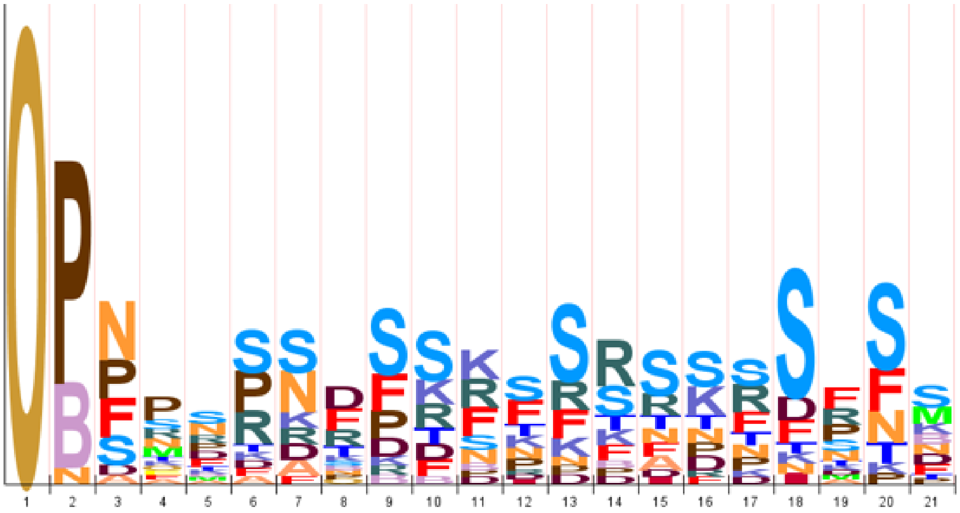


Figure 1. Dialogue mode sequence logo for top 10% of tutorial sessions in terms of largest learning gains. The average length of such sessions in terms of mode switches is 21. (Note for the figure: AAAI Copyright - reprinted with permission from AAAI.)

Heatmaps for Stationary Distribution Analysis of SKIVE Elements

When participating in internships, besides content knowledge, students need to master their target profession's skills, knowledge, identity, values, and epistemology (SKIVE or epistemic frame elements). A virtual internship environment offering a virtual equivalent of or something very close to an actual, real-life internship needs to monitor the acquisition of SKIVE element mastery.

We have explored a novel way to monitor, assess, and visualize students' mastery of the SKIVE elements in terms of stationary distributions of the states of a SKIVE Markov process in which there is a state for each of the SKIVE elements (Rus, Gautam, Bowman, Graesser, & Shaeffer, 2017). We rely on students' activation of SKIVE elements during their conversations with other players in the virtual internship to characterize the underlying Markov process and infer the stationary SKIVE distribution for each intern.

We use a Markov process approach to infer the stationary distribution of SKIVE elements based on an analysis of learners' conversations with other players, e.g. a mentor or intern, in virtual internships. Markov processes are characterized by a set of states, which in our case are the SKIVE elements, and a set of transition probabilities, which we derive from analyzing the activation of SKIVE elements during the virtual internship conversations.

We focused on engineering virtual internships, such as Nephrotex (NTX) and RescueShell (RS), in which students research and create multiple engineering designs (Bagley & Shaffer, 2009). As part of their engineering design process, the students regularly engage in online conversations with teammates and mentors. We infer their current SKIVE mastery based on these conversations.

Elements of the engineering epistemic frame are operationalized as discourse codes in order to detect when students activate SKIVE elements during conversations. While an empirical distribution could be derived by computing the relative proportion of each activated SKIVE element, our goal is to infer the true or stationary distribution of SKIVE elements for each student by modeling students' epistemic frames as Markov processes. For this purpose, every utterance of a conversation recorded in virtual internships is being annotated with binary codes indicating whether a particular SKIVE element is present or absent in the utterance. That is, whether the student activated the corresponding SKIVE elements during his conversational moves.

Once we inferred the SKIVE epistemic frame in terms of a stationary distribution for each student/intern, we compared students' SKIVE epistemic frames against each other and against an average epistemic frame distribution obtained by computing an average of the stationary distributions of students' epistemic frames. We compared the epistemic frame distributions using Kullback-Leibler (KL) divergence. Heatmaps visually summarizing the KL divergences of student SKIVE element distributions were then generated (see Figure 2). It should be noted that we used data from the virtual internship Nephrotex in which groups of students work together on a design problem, e.g. designing filtration membranes for hemodialysis machines, with the help of a mentor. In the Nephrotex dataset, there are 25 players divided into five groups. Each group is assigned a virtual room to work together on a task. The dataset consists of a total of 2,970 utterances with an average of 37 utterances per room. The utterances are coded with 20 SKIVE elements. While analyzing the dataset, we found that some of the utterances do not contain any SKIVE elements, hence a row attributed to that utterance has zero counts across all SKIVE elements. We handled such scenarios with two different approaches. In the first approach, we discarded all the utterances with all-zero counts. In another approach, we introduced a dummy state, called no-SKIVE state, which indicates a state when no SKIVE element was activated by a student in an utterance.

In the heatmaps, student players are sorted based on their average KL score with other players. The left vertical color bar in the maps show the intensity of the user's average KL score in sorted order. Some players have similar distributions, e.g., those shown in the lower left corner of the heatmaps in Figure 2 (from Rus, Gautam, Bowman, Graesser, & Shaeffer, 2017).

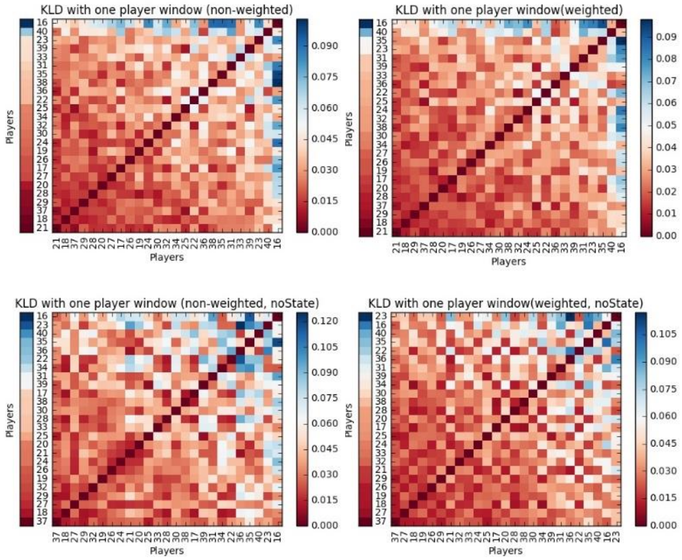


Figure 2. Examples of heatmaps showing KL-divergence of distribution SKIVE elements. (Note for the figure: AAAI Copyright - reprinted with permission from AAAI.)

The lower left corner corresponds to lower divergence scores. When using a no-SKIVE state and weights (less weight is given to transitions derived from utterances far apart), the distributions of SKIVE elements between players seem to be more similar as shown in the heatmaps on the lower right-hand side of Figure 2. Furthermore, adding a no-SKIVE state (bottom left and bottom right) revealed that some of the players move from an upper position, corresponding to a higher average divergence score, to lower-divergence positions in the heatmap. Those that move to lower-divergence positions are more likely to have utterances in which no SKIVE elements are activated.

Our work provides a rigorous way of describing students' mastery of the SKIVE elements in terms of stationary distributions as opposed to empirically derived distributions. Furthermore, the heatmap visualizations allow researchers to interpret differences among students in terms of SKIVE elements mastery. For instance, in our work (Rus, Gautam, Bowman, Graesser, & Shaeffer, 2017), through the proposed heatmap visualization, we discovered that some students play a more managerial/coordinator role as their utterances focus more on process and conversational management topics as opposed to the epistemic frames elements.

Recommendations and Future Research

Visualizing information, e.g., information extracted from learning activities or at least digital records of such activities, could offer an intuitive and elegant view of learning activities and of learners' state (whether cognitive or epistemic state or of another nature such as affect). Furthermore, when paired with additional data analysis steps before generating the visualizations, the resulting visualizations can be an extremely powerful way to synthesize and understand learning activities. In particular, we exemplified the use of Sequence Logos to visualize distributions of sequences of actions and strategies over many one-on-one tutorial sessions and also the use of heatmaps to compare stationary distributions of learners' epistemic elements in virtual internships. The Sequence Logos are an efficient visualization tool for representing the distribution of various observations over discrete time. Heatmaps are very good at illustrating intensities of a phenomenon such as similarities and dissimilarities between two sets of items which in our case were distances among stationary distributions or the relationship among two variables such as SKIVE stationary distributions of two students. The proposed visualizations can be applied to other similar learner and learning data as long as they are representing sequential data in the case of Sequence Logos and intensities of a phenomenon or two variables in the case of heatmaps.

Such visualizations could be extremely useful for administrators or parents or other stakeholders who are less familiar with numerical reports of data analyses. For this reason, the Generalized Intelligent Framework for Tutoring (GIFT) should include a variety of visualization components to present intuitive visualizations of various learning aspects that GIFT mediates to various stakeholders. In order for visualizations of the kind we proposed here to be offered as GIFT features, there is a need to integrate a number of components into GIFT such as dialogue act classification, automated detection of SKIVE elements from dialogues, and stationary distribution inference components. Furthermore, there is a need to incorporate sequence logo generation libraries (there are some already available) and heat map generation libraries (there are some already available).

Conclusions

This chapter described two visualization techniques for sequential information in learning environment dialogues. The dialogues we analyzed were exchanges among human learners and human tutors or mentors. Furthermore, in one case, the tutorial dialogues were between one learner and one tutor (that is, between two people, not necessarily the same, i.e., the learner and tutor were most likely different in different sessions) whereas in the other case the dialogues were among a group of learners and a human mentor. The bottom line is that in both cases the interactions can be viewed as mixed sequences of actions by the tutors/mentors and tutees/learners. In order to visualize those sequences of actions, we described a method based on Sequential Logos and one based on visualizing differences between stationary distributions using heatmaps. The former provides in an intuitive graphical form a summary of sequences of pedagogical strategies used by tutors. The heatmap visualizations can help with better understanding the role students play in multi-party conversations in virtual internship environments. The proposed techniques could eventually be used for any sequential interaction data in education environments, not only dialogue interactions.

ACKNOWLEDGMENTS

This research was partially sponsored by the National Science Foundation under award The Learner Data Institute (award #1934745) and an award from the Department of Defense (U.S. Army Combat Capabilities Development Command - Soldier Center). The opinions, findings, and results are solely the authors' and do not reflect those of the funding agencies.

References

Austin, J. L. (1962). How to do things with words. Oxford University Press.

Bagley, E., & Shaffer, D. W. (2009). When people get in the way: Promoting civic thinking through epistemic gameplay. *International Journal of Gaming and Computer-mediated Simulations*, 1, 36-52.

Cade, W., Copeland, J., Person, N., & D'Mello, S. (2008). Dialogue modes in expert tutoring. In *Intelligent tutoring systems*, 470–479. Springer.

Chesler, N. C., Bagley, E., Breckenfeld, E., West, D., & Shaffer, D. W. (2010, June). A virtual hemodialyzer design project for first-year engineers: An epistemic game approach. *In ASME 2010 Summer Bioengineering Conference* (pp. 585-586). American Society of Mechanical Engineers.

Guo, S., Jin, Z., Gotz, D., Du, F., Zha, H., & Cao, N. (2018). Visual Progression Analysis of Event Sequence Data. *IEEE Transactions on Visualization and Computing Graphics*, 2018, Aug 20.

Maharjan, N., Rus, V., & Gautam, D. (2018). Discovering Effective Tutorial Strategies in Human Tutorial Sessions. In Proceedings of *The Thirty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS-31)*, Melbourne, Florida, USA. May 21-23 2018.

Morrison, D., Nye, B., Samei, B., Datla, V., Kelly, C., & Rus, V. (2014). Building an intelligent pal from the tutor. Com session database phase 1: Data mining. In Proceedings of The 7th International Conference on Education Data Mining, July 4-7, 2014, London, UK

Rus, V., Maharjan, N., & Banjade, R. (2015). Unsupervised Discovery of Tutorial Dialogue Modes in Human-to-Human Tutorial Data, *The 3rd Generalized Intelligent Tutoring Framework Symposium*, June 17-18, Orlando, FL.

Rus, V., Niraula, N.B., Maharjan, N., Banjade, R. (2015). Automated Labelling of Dialogue Modes In Tutorial Dialogues. In Proceedings of *the 28th International Conference of the Florida Artificial Intelligence Research Society Conference* (FLAIRS), Hollywood, Florida. May 18–20, 2015.

Rus, V., Gautam, D., Bowman, D., Graesser, A., & Shaeffer, D. (2017). Markov Analysis of Students' Professional Skills in Virtual Internships. In Proceedings of *the 30th International Conference of the Florida Artificial Intelligence Research Society Conference*, Marco Island, Florida, May 22–24, 2017.

CHAPTER 19 – DATA VISUALIZATION WITHIN SIMULATED GAMES

Keith Brawner[1] and Scott J. Ososky[2]
U.S. Army Combat Capabilities Development Command (DEVCOM) – Soldier Center – Simulation and Training Technology Center[1], Microsoft Corporation[2]

Introduction

The world, in general, and intelligent tutoring systems, in specific, are able to create more data than can be sufficiently analyzed. To quote Herbert Simon – "In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it." (Simon, 1996, pp.40 - 41). The world is quickly becoming "information-rich" as Simon would put it, but an information-rich world is not necessarily an effective one.

Modern intelligent tutoring systems, serious games for learning, and video games share many things in common – they have players taking actions, navigating through environments, encountering obstacles, overcoming (or quitting), and further activities within simulations. As an example, the Artificial Intelligence for Interactive Digital Entertainment (AIIDE) conference regularly has a track on learning games alongside its entertainment games track. Both environments face similar problems, in that the systems they generate can have many differing measurements, but the question of which measurements and visualizations are most effective for achieving outcomes is not clear. A simple strategy of measuring everything is not sufficient towards an end goal.

Further, modern military simulations require the generation of synthetic data in order to train computational models, such as artificially intelligent agents or automatic after action review summaries, which can create vast volumes of data quickly ("Synthetic Training Environment", n.d.). This problem is similar to the beta launching of a new game product, where suddenly many players simultaneously crowdsource testing and interactions. The experiences of both synthetic agents in an existing environment and new players in a new environment can be difficult to predict, which is part of the reason we rely on data collection in order to assess the experience of the use. While you can run many simulations or test many variations of user contact, it is not clear what to *measure* in order to improve the overall system.

Environments of Interest

Modern video games are data rich systems, and a worthwhile domain to examine when considering best practices for data visualization for training applications. There are at least three aspects of video games that produce exceedingly useful data visualizations: player experiences, viewer experiences, and post-play analysis; think of the first two as real-time visualization. Even the most seemingly simplistic games require players to quickly process information and react to unfolding situations. As a simple example, consider the player-controlled avatar's health in a game, which could be represented by a red-bar, a set of hearts, a percentage, and so on. Information like that is usually found in the heads up display (HUD). Alternatively, that information could also be conveyed with increasing the appearance of damage on a vehicle, the change in the stance of a player character, or increasing the intensity of a red hue on a screen. Numbers sometimes accompany those cues (especially in a Role Playing Game (RPG) or strategy game), but are rarely ever the first or only cue that a player will attend to. What is so interesting about representing data like player health is that it is arguably tacit knowledge for anyone that has played a game. Some cues are so intuitive, like the use of hearts in Legend of Zelda series games, that it needs little explanation even to non-game players. Full, red hearts are *good*, empty hears are *bad*. This simple yet effective visualization lets players know how much health they have, how much they have lost, the damage output of an enemy, and amount of health restored by helpful game items.

What these visualizations do is help the player to be able to see, at a glance, what is happening to them, around them, in the game world, and so on. When moving from a single-player game to a multi-player game to an electronic sports (eSports) game or a massively multi-player online role playing game (MMORPG), information density tends to increase and the time available to process that information decreases. For games that are more information-dense, think of all of the different questions that a player can answer simultaneously just by examining the icons, graphics, numbers, and colors in a game HUD: Where am I? How much health do I have? What am I carrying? What resources do I have? Which direction am I facing? Where am I going next? Where are my teammates? Who is talking to me? What is my mission? How much time is left? Where am I looking/aiming? Am I winning or losing a match?

Later in this chapter, we briefly touch upon large scale simulation. In keeping with game-inspired cross-disciplinarily work, we also consider a third case of special visualizations built for game *viewers*, specifically those watching eSports (Hamari & Sjoblom, 2017). Not unlike traditional sports, these broadcast events provide additional information to the viewer to help them understand what is going on (Kokkinakis et al., 2020).

Whether for an individual player or a viewer, the fact that game HUDs and other cues are so effective in communicating information is no accident. The end product is based on a considerable amount of development effort, supported by user research and data science. Those disciplines create, analyze and visualize data about games in order to improve many game aspects like player engagement, player satisfaction, monetization, game challenge, pace, and/or balance, just to name a few. As an example, Wired magazine ran an article over a decade ago about the use of player data to tune the balance of a multi-player map in Halo 3, a first-person shooter game (Thompson, 2018). The team examined where player deaths occurred and plotted those as a heat map over a top-down representation of the map itself. The conclusion was that one of the two competing teams had an advantage on the map based on a team's starting position on the map. The map was adjusted in order to balance the challenge of starting from either side of that particular game map.

At the core, the work is driven by a question that needs an answer. In the case of Thompson's (2018) work, the question was probably something like, "is this map balanced, to offer a similar challenge to both teams?". Data about the locations of player deaths supplied an answer, when consumed visually. Similarly, training simulations will have many questions that need to be answered, and those questions should drive the data sources to measure and visualizations to build. Going back to the multi-player heatmap example, the visualization could have been adjusted to include a dimension of time to understand how the action unfolds on a map as teams compete. Or, perhaps the player death icons could have been color coded to get a sense of the weapons with which players were being defeated. Or, the heat map could have been reimagined to show a player's position when they defeated another player (instead of where the original player perished). The questions asked from the gaming perspective are frequently reframed or mirrored in military examples, where instead of asking "how can this map be made more balanced?" the mirrored question is "what starting position unbalances this map to favor my team?".

Ultimately, all of these viable options provide little value unless there is an underlying need for them. Games provide an endless source of inspiration, in both examples to be examined and in research publications on the topic to inspect. Plenty of information is available, including archival talks from the Game Developers Conference as well as the bi-annual Game User Research summit.

The Generalized Intelligent Framework for Tutoring (GIFT) has been compatible with the Army's game for training Virtual Battlespace 3 (VBS3) since its first official version (Sottilare, Brawner, Sinatra, & Johnston, 2017; Simulations, 2015), which was chosen for programmatic reasons. At the time of writing, the GIFT development team is in the middle of switching the game engine to the game engine being developed by the Synthetic Training Environment (STE) Cross Functional Team (CFT), collectively referred to as the STE-Common Synthetic Environment (STE-CSE).

The STE-CSE has an intelligent agent/enemy requirement. A derived part of the requirement means that it will be possible to enter in battle plans and test them against the types of agents that have been trained in the simulation. Taking this a step further, it would be possible to have computer-generated battle-plans against the computer-generated forces (Loper & Turnitsa, 2012). These mock battles can be used in order to generate insights for live battles. This combination of requirements and items means that the number of training events and simulated battles requiring visualization is going to significantly increase.

The Purpose of Metrics is to Answer Questions about Gameplay

By default, each of the environments above comes packaged with instrumentation to answer questions about itself. In all of the environments above, however, the questions that can be answered by the default data are not necessarily the questions that are most relevant. These questions are items which are appropriate to the game, but not necessarily important to the user, such as:

- What happened?
  - Instrumentation on player/enemy positions, by timestep
  - Replay capability
- How did the player do?
  - Player performance metrics (kills, deaths, attempts, territory)
  - Intra-player metrics

In terms of video games and military simulators, however, examples of more interesting questions to answer are questions such as:

- Is the game fun, or being enjoyed?
- What is the player doing wrong?
- How can the player improve?
- What is an appropriate amount of difficulty for this particular player?
- Is the players' response to a new situation an appropriate one?
- Are there options that the player has neglected to notice, or use?

A User-Centered Approach to Data Visualization Design

The questions referenced in the previous section additionally require succinct manners of presentation back to the controlling users. The simplicity of the initial question set allows for a simplicity of visualization – showing a replay with some statistics. The more advanced and interesting questions, however, do not have a simple visualization scheme – how do you show 1,000 replays? What is the best way to show a player that they are neglecting options chosen by others?

When planning for data visualizations based on activities and behaviors within synthetic environments or games for training, a user centered approach is recommended. Such an approach should be taken up at the earliest possible moment in the planning and development of the training or learning environment. The *users*, with respect to data visualizations, are not necessarily the players, trainees, or learners; rather the users are the *consumers of the data* that is generated by the interaction of the learner(s) with the game or simulation. Consumers of data visualizations may include the learners, but will also include instructors, administrators, researchers, and anyone else having an interest in the outcomes of those interactions.

Data visualization consumers should be involved early and often in the development of a simulation/game, training program, etc. to understand the types of questions that they are trying to answer by consuming data. For instance, if you were building a marksmanship simulator an instructor might want to know: What did the trainee do? What did their weapon do? What were the environmental conditions? What was the outcome of each shot? (Johnson, 2001) Notice that those questions are system agnostic. They are just as applicable to a simulated or game environment as they are to the activity or event that is being simulated. Those questions seem intuitive and reasonable at face value, but where do those questions actually come from?

Those questions are system agnostic because they are not driven by a game or simulation (i.e., What *could* we create?), but questions that come from the user's needs (i.e., What *should* we create?), or their jobs to be done (JTBD). The theory surrounding JTBD has been around for about 20 years and the concept is relevant to user-centered design. The basic premise is that customers buy products in order to achieve a certain outcome, a “job” that they are trying to do (Ulwick, 2017). If you have not seen this idea put into practice, bookmark this page, set your book (or computer) down, then go and search the internet for “Clay Christensen milkshakes video”. We will be here when you are done, no rush.

So, while this theory might have been borne out of thinking from business and marketing (i.e., to sell more milkshakes), there are connections that can be made to the current problem space. Think of a data visualization like a product, one that is competing against other “products”. A good question to ponder might be, why would a user “hire” my visualization over something else, such as a simple video recording, or interviews with the learners, or a “match” summary, or their own intuition? An approach to visualization design is recommended in which the goal is to help the data consumer do their job in a more efficient way than those competing products.

In the current context, the JTBD for a marksmanship instructor might be, “Evaluate performance of trainee” or “Identify areas of improvement for trainee” or “Provide feedback to trainee”. In a more abstract sense, that instructor might have other jobs like, “make it to my next engagement on time” or “get X number of participants through this training program by the end of the week”. In order to do those things, the marksmanship instructor (a data consumer), will need answers to questions like those identified earlier (What did the trainee do? What was their weapon doing?). With those jobs and questions in mind, it becomes much easier to identify what needs to be measured and why. Instead of trying to boil the ocean of data to measure everything, we are focusing on making a really great cup of coffee – boiling only what is needed.

Involve users early and often to understand their needs (Abras, Maloney-Krichmar, & Preece, 2004). It is important enough to repeat it in this chapter. Games and simulations have the capability to generate completely unmanageable sets of data including, but not limited to imagery, audio, communication logs, user behaviors, use of objects by users, object characteristics and behaviors, interactions between objects, behaviors of non-player characters (NPCs), player state changes, world state changes, and the list goes on (Medler & Magerko, 2011). A data consumer does not want to see all of it, nor would they have the time to make sense of all of it. Returning to the marksmanship example, what specific aspects of a simulation or game might be useful to track or measure, given the instructor's jobs to be done, and associated questions? It may be important to attend to: the way the user was breathing, the angle at which they were holding their weapon, the rotation of the weapon, how much pressure was on the trigger (and when), how many shots were fired, and the outcomes of those shots. Alternatively, a training effectiveness researcher may have questions requiring more aggregate views of data across multiple trials or groups of trainees, adding a layer of data about the data. Perhaps a better question is, what can you safely not measure? That is - what information is not useful for a particular audience?

The key takeaway is, that by knowing what to track, record, and measure ahead of time, those can become requirements that go into a development plan for a game or simulation. Equally important, you will also know what *not* to track. There may be situations in which telemetry "hooks" (Drachen, 2015) may need to be built into a game or simulation, which, like anything else, may become more costly to implement as the project gets further into development.

To that end, it may be desirable to mock up potential data visualizations and gather feedback from potential users to determine the extent to which they are useful and desirable. Doing this early in the development process allows data visualization designers to make changes and then re-evaluate, as necessary. Better to let the limit of a data visualization be one's effort and imagination rather than have it be constrained by time, systems, or resources.

New Approaches are Much Better

New approaches are better – and better targeted – to ask the types of questions of relevance to the outcomes. They include the following:

- In-depth metrics of player performance on multiple fronts. For example, not a "win/lose", but a "did they use the weapon feature" and a "how accurate is their shooting?"
- Multi-phasic replay
- Heatmaps
- Over/under performance metrics

Where large scale simulation is concerned, there are likely needs that go far beyond what video games have to offer. In large scale simulations, there will likely be data about different types of entities, events, and outcomes. There will be data about what the artificial intelligence (AI) did, and what the interactions between the entities looked like. All of this will likely result in a dashboard containing multiple visualizations of the data arranged in a useful hierarchy (tested and retested with end-users, of course). With all of that in mind, there is one place in which games could offer guidance, and that is in the perspective of the *viewer*. The viewer is not an active participant in a game, but is important to the game industry with respect to online streaming and particularly eSports (Charleer et al., 2018).

As an example, think about what a match of a 100-player Battle Royale game like Fortnite or Player Unknown's BattleGrounds (PUBG) looks like to a player including what they see in the HUD and what they see at the conclusion of a match. Ultimately that one player's perspective provides a very narrow window into what happened in that match. Now, think about the viewer's perspective, this is someone who tuned in (likely online) to watch a broadcast of a match of one of those games unfolding in real time. Not only do they require more information, they are also able to see things that would otherwise be considered "cheating" if those things were available to the game players. For instance, the viewer can spectate any player they wish, or switch to a birds-eye view ("Pubg - Game 1", 2017). All nearby players are outlined in yellow, even behind cover. All players have cards above them showing their weapon and health. A lot of the visualizations are there to help the user understand what is going on, like "who is left in the match?", "what player is that?", "What is their health / weapon", "What team are they on?", "Where are the other players in relation to them”. All of this information is shown to the viewer, in addition to what the player sees.

This is not unlike the type of visualizations that one might see in a broadcast of traditional sports like baseball, where information is provided to help the user to understand the speed of the pitch, and where it went in relation to the strike zone. In football, lines are superimposed on the field to indicate things like the first down marker, or a kicker’s field goal range. In both American football and simulated football (i.e. Madden NFL), there are also stats and Gamebooks compiled for just about every measurable aspect of the game (e.g., playtime percentage) (The National Football League, 2015). Again, it is little more than data unless there is a question that needs to be answered, but it is available nonetheless.

For large-scale simulation, one might assume that the data that is generated will be vast, and the visualizations numerous. Perhaps all of it is important enough to include in a data explorer or dashboard solution. With that in mind, consider at the very least the notion of progressive disclosure (Nielsen, 2016). In Human Computer Interaction (HCI), this refers to reducing the amount of information shown at once, placing the most important information up front, and providing additional elements on the user’s request (e.g., Settings and Advanced Settings). Applied to large-scale simulation data visualization, consider what is most important to show right up front, and give users the power to explore as needed to consume additional information.

Design Recommendations

At the risk of being trite – it is important to consider the end user up front and early in the design process. Consider what insights are to be generated in order to meet the ends that you are seeking. Questions that should be considered about the end user, for every visualization, are:

1. Who are the consumers (users) of data visualizations, what are their roles?
2. What are their jobs to be done in their roles?
3. What questions are asked to complete those jobs?
4. What media and/or data is needed to answer those questions?
5. How will that media and/or data be tracked, recorded, or measured within the game/simulation?
6. Draft sample visualizations, test with users, evaluate, adjust, and re-test.
7. Remember, why would the user want to hire your visualization over something else?

To provide specific answers to these questions in regard to GIFT, a place to begin would be with the following answers:

1. Training managers, managing training for a small group for which they are responsible.
2. Assigning training and managing training events.
3. In planning - “What does my group need to train on?”. In preparing – “What items should they experience?”. In execution – “How are they handling the scheduled events?”. In assessing – “How did they do?”.
4. In planning – a view of the various strengths and weaknesses of the group and as individuals. In preparing – A mapping between available content resources and the selected strengths/weakness of the team. In execution – a live view of performance towards all objectives. In assessment – A grading, replay, and mapping of key successes/weaknesses.
5. Answered above.
6. Left as an exercise.
7. Keep in mind that this is competing against pen, paper, and live observation. All visualizations should show something more than looking at a report card and watching from a single camera view.

References

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications, *37(4)*, 445-456.

Charleer, S., Gerling, K., Gutiérrez, F., Cauwenbergh, H., Luycx, B., & Verbert, K. (2018, October). Real-time dashboards to support esports spectating. In Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (pp. 59-71).

Drachen, A. (2015). Behavioral telemetry in games user research. In Game User Experience Evaluation (pp. 135-165). Springer, Cham.

Hamari, J. & Sjöblom, M. (2017), "What is eSports and why do people watch it?", Internet Research, Vol. 27 No. 2, pp. 211-232. https://doi.org/10.1108/IntR-04-2016-0085

Johnson, R. F. (2001). Statistical measures of marksmanship (No. USARIEM-TN--1/2). ARMY RESEARCH INST OF ENVIRONMENTAL MEDICINE NATICK MA MILITARY PERFORMANCEDIV.

Kokkinakis, A. V., Demediuk, S., Nölle, I., Olarewaju, O., Patra, S., Robertson, J., ... & Hughes, P. (2020) DAX: Data-Driven Audience Experiences in Esports.

Loper, M. L. & Turnitsa, C. (2012). History of combat modeling and distributed simulation. *Engineering Principles Of Combat Modeling And Distributed Simulation*, 331-355.

Medler, B., & Magerko, B. (2011). Analytics of play: Using information visualization and gameplay practices for visualizing video game data. *Parsons Journal for Information Mapping*, *3*(1), 1-12.

Nielsen, J. (2006). Progressive disclosure. Retrieved from https://www.nngroup.com/articles/progressive-disclosure/

Pubg - Game 1 - Final - Iem Oakland Pubg Invitational. (2017). Retrieved from https://youtu.be/ju0ktwX6z0Y

Simon, H. A. (1996). Designing organizations for an information-rich world. *International Library of Critical Writings in Economics, 70*, 187-202.

Simulations, B. I. (2015). VBS3: The future battlespace. Retrieved February 19, 2015.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT). *GIFTtutoring. org*.

Synthetic Training Environment (STE). (n.d.). Retrieved from https://asc.army.mil/web/portfolio-item/synthetic-training-environment-ste/

The National Football League. (2015). Super Bowl XLIX National Football League Game Summary. Retrieved from http://www.nfl.com/liveupdate/gamecenter/56502/SEA_Gamebook.pdf

Thompson, C. (2018). Halo 3: How Microsoft labs invented a new science of play. Retrieved from https://www.wired.com/2007/08/ff-halo-2/

Ulwick, A. W. (2017). Outcome-Driven Innovation®(ODI): Jobs-to-be-Done Theory in Practice. Strategyn, LLC Whitepaper.

Chapter 20 – Data Analytics and Visualization for xAPI Learning Data: Considerations for a GIFT Strategy

Shelly Blake-Plock[1], Will Hoyt[1], Cliff Casey[1], and Diego Zapata-Rivera[2]
Yet Analytics, Inc.[1], Educational Testing Service (ETS)[2]

Background

When implementing Experience API (xAPI; Advanced Distributed Learning Initiative, 2013), context matters. xAPI is a data specification that enables an interoperable capability to capture and store activity data produced during learning and training. The xAPI Profiles specification (Advanced Distributed Learning Initiative, 2017) provides a structured approach to the documentation of the vocabulary concepts, extensions, statement templates, and patterns that support the development of controlled and domain-specific xAPI implementations. Because the design of an xAPI Profile influences the quality and value of the data available for reporting and data visualization downstream, it is important that the design of xAPI data, xAPI Profiles, and subsequent data visualizations are taken into consideration holistically.

The Data Analytics and Visualization Environment — DAVE — (Yet Analytics, 2018), which is currently a beta project at the Advanced Distributed Learning (ADL) Initiative (Blake-Plock, 2020), is comprised of a framework of specifications, a language for learning analytics algorithms, and a software reference model (Yet Analytics, 2020). The DAVE framework is built for the unique demands of the learning record store (LRS) and is designed to support analytics and visualization across ADL's Total Learning Architecture (TLA), including as it relates to Intelligent Tutoring Systems (ITSs) such as the Generalized Intelligent Framework for Tutoring (GIFT).

In the context of GIFT (U.S. Army Research Lab, 2012), the alignment of xAPI data that tracks activity through the intelligent tutoring paradigm with capabilities that provide for the automation of visualizations, alerts, and reports is beneficial. These capabilities provide the framework that is necessary to derive insight both into learner behavior as well as into the decisions being made by the intelligent tutor within the event-based context of the learning activity. Previous work regarding the application of xAPI to GIFT includes a description of the capabilities of xAPI in regards to supporting the development of interoperable ITSs in GIFT (Johnson, Nye. Zapata-Rivera, & Hu, 2017).

In general, the xAPI Profile specification provides a means to standardize and make use of both xAPI vocabularies as well as patterns and templates of activity common to a learning domain. The DAVE specification is designed to consume xAPI data that is aligned to xAPI Profiles. Therefore, an efficient way to establish a valuable and semantically interoperable data visualization and reporting system for GIFT may be to establish a sound protocol for the design and implementation of ontologically-accurate xAPI Profiles in the intelligent tutoring domain — this could mean profiling the technology and domain of the ITS itself or could be applied to independent activities occurring within an ITS environment such as math instruction versus language instruction. This protocol would be built to align with the objectives produced through GIFT research, and the end result should be data schematics designed for GIFT that fully leverage xAPI.

There are several defining features in semantic data that are germane to a discussion of event-based visualizations. At the root of this is the concept of the Resource Description Framework — RDF — (W3Schools) which at a high-level can be understood as a clustering — called a triple — of an actor, a verb, and an object used as a way to describe things on the Web. xAPI is based on the RDF concept. Ontologies (W3C Vocabularies), at a high-level, leverage Linked Data (W3C) to create context around those triples. xAPI Profiles support JSON-LD (W3C) to promote the use of this contextualizing capability.

Structured data and technologies leveraging Linked Data provide a consistency of meaning that is machine readable. This data can support advanced analytics by affording the navigation and resolution of related data across the Web. Ontologies are the basis of this data. The design of an ontology will have a consequential effect on the accessibility of meaning, and therefore the meaningfulness of the visualization of event-based data in any given context is governed by it. It is important from the perspective of data visualization to take into consideration the quality of the design of a given ontology. Of special interest with regard to GIFT is the xAPI Ontology (Advanced Distributed Learning Initiative, 2018), which represents the xAPI Statement Data Model as RDF Classes and Properties.

The design and implementation of the xAPI Ontology and related ontologies across the linked data spectrum will have an impact upon the options for data visualization of GIFT xAPI data downstream. In designing a data regime for GIFT that will be able to produce relevant data visualizations in the future, a goal must be to support the design of ontologies and data models that sustain the automation of data visualizations at scale. These ontologies and models should also provide for a degree of contextualization that keeps the data, resulting visualizations, and reports relevant to the mission. Event-based data that potentially could stream through an ITS include the activities that actually transpire from amongst the activities that are possible. These activities are nodes on a complex web of possibilities that comprise the parameters of a given learning experience.

In the context of intelligent tutoring, a proper ontology could provide a key resource for automated data visualizations as it would limit the disconnect between initial data production and the interpretation of that data downstream. Therefore, fed by data structured in accordance to the ontology, the downstream data visualizations could both portray meaning and provide insight into characteristics and patterns within a given event or series of events contextualized by the ontology. Individual xAPI Profiles are used to model distinct experiences or capabilities, therefore in the case of xAPI data, this contextualization would relate to the modeling purposes of the specific xAPI Profile at hand. The downstream data visualization could potentially offer insight into what is underlying the tutor's decisions vis-a-vis the rendering of relevant characteristics and patterns as described by the xAPI Profile and then as rendered as xAPI data tracked from the actual performance of the ITS.

At a high level, the flow of data is:

- xAPI Ontology and xAPI Profile governs the model
- Instrumented content delivered from xAPI Learning Record Provider to LRS
- Validation of xAPI and storage of and access to xAPI data via the LRS
- xAPI data consumed by DAVE, queried, and visualized
- Visualizations exported to dashboards, business intelligence tools, and reporting tools

A challenge in event-based data design is the creation of ontologies and Linked Data resources that accurately reflect the context of semantic information represented by context-rich learning and training events as well as experiences which are not categorized essentially as learning and training experiences, but have learning outcomes (e.g., informal and in-the-flow learning that occurs on the job in response to novel requests). The quality of ontology and data design has a direct impact upon an analytics consumers' ability to visualize, report on, and gain insight from event-based data. With regards to xAPI, this will have an impact upon the ability of a Learning Record Consumer — LRC — (Advanced Distributed Learning Initiative, 2017) to effectively provide discovery of or access to insight to meaning within any xAPI data set. Software applications that are built on the properties of the DAVE framework would generally be classified as LRCs.

Data visualization, in its current business usage, is designed for humans, not machines. However the processes that produce digital data visualizations, especially automated ones (such as persistent visualizations and visual alerts), are machine, not human processes. This is important because event-based data visualization is generally a problem of scale and data logistics more than a problem of graphic representation. In short, if the former is not addressed, the later does not matter. ITSs will rely on scalable infrastructure and will create scale issues with regard to an exponential rise in the amount of learning and training data made available to them. With regard to scale, consider that 100,000 learners engaging in typical Shareable Content Object Reference Model (SCORM) formatted courses over the course of a month might produce less event-based data than a single learner on a single day engaged in a hybrid simulation that tracks biometrics against digital activity. Considering all of the components and processes that lead to downstream visualization, this is perhaps the matter that should be of greatest interest to the implementer of such a system. It is important to note that even when planning for machine scale, it is necessary to optimize for human decision making. For this reason, both machine and human oriented processes should be aligned in order to provide humans with the data needed to support their decisions (Zapata-Rivera, Graesser, Kay, Hu, & Ososky, 2020).

Does the DAVE Framework support the GIFT/xAPI Alignment?

Johnson et al. (2017) recommended that alignment of GIFT and xAPI occurs in the following five areas:

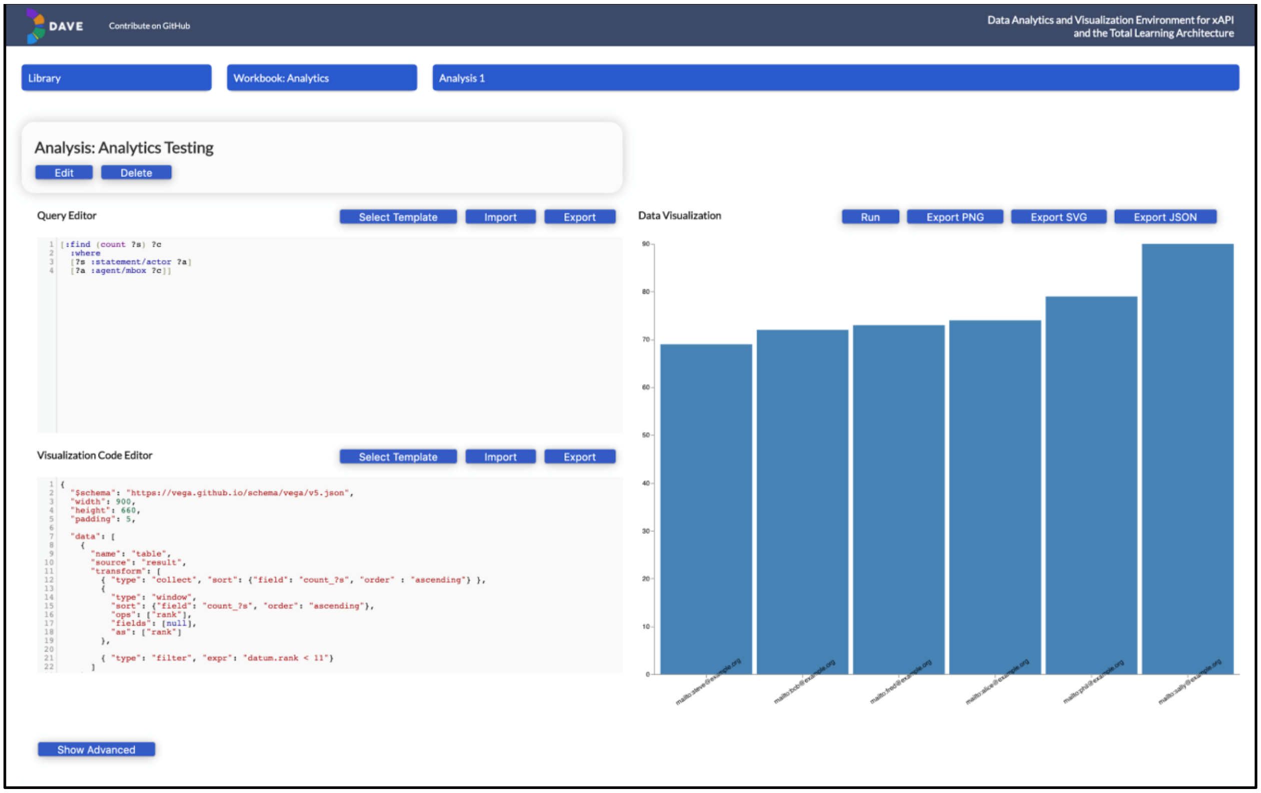
- fine-grained achievement data that allows GIFT to track performance data across different contexts
- metrics to compare the effectiveness of different interventions
- developing an assessment vocabulary in GIFT
- supporting the development of an effective model for competencies and their decay
- an assessment profile to track both formal and informal learner experiences

This book chapter considers the ways in which the DAVE framework either a) provides, b) could be iterated to provide, or c) would not directly provide analytics and data visualization capabilities supporting the recommended alignment.

Results

The DAVE framework makes extensive use of the availability of rich data offered by ontologies and xAPI Profiles. DAVE is being designed to provide a common means through which Department of Defense (DoD) and Federal Stakeholders could analyze, interpret, and visualize micro-level behavior-driven learning aligned to the technical requirements of xAPI, xAPI Profiles, and the TLA. Featuring a query editor and a visualization code editor, DAVE provides a user (an analyst, instructional designer, learning ecosystem systems engineer, or learning enterprise executive) with the ability to customize xAPI data analysis, prototype visualizations, and export data from the subsequent visualization code to persistent reports. While the beta version of the DAVE software application is lightweight and can be deployed in any location, including browser-only with no server side implementation, the technologies and protocols chosen for the beta can be scaled to handle very large streaming datasets, and can be modified to distribute workloads horizontally across a distributed computing environment. In an effort to build analytics specifically for the unique case of xAPI, the DAVE framework provides scaffolding for the design of algorithms meant to leverage the attributes found in xAPI data statements (Yet Analytics, 2019). In reviewing capabilities as aligned to Johnson’s five areas, mentioned above (Johnson et al., 2017), the DAVE beta reference model software application was used within a browser on commodity hardware — in this case a MacBook Pro — in order to demonstrate a baseline use case.

In an attempt to simulate the type of variety that may be generated by a proper xAPI Profile built specifically for GIFT, xAPI datasets for the demonstration of DAVE were generated using the open source Apache 2.0 Data and Training Analytics Simulated Input Modeler (DATASIM; Yet Analytics, 2020) and specifically the Tactical Combat Casualty Care (TC3)/Care Under Fire/Hemorrhage Control xAPI Profile. Note that DATASIM is an Apache 2.0 open source application designed to generate realistic xAPI Statement datasets from a conformant xAPI Profile, a set of Actors, and alignments between the Actors and components of the xAPI Profile. See Figure 1 for a screenshot of the beta version of DAVE displaying a query and the resulting visualization of data aligned to a subset of the Tactical Combat Casualty Care / Care Under Fire / Hemorrhage Control xAPI Profile.



```
[:find (count ?s) ?c
  :where
  [?s :statement/actor ?a]
  [?a :agent/mbox ?c]]
```

```
{
  "$schema": "https://vega.github.io/schema/vega/v5.json",
  "width": 900,
  "height": 660,
  "padding": 5,

  "data": [
    {
      "name": "table",
      "source": "result",
      "transform": [
        { "type": "collect", "sort": {"field": "count_?s", "order" : "ascending"} },
        {
          "type": "window",
          "sort": {"field": "count_?s", "order": "ascending"},
          "ops": ["rank"],
          "fields": [null],
          "as": ["rank"]
        },

        { "type": "filter", "expr": "datum.rank < 11"}
      ]
```

Figure 1. The Data Analytics and Visualization Environment (DAVE) beta version. This screenshot shows queries of learner engagement over time using a subset of the Tactical Combat Casualty Care (TCCC | TC3) Care Under Fire (CUF) Hemorrhage Control (HC) xAPI Profile.

The analytics and data visualization capabilities of the DAVE framework and software reference model provide the user with the ability to achieve some of the GIFT-to-xAPI alignment goals recommended by Johnson et al. (2017), namely:

- Through xAPI result fields, there is the ability to gather fine-grained achievement data to track performance across contexts, including across platform modalities. For example, in the tested use case it was possible to track performance against each possible activity, and the durations of those activities, that could occur within a learning environment for the TC3 Hemorrhage Control curriculum containing both video engagement content and digital assessment.

- There is the mathematical ability to compare metrics across fields including within the context of interventions. The DAVE framework provides a full range of mathematical functions:

  abs(); acos(); acosh(); asin(); asinh(); atan(); atan2(); atanh(); cbrt(); ceil(); clz32(); cos(); cosh(); exp(); expm1(); floor(); fround(); hypot(); imul(); log(); log10(); log1p(); log2(); max(); min(); pow(); random(); round(); sign(); sin(); sinh(); sqrt(); tan(); tanh(); trunc()

  which can be implemented in queries across all attributes of the xAPI data set, including:

  :language-map/language-tag; :language-map/text; :extension/iri; :extension/json; :account/name; :account/homePage; :agent/objectType; :agent/name; :agent/mbox; :agent/mbox_sha1sum; :agent/openid; :agent/account; :group/objectType; :group/name; :group/mbox; :group/mbox_sha1sum; :group/account; :group/member; :group/account; :verb/id; :verb/display; :interaction-component/id; :interaction-component/description; :definition/name; :definition/description; :definition/correctResponsesPattern; :definition/interactionType; :definition/type; :definition/moreInfo; :definition/choices; :definition/scale; :definition/source; :definition/target; :definition/steps; :definition/extensions; :activity/objectType; :activity/id; :activity/definition; :statement-reference/objectType; :statement-reference/id; :score/scaled; :score/raw; :score/min; :score/max; :result/score; :result/success; :result/completion; :result/response; :result/duration; :result/extensions; :context-activities/parent; :context-activities/grouping; :context-activities/category; :context-activities/other; :context/registration; :context/instructor; :context/team; :context/contextActivities; :context/revision; :context/platform; :context/language; :context/statement; :context/extensions; :attachment/usageType; :attachment/display; :attachment/description; :attachment/contentType; :attachment/length; :attachment/sha2; :attachment/fileUrl; :sub-statement/actor; :sub-statement/verb; :sub-statement/object; :sub-statement/result; :sub-statement/context; :sub-statement/timestamp; :sub-statement/attachments; :sub-statement/objectType; :statement/id; :statement/actor; :statement/verb; :statement/object; :statement/result; :statement/context; :statement/timestamp; :statement/stored; :statement/authority; :statement/version; :statement/attachments

The DAVE framework could be iterated to provide:

- An assessment vocabulary aligned to xAPI Profiles with the ability to identify which parts of a GIFT xAPI vocabulary are most relevant to real-world usage. For instance, DAVE can provide engagement information correlating assessment vocabulary with the digital event stream of an on-the-job workflow. Key here would be either to coerce the on-the-job data stream into xAPI (for the most efficient use of the DAVE framework) or to run the workflow and the xAPI data as separate streams in a common streaming architecture and then run DAVE in parallel with other business intelligence processes.

Where it is difficult to see where the DAVE framework directly corresponds to the alignment goals is with regard to:

- Where competency or assessment data are made available through reference xAPI statements; the ability to query for the purpose of comparing strategies, interventions, and decay rates across both formal and informal learner experiences.
- An assessment profile to track both formal and informal learner experiences.

There is potential to leverage the DAVE specification in parallel to a competency framework for the purpose of meeting the penultimate goal. Though the more efficacious way to accomplish this may be an alignment between an xAPI Profile and a specific competency framework. DAVE would then consume the xAPI Profile. With regard to the assessment profile, DAVE may provide analytics capabilities related to the outcomes of the data, but the necessary work itself would more likely be a matter of xAPI Profile design. The needed xAPI Profile would include statement templates and patterns matched to the assessment objectives and the range of possible activities, learner pathways, and outcomes. An xAPI Profile can provide a competency-determining system with the information it may need, but in and of itself, an xAPI Profile — and by extension, DAVE — would not perform that computation in any way.

Note that team activity is an accessible attribute of xAPI data, and therefore can be visualized by the DAVE framework or any system built-for and properly implementing xAPI. The more nuanced view is whether specific data visualizations are built to address this available data asset (just like any asset or attribute of xAPI). Team-oriented analytics for divisions, departments, teams, access hierarchies, and a variety of personae (including learner teams, instructor and executive teams) are accessible via xAPI and may be considered in the design of xAPI Profile statement templates.

It is useful to note that in a broader sense, when designing analytics for xAPI, in addition to the semantic issues, of high importance are:

- ID management
- Identity and Privacy
- Authorization and Data Management
- Ability to manage mobility within the workforce

A standardized process of ID management with an eye towards consistency of identity and maintenance of privacy will help to alleviate many of the traditional single-view challenges, such as needing to join actors who use different email addresses as IDs across different platforms. In addition to assuring good business practices, designing data models with an eye towards the way that data will be accessed and managed down the line will improve the ability to provide something that will have relevance to the metrics measured by those constituencies. In environments where rates of attrition or mobility within the workforce is high, it is valuable to consider the portability of data and the means of assuring validity and conformance of the data in different future lines of operation. This discussion is relevant to data visualization as a chart is only as meaningful as the data that is put into it. Generally speaking in regard to automated visualizations, a lesson learned is that a visualization may mask, but cannot hide poor data design. Consideration given to the discussed “operational” functions of data will provide better results with regards to the goal of better insight into learning. Operational functions should not be treated as a second tier to learning and assessment functions.

Discussion

Based on the current work, the lesson most pertinent to GIFT is that downstream visualizations will be the result of upstream data design decisions. The most important consideration is not what visualization to use, but rather to ensure that the data is designed in a way that is consistent and can leverage all of the desired functionality, such as Linked Data, semantic resources, and operational workflows. Data visualization is only as good as the underlying data in both the shape and content of that data.

Beyond the matter of ontologies and data design, in order to increase interoperability, the data visualization assets serving GIFT should follow an exportable open source specification such as the Vega visualization grammar (Vega, 2018). The value add is in the ability to export and share the underlying code of data visualization templates no matter what underlying database technology is used. In the case of DAVE, a common visualization grammar provides the ability to easily share and remix data visualization code built on pre-existing xAPI queries.

Recommendations and Future Research

To achieve the GIFT-to-xAPI alignment goals discussed above, the most interoperable solution would be to design and develop an xAPI Profile Suite for GIFT. If the business need for GIFT is to visualize metrics such as leadership and initiative, supporting behavior, communication, and information exchange, as well as the performance of and progress of individuals and teams towards learning objectives, the underlying event-based data should leverage xAPI Profiles. As xAPI Profiles can model both macro and micro level activity, technologies such as GIFT, which are generally agnostic in terms of the domain of the content that passes through the platform, can benefit from a portfolio approach to xAPI Profiles, something akin to an xAPI Profile Suite. This suite consists of three levels of modeling:

- A high-level Administrative xAPI Profile which models the generalized capabilities of GIFT including core functionality and expected behavior in any use case.
- A secondary-level Pedagogical xAPI Profile which models the strategic instructional capabilities of GIFT so that an author of content to be used with GIFT could assign activities to be perceived in the context of certain expected patterns and statement templates.
- A tertiary-level Domain xAPI Profile which models the specific learning activity for whatever domain and related content is being applied. It is expected that there could eventually be as many third-tier applications as there are career-fields served by GIFT.

Note that the xAPI Profiles for GIFT would not be limited to a verb registry. Rather, the suite would model the entirety of currently understood GIFT capabilities and provide an easy way to append and add Profile-based content and contextualization as GIFT use cases mature. This approach can provide the capability of modularly modeling any possible use case for GIFT. It is applicable to distributed learning, synthetic training environments, and simulation-based training.

Additionally, note that because ontologies can affect data visualization, any GIFT resources depending on Linked Data functionality would be affected by any ambiguity in the ontologies serving GIFT. Therefore, it would be highly recommended to audit those ontologies for clarity, purposefulness, and validation against any xAPI Profiles that GIFT employs. For example, the namespaces of an ontology must be configured so as to allow the distinction between IDs. Lacking to do so would inhibit GIFT's ability to properly access information across Linked Data necessary to perform functions including any automated data visualization that depends on external linked sources of data. Additionally, class hierarchies will affect the ability to establish a context using JavaScript Object Notation for Linked Data (JSON-LD), particularly in using xAPI Profiles to support the establishment of patterns which can inform machine-led automation, intervention, alerts, and the like. The use of improper subclass axioms and components of statements inhibits the proper query behavior and will limit visualization capabilities in the space of event-based data. For these reasons, it is recommended that an audit be done of any ontologies and Linked Data resources to be used by GIFT and validation be done on any relevant xAPI Profiles. Note that an element of this audit is to make sure that the ontology and Linked Data resources will be available to generate the visualizations that different users will need.

To support cross-platform and multi-domain applications, there should be clear and useful links between the xAPI Ontology and other ontologies such as the xAPI Profiles Ontology so as to convey meaning. Related both to scalability and readability, the ontologies should be usable by modern high-performance triplestores. A triplestore is a specialized database designed to provide for the storage and retrieval of semantic triples (as mentioned above). Ill-constructed ontologies will adversely affect the ability of GIFT, DAVE, and other systems to leverage necessary data assets at scale and will limit the capabilities, usefulness, and confidence of data visualizations produced. Further, ontologies and Linked Data resources should be established for the purpose of supporting human decision making. Additionally, because GIFT will touch data across a wide array of learning use cases, it is recommended that a suite of xAPI Profiles be designed to provide the ability to fully model any given use case. Understanding that not every possible use case is — or necessarily can be — known, the approach to the GIFT xAPI Profile Suite should be to leverage modularity and extensibility.

Note that data visualizations themselves are brittle and that without a good metric outcomes strategy to support the data that infuse them, they will either break or produce bad data. Therefore, the GIFT xAPI Profile Suite — especially the Administrative Profile — should be designed to take into consideration the technical workflow of GIFT itself. The GIFT workflow will influence the ways that data are made accessible and will provide context in every instance of how GIFT was applied in the learning scenario.

Conclusions

Effective data visualization is dependent upon effective data design. Data visualization can be graphically sophisticated, sound, and aesthetically pleasing and yet present inaccurate insights. Data design does not begin in the authoring of activity. Data design begins with the ontologies that support the structure and capabilities of the data model.

In regard to the recommendations of Johnson et al. (2017) to align GIFT and xAPI, the DAVE framework provides:

- fine-grained achievement data that allows GIFT to track performance data across different contexts
- metrics to compare the effectiveness of different interventions

The DAVE framework could be iterated to provide:

- an assessment vocabulary in GIFT

The DAVE framework would not directly include, but could potentially be leveraged to support the design and development of:

- an effective model for competencies and their decay
- an assessment profile to track both formal and informal learner experiences

The ontologies and data models underlying event-based data affect the shape and quality of the data, and therefore influence the quality and usefulness of data visualization. In the case of automated data visualization based on queries traversing JSON-LD, confidence in the data visualization can only be as sound as the quality of the ontology or ontologies. GIFT, as an application central to the learning enterprise, should require a data strategy aligned to the event-based data capabilities of xAPI and the domain-modeling capabilities of xAPI Profiles. Aligning GIFT with xAPI is of foremost importance with regard to producing meaningful data, measurable metrics, accessible data visualizations, and reports.

References

Advanced Distributed Learning Initiative. (2013). xAPI-Spec. Retrieved version 1.0.3 on April 11, 2020, from https://github.com/adlnet/xAPI-Spec.

Advanced Distributed Learning Initiative. (2017). xAPI-Spec: LRC. Retrieved April 11, 2020, from https://github.com/adlnet/xAPI-Spec/blob/master/xAPI-About.md#def-learning-record-consumer.

Advanced Distributed Learning Initiative. (2018). xAPI-Ontology. Retrieved April 11, 2020, from https://github.com/adlnet/xapi-ontology.

Advanced Distributed Learning Initiative. (2017). xAPI-Profiles. Retrieved version 1.0 on April 11, 2020, from https://github.com/adlnet/xapi-profiles.

Blake-Plock, S. (2020). Technical Report: Data Analytics and Visualization Environment for xAPI and the Total Learning Architecture. Advanced Distributed Learning Initiative.

Johnson, A., Nye. D. B., Zapata-Rivera, D.,& Hu, X. (2017). Enabling Intelligent Tutoring System Tracking with the Experience Application Programming Interface (xAPI). In R. Sottilare, A. Graesser, X. Hu, and G. Goodwin (Eds.). *Design Recommendations for Intelligent Tutoring Systems: Volume 5 - Assessment Methods*. Orlando, FL: U.S. Army Research Laboratory. ISBN 978-0-9893923-9-6. 41–45.

U.S. Army Research Lab. (2012). Generalized Intelligent Framework for Tutoring. Retrieved version 2019-1 on April 11, 2020, from https://www.gifttutoring.org/projects/gift/wiki/Documentation_2019-1.

University of Washington Interactive Data Lab. (2018). Vega: a Visualization Grammar. Retrieved April 11, 2020, from https://github.com/vega/vega.

W3C. JSON-LD 1.1. Retrieved April 11, 2020, from https://www.w3.org/TR/json-ld11/.

W3C. Linked Data. (2020). Retrieved April 11, 2020, from https://www.w3.org/standards/semanticweb/data.

W3C. Vocabularies. Retrieved April 11, 2020, from https://www.w3.org/standards/semanticweb/ontology.html.

W3Schools. XML RDF. Retrieved April 11, 2020, from https://www.w3schools.com/xml/xml_rdf.asp.

Yet Analytics. (2020). DATASIM: Data and Training Analytics Simulated Input Modeler. Retrieved April 11, 2020, from https://github.com/yetanalytics/datasim.

Yet Analytics. (2018). DAVE: The Data Analytics and Visualization Environment for xAPI and the Total Learning Architecture. Retrieved April 11, 2020, from https://github.com/yetanalytics/dave.

Yet Analytics. (2019). Data Analytics and Visualization Environment for xAPI and the Total Learning Architecture: DAVE Learning Analytics Algorithms. Retrieved April 11, 2020, from https://github.com/yetanalytics/dave/blob/master/docs/main.pdf.

Yet Analytics. (2020). DAVE Beta Reference Model. Retrieved April 11, 2020, from https://yetanalytics.github.io/dave/.

Yet Analytics. (2020). Tactical Combat Casualty Care (TCCC | TC3) Care Under Fire (CUF) Hemorrhage Control (HC) Experience API (xAPI) Profile. Retrieved April 11, 2020, from https://github.com/yetanalytics/datasim/blob/master/dev-resources/profiles/tccc/cuf_hc_video_and_asm_student_survey_profile.jsonld.

Zapata-Rivera, D., Graesser, A., Kay, J., Hu, X. & Ososky, S. J. (2020). Visualization Implications for the Validity of ITS. In *Design Recommendations for Intelligent Tutoring Systems: Volume 8 – Data Visualization*.
Orlando, FL: U.S. Army Research Laboratory.

Chapter 21 – Utilizing Open-Source Data Analytics Software in Visualizing Performance in GIFT and Commercial Applications

Lina Brihoum, Zachary Heylmun, Mike Kalaf, Christopher Meyer, and Lucy Woodman
Synaptic Sparks, Inc.

Background

The Generalized Intelligent Framework for Tutoring (GIFT) software suite, with information available online at www.gifttutoring.org, was first deployed as downloadable software in May of 2012 (Sottilare, Brawner, Goldberg, & Holden, 2012). Since that time, the GIFT software suite has evolved to meet computer-based tutoring system (CBTS) mission goals as part of the Army Research Laboratory – Human Research and Engineering Directorate (ARL-HRED), and US Army Combat Capabilities Development Command (DEVCOM) – Soldier Center. As GIFT has evolved to meet additional research goals, specifically when GIFT began to provide an online server instantiation (GIFT Cloud) at cloud.gifttutoring.org, it became beneficial to track online user behavior. The process of tracking raw unique user behavior creates the materials used for Data Analytics. The science of Data Analytics, properly applied, indirectly supports research goals of adaptive tutoring by allowing scientists to more-closely observe, test, and verify user behavior in a software suite. This chapter shall now describe one such Data Analytics tool that is, at the time of this writing, used in conjunction with GIFT Cloud. It shall also present trade studies that were performed, inform the reader of the results, and finally conclude with recommendations for further satisfying future mission goals of US Army DEVCOM Soldier Center.

Piwik – An Open Source Data Analytics Tool

Piwik, now known as Matomo, is an open source analytics application (https://github.com/matomo-org/matomo) that was integrated with GIFT Cloud and deployed in January of 2017. Piwik version 3.0.2 was the final version deployed to track user behavior with GIFT Cloud and is also currently available online for GIFT system administrator and research scientist use. Many of the figures, diagrams, and data points mentioned in this chapter come directly from GIFT Cloud's aggregated collection of anonymized user behavior over the course of the last 3 years.

When considering Data Analytics, application developers have been tempted in the past to approach the problem through tracking everything possible that a user can do. Actions that a user does (or does not) perform, such as button clicks, idle time, mouse hovering, and hundreds of other actions have been trackable since the early 2000s. Approaching the problem of optimizing Data Analytics best practices, however, has proven that behavioral scientists do not need to drown in their own data. As Fiaz, Asha, Sumathi, and Navaz (2016) summarized, "Dealing with large volumes of data, often leads to chaos and misinterpretations between the data statistics, thus visually grouping together with many data points significantly enhances the quality of the data analytics and provides a much convenient [*sic*] way for the business executives, data scientists in understanding the relationships between the data" (p. 2802).

Data Analytics tools such as Piwik address the problem of an overabundance of user behavior data in much the same way as an intelligent tutoring system (ITS) such as GIFT addresses teaching. Drowning either a behavioral scientist or student in information is rarely the answer. Instead, engineers responsible for integrating Piwik with GIFT considered that an ITS is only successful if both the student and the human teacher understand the data presented to them. Because of the volume, scope, and variation of the data that both GIFT and Piwik records, both tools evolved to allow for smarter tracking, clustering, and visualization of user behavior.

Thanks to the internet and big data, the military is ingesting exponentially more data now than in prior decades. Further, this data is largely unstructured and difficult to interpret. Leaders do not have the luxury of failing to properly interpret data; as the correct information is paramount to the mission (Kulshrestha, 2015).

Piwik was chosen because of the application's flexibility in tracking different types of user actions. As the use of mobile devices increases, data sources increase in volume and diversity. Unfortunately, these sources do not produce data that is ready for scientific analysis (Kennedy, King, Lazer, & Vespignani, 2014). This presents the unique challenge of "harmonizing and extracting meaningful features from a variety of data streams" (Allum, Denman, Kim, & Metzler, 2016, p. 17). Tools such as Piwik allow for the aggregation of these various data streams into one coherent data set for analysis. Thus, Data Analytics and supporting applications are critical to consider for most future software development efforts.

Finally, it should be noted that there are two other types of Data Analytics used with GIFT Cloud other than Piwik. GIFT Cloud also has logs and data from the GIFT Monitoring tool that executes on the server instance which provides a higher degree of technical data concerning GIFT software modules that may be correlated with user actions. And, GIFT Cloud is hosted through Amazon Web Services (AWS) with specific configurations to enable a service called CloudWatch (aws.amazon.com/cloudwatch/) for system resource analytics. While beyond the scope of this chapter, the reader is encouraged to ask the GIFT team for more information on these topics if desired.

Methods

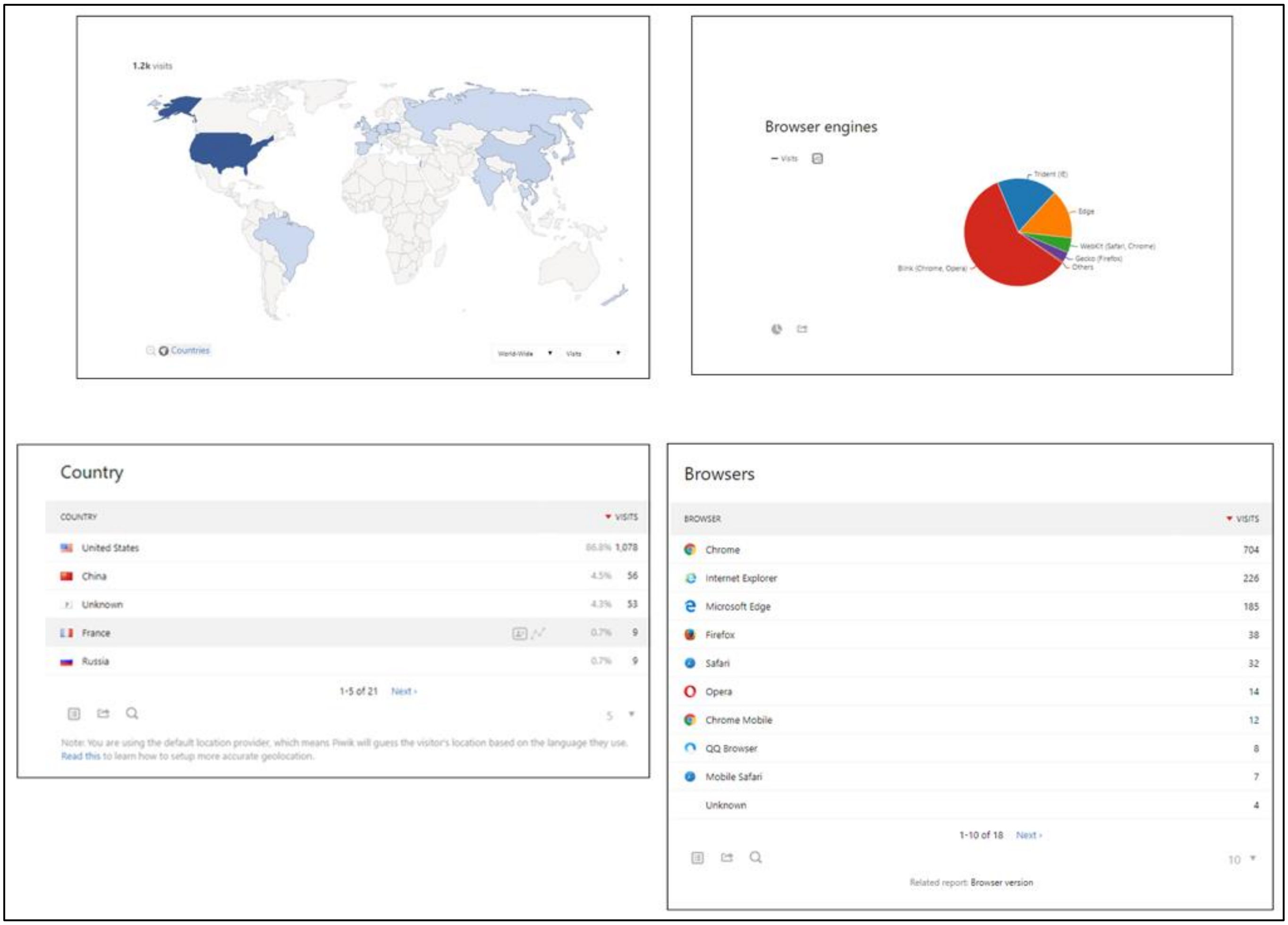
The GIFT team wished to research and observe user behavior while the community was interacting with GIFT Cloud. The integration of a Data Analytics tool such as Piwik allowed the GIFT team to examine user behavior and record concrete data from January 2017 and on, mainly at the web page level. The GIFT team also had objectives to relate application-level decisions to user behavior such as user interface layouts, and end user goals.

With these high-level goals in mind, the primary method used to begin gleaning useful information from Piwik was to aggregate the highest-level data first into a usable format. Piwik allowed a wide variety of information to be collected such as Country of Origin, Browser, Operating System, and Continent. Collecting this high-level information as a first step allowed behavioral scientists to be able to view information that had been assumed but never observed before about the GIFT user base. In addition, integrating this high-level information served as an introductory task to GIFT software engineers who were combining GIFT with Piwik for the first time.

As the GIFT team began interacting with Piwik, the most important item from a future validity standpoint was discovered. Version 3.0.2 of Piwik clusters certain batches of information into a data element called a "Visit." For certain data analytics categories such as Country or User Actions, Piwik v3.0.2 clustered similar IP addresses into a single Visit. This behavior led to some inaccuracy regarding the total number of participants in a country or interacting with GIFT. For example, during one GIFT interaction with participants from China, Piwik reported Visits from over 10 community members as a single Visit. This is due to the fact that the members were likely under the same lab IP address.

Events such as these revealed two things: First, data anomalies can and do occur. The data should be treated as guidelines, and the more critical a decision is, the more caution must be had when consulting the data. Secondly, as Piwik and future open source or Commercial-Off-the-Shelf (COTS) applications are integrated with any military product, said software may not necessarily operate with the level of confidence that is needed in mission critical environments without further development and investment.

Shown below in Figure 1 are the aggregated totals for roughly 70% uptime (the amount of time that the Piwik software was actively monitoring users while GIFT Cloud was also running and operational) during 2017, 2018, and 2019 Piwik GIFT Cloud operations. Renewal of expired web security certificates and a few instances of software update incompatibilities were responsible for the missing uptime. Future versions of Piwik and other Data Analytics tools allow for email notifications or other forms of communication to be automatically sent out upon radio silence, but it falls to GIFT software engineers to manually monitor data collection for this instantiation.



Country

COUNTRY		▾ VISITS
United States	86.8%	1,078
China	4.5%	56
Unknown	4.3%	53
France	0.7%	9
Russia	0.7%	9

1-5 of 21 Next ›

5 ▾

Note: You are using the default location provider, which means Piwik will guess the visitor's location based on the language they use. Read this to learn how to setup more accurate geolocation.

Browsers

BROWSER	▾ VISITS
Chrome	704
Internet Explorer	226
Microsoft Edge	185
Firefox	38
Safari	32
Opera	14
Chrome Mobile	12
QQ Browser	8
Mobile Safari	7
Unknown	4

1-10 of 18 Next ›

10 ▾

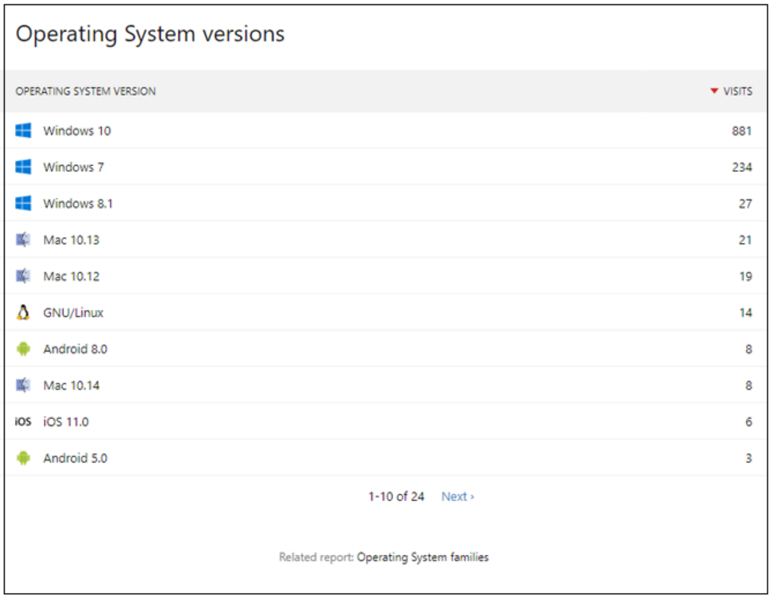
Related report: Browser version

Figure 1.Country Visitors and Browser of Choice for the GIFT Community

Again, while the exact numbers were interesting in the fact that they began to paint the picture of GIFT usage, of more initial importance may be the percent breakdown and variety of data. As the GIFT team assumed, the vast majority of usage comes from inside the United States, followed second by China, and third by masked-IP addresses. The percent breakdown came down to a 90% / 5% / 5% split, and this type of knowledge can start to add credence to claims of needing to cater to certain audiences, or expand usability through features such as internationalization (i18n) to capture more foreign users if those happen to be mission goals of GIFT at a high level. Finally, of amusing note in Figure 1 is Google Chrome's majority claim to the GIFT community's selection of browser. Technically, GIFT began development as an

Internet Explorer-only compatible program but due to natural community pressure and now the supporting Piwik numbers, compatibility with Google Chrome can only be said to be of undeniable value.

Figure 2 below shows the Operating System (OS) breakdown of GIFT Cloud users. Windows OS devices accounted for more than 90% of community usage, but even with a Windows-focused development for GIFT it was interesting to note the number of Mac, Linux, and Mobile devices visiting GIFT Cloud over time. This final piece of information concluded the GIFT team's initial data efforts with Piwik and justified further research and development efforts to collect finer-grained user behavior data.



Operating System versions

OPERATING SYSTEM VERSION	▼ VISITS
Windows 10	881
Windows 7	234
Windows 8.1	27
Mac 10.13	21
Mac 10.12	19
GNU/Linux	14
Android 8.0	8
Mac 10.14	8
iOS 11.0	6
Android 5.0	3

1-10 of 24 Next ›

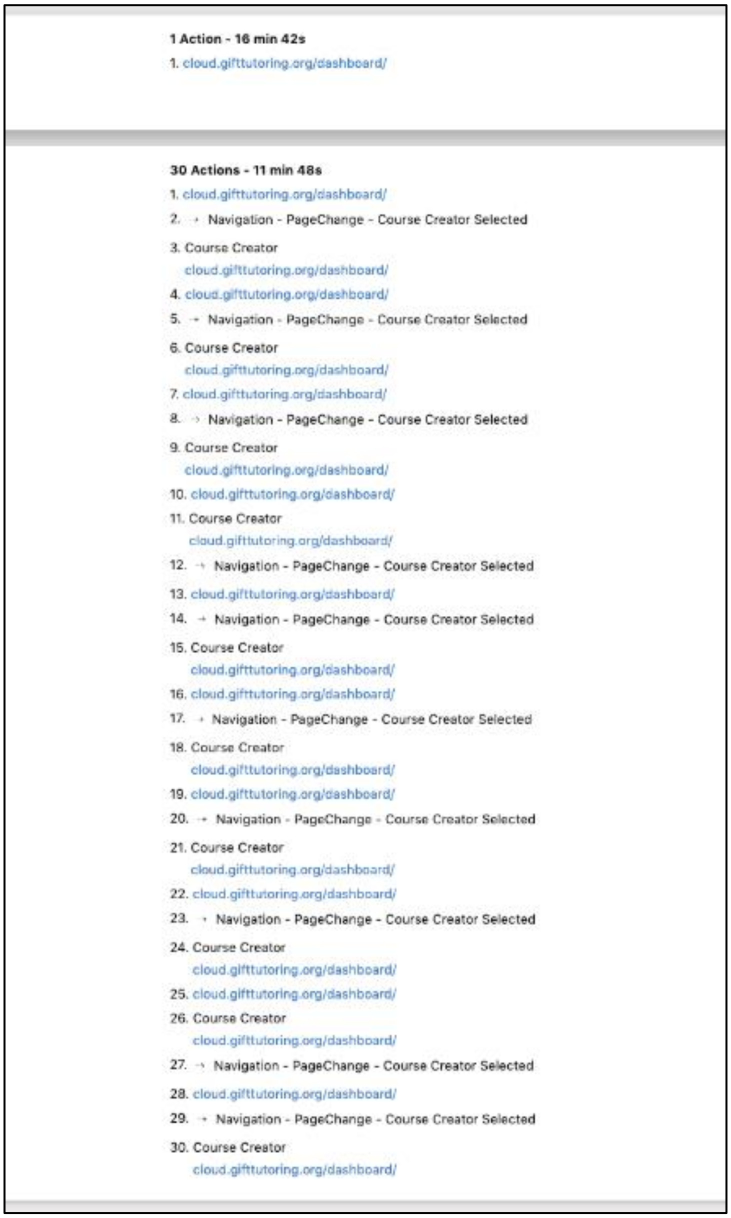
Related report: Operating System families

Figure 2. GIFT Cloud Operating Systems of Choice

Results

Continuing the above trend, behavioral scientists and engineers on the GIFT team were able to engineer the Piwik tracking system to sift through the potentially infinite set of data. The data set that resulted was focused first on tracking overall user page navigations in GIFT Cloud and focused secondly on user actions per page.

As shown in Figure 3 below, the Piwik system recorded and tracked an anonymized user through an 11 minute and 48 second session. As Data Analytics science is also an art, GIFT researchers may conclude that this user was a course author; due to the 15 different page navigations from the Course Creator to the Dashboard and vice-versa over the session. Furthermore, the user's session revealed that they were taking specific actions to save, load, and verify their course during that time period. After viewing this user behavior data alongside thousands of others, a scientist becomes able to gauge hypothetical use cases such as assuming users may be new or lost if there are too many navigations between pages without cohesive actions. Comparing the user behavior data in Figure 3 to other user experiences, the data revealed that there were two distinct user categories: the users that know exactly what they are doing, and the users that appear to be using the system for the first time. Very little user behavior data that demonstrated "journeyman" (mid-level) skill levels with GIFT Cloud were present. This suggests that new users force their way through the first few experiences and then do not perform the necessary training and study to become expert level users. This group may constitute students who are required by a professor to experiment with an ITS, a new user base trying GIFT Cloud for the first time, or other novice community stakeholders.



1 Action - 16 min 42s

1. cloud.gifttutoring.org/dashboard/

30 Actions - 11 min 48s

1. cloud.gifttutoring.org/dashboard/
2. → Navigation - PageChange - Course Creator Selected
3. Course Creator
   cloud.gifttutoring.org/dashboard/
4. cloud.gifttutoring.org/dashboard/
5. → Navigation - PageChange - Course Creator Selected
6. Course Creator
   cloud.gifttutoring.org/dashboard/
7. cloud.gifttutoring.org/dashboard/
8. → Navigation - PageChange - Course Creator Selected
9. Course Creator
   cloud.gifttutoring.org/dashboard/
10. cloud.gifttutoring.org/dashboard/
11. Course Creator
    cloud.gifttutoring.org/dashboard/
12. → Navigation - PageChange - Course Creator Selected
13. cloud.gifttutoring.org/dashboard/
14. → Navigation - PageChange - Course Creator Selected
15. Course Creator
    cloud.gifttutoring.org/dashboard/
16. cloud.gifttutoring.org/dashboard/
17. → Navigation - PageChange - Course Creator Selected
18. Course Creator
    cloud.gifttutoring.org/dashboard/
19. cloud.gifttutoring.org/dashboard/
20. → Navigation - PageChange - Course Creator Selected
21. Course Creator
    cloud.gifttutoring.org/dashboard/
22. cloud.gifttutoring.org/dashboard/
23. → Navigation - PageChange - Course Creator Selected
24. Course Creator
    cloud.gifttutoring.org/dashboard/
25. cloud.gifttutoring.org/dashboard/
26. Course Creator
    cloud.gifttutoring.org/dashboard/
27. → Navigation - PageChange - Course Creator Selected
28. cloud.gifttutoring.org/dashboard/
29. → Navigation - PageChange - Course Creator Selected
30. Course Creator
    cloud.gifttutoring.org/dashboard/

Figure 3. Anonymized User Visit Navigations

Figures 4 and 5 below present further tracking data about Visits made to GIFT Cloud. Looking at the "Visits per number of pages" data, GIFT researchers were able to conclude that the majority of users go to GIFT Cloud for a singular purpose, as indicated by > 90% of the visits limited to one to two pages. Furthermore, the distribution of time spent on a total visit was evenly spread between visits lasting anywhere from 2 minutes to more than 30 minutes.

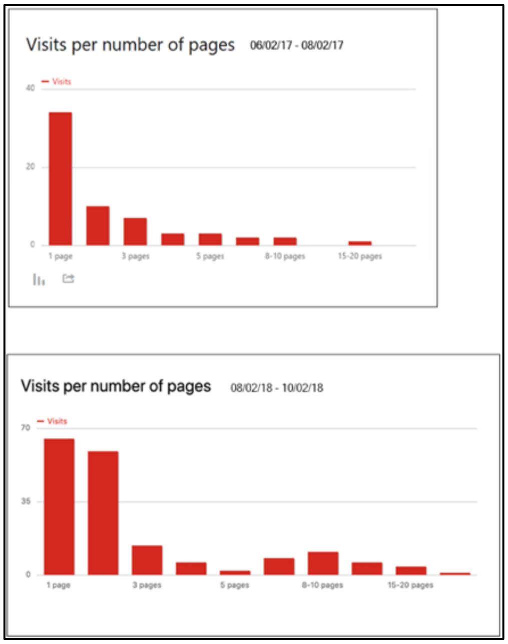


Figure 4. Visits Per Number of Pages

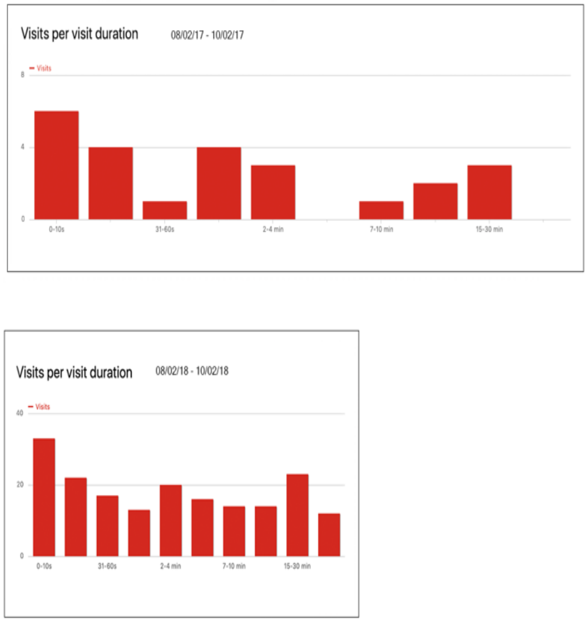
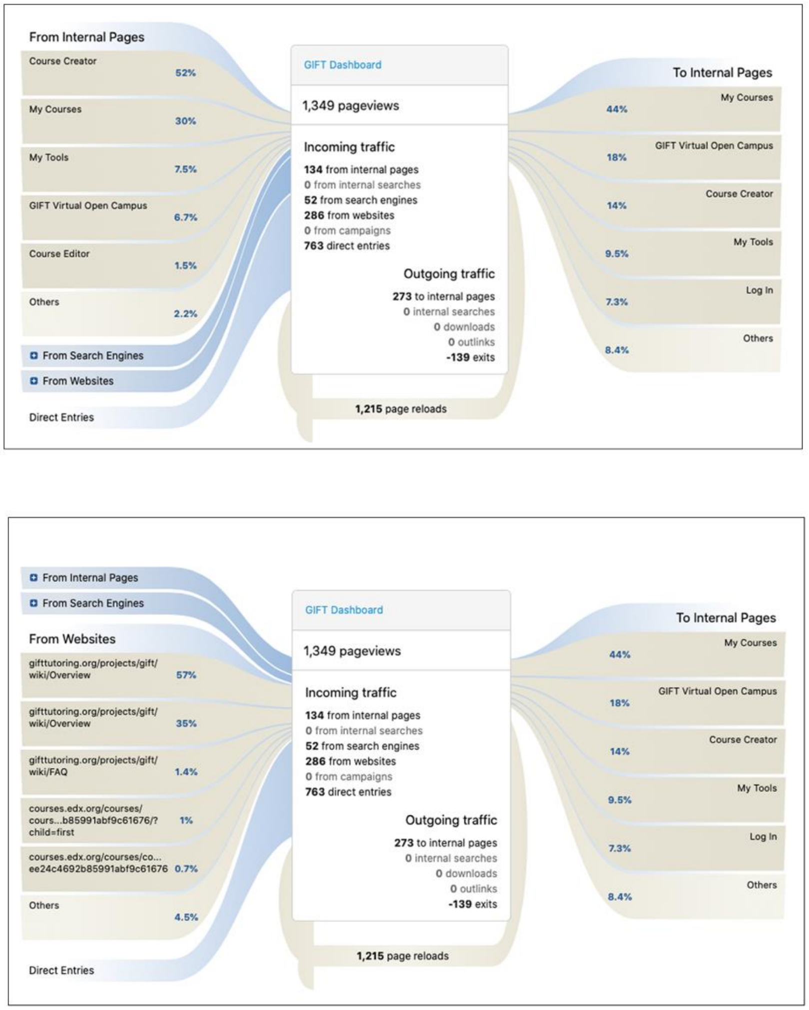


Figure 5.- Visits and Durations

There is much more data available from GIFT Cloud Data Analytics than has been shown here. The reader is encouraged to contact the GIFT team for further information if additional user behavior data and discussion are desired.

Discussion

Obtaining all user behavior data at a level of granularity correct for the application in question is an art. When properly refined by Data Scientists, this can result in some of the most useful visualizations, such as the funnel. A funnel is any series of steps taken by a user to perform a task (Melo & Machado, 2019). Data Analytics' funnels can be created when certain requirements are fulfilled such as being able to track an individual user, understand what actions they take within a page, and finally, understanding where a user navigates to. With this input-process-output data tracking format, funnels such as the ones presented below in Figure 6 can reveal some of the most telling conclusions about an application's user base. These funnels helped the authors reach several conclusions regarding user behavior that are open for discussion within the community.



GIFT Dashboard

1,349 pageviews

Incoming traffic

134 from internal pages
0 from internal searches
52 from search engines
286 from websites
0 from campaigns
763 direct entries

Outgoing traffic

273 to internal pages
0 internal searches
0 downloads
0 outlinks
-139 exits

Figure 6. GIFT Dashboard Funnel Data Visualization

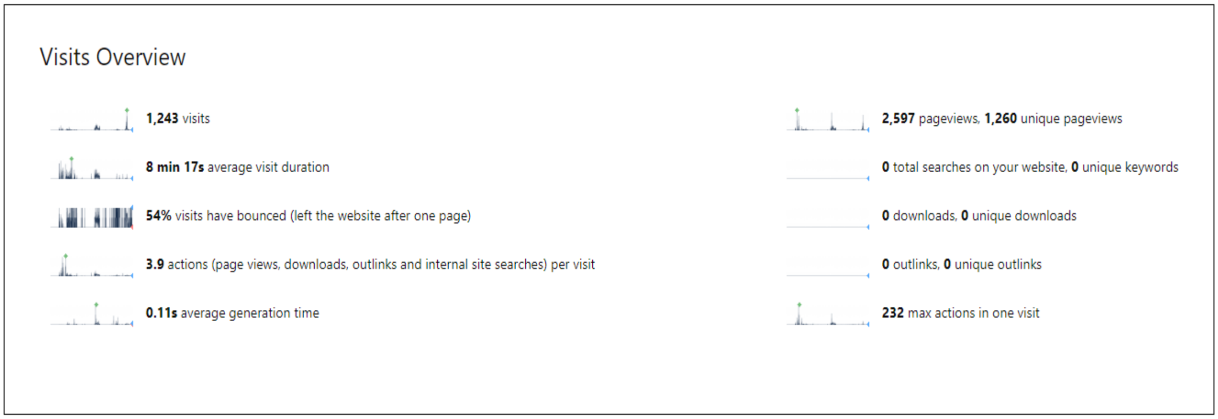
Part of Data Analytics science is statistical analysis. In an elementary sense, one can read the above information in Figure 6 as, "For every X% of users that arrived at Page XYZ (GIFT Dashboard in this case), Y% took non-trivial actions and then navigated to Page ABC." Using the above information, the reader may reach the conclusion that the gifttutoring.org website accounts for a staggering 93.4% of referral traffic to the GIFT Cloud landing page. This may be appropriate for GIFT as a military research effort, but in a commercial market, a company would aim for a larger variety of external referrers. The commercial sector also uses these types of funnels to understand what type of marketing is resulting in which variety of referrals, without the need for interpretation from a marketing representative with potential conflicts of interest.

Roughly 50% of total user visits resulted in what Data Analytics calls a "bounce," or reaching a page and then closing a browser window. These users were excluded from the data sets and funnels presented above. As shown above in the top-half of Figure 6, by the time a user has successfully logged in, the authors are confident they are "invested" in the GIFT Cloud experience. Assuming the above funnel is valid user information, an "invested" user is one that logs in, arrives at the My Courses landing page, and then navigates to a page of intended destination for the user's specific goals. Shown in the bottom-half of Figure 6, in the case of the GIFT Cloud website (from September 2016 to December 2019), the authors noticed 44% of users navigating to the My Courses page after arriving at the Dashboard. The arrival at the Dashboard was followed by a near-equal distribution of all other links with bounce rates less than 35%. Results such as these heavily imply that users are able to cleanly navigate to their chosen pages without cyclical navigation occurring – a user behavior otherwise known as aimlessly entering and exiting web pages with no clear goal.

The number of 1-to-1 funnels corresponds to the number of pages within a web application, and thus can start to be overwhelming as the number of pages increase within a system. However, key discussion points about GIFT Cloud centered around ensuring a high-quality user experience that is intuitive, effective, and can be measured for future scientific research and experimentation.

Conclusions

The inclusion of a Data Analytics Visualization tool such as Piwik was shown to be of value to the GIFT team. Using Piwik, the team was able to implement everything from high level, anonymized user tracking procedures to obtain information such as country of origin, down to low-level information such as an individual user button-click during a Visit. The ability to concretely track, organize, and customize user behavior data is one of the few and only ways to stop making assumptions about your users and see what is really going on. The GIFT community has continued to grow over the years, as demonstrated by a final aggregational subset below in Figure 7. The number of unique, non-trivial visits has increased steadily year by year. The average session is 8 minutes and 17 seconds, with an acceptable bounce rate of 54% (the best of million-dollar optimized commercial sites aim for 35%). Continuing to update and use modern Data Analytics tools is of the utmost importance for any software application, and the authors look forward to further integrations and efforts experimenting with successful user experiences.



Visits Overview

1,243 visits

8 min 17s average visit duration

54% visits have bounced (left the website after one page)

3.9 actions (page views, downloads, outlinks and internal site searches) per visit

0.11s average generation time

2,597 pageviews, **1,260** unique pageviews

0 total searches on your website, **0** unique keywords

0 downloads, **0** unique downloads

0 outlinks, **0** unique outlinks

232 max actions in one visit

Figure 7. Aggregation of GIFT Cloud Piwik User Behavior Data Statistics from September 2016 to December 2019

References

Fiaz, A.S., Asha, N., Sumathi, D., & Navaz, A.S. (2016). Data visualization: Enhancing big data more adaptable and valuable. *International Journal of Applied Engineering Research*, 11(4), 2801-2804.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: Traps in big data analysis. *Science, 343*(6176), 1203–1205.

Melo, P. N., Machado, C. (2019). *Business intelligence and analytics in small and medium enterprises*. Boca Raton, FL: CRC Press.

Metzler, K., Kim, D.A., Allum, N., & Denman, A. (2016). *Who is doing computational social science? Trends in big data research.* London, UK: SAGE Publishing.

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The generalized intelligent framework for tutoring (GIFT). *Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate (ARL-HRED)*.

Chapter 22 – Considerations for Learner, Instructor, and Researcher Roles with a Team Emphasis in the Generalized Intelligent Framework for Tutoring

Anne M. Sinatra[1], Scott J. Ososky[2], and Cheryl I. Johnson[3]
U.S. Army Combat Capabilities Development Command – Soldier Center – Simulation and Training Technology Center[1], Microsoft[2], Naval Air Warfare Center Training Systems Division[3]

Introduction

The Generalized Intelligent Framework for Tutoring (GIFT) is a research-based intelligent tutoring system framework that has been being developed by the US Army (Sottilare, Brawner, Sinatra & Johnston, 2017, gifttutoring.org). While early research in GIFT focused on the individual learner, there has always been the goal of scaling up to team and collaborative tutoring. One of the unique aspects of GIFT is that it is domain-independent, such that the framework itself is reusable, and the instructor or Subject Matter Expert (SME) provides the specific content that will be used during the tutoring interaction. GIFT also can be integrated with many different external applications including serious games such as Virtual Battle Space 3 (VBS3). The flexibility of GIFT ensures that it is a highly relevant and highly applicable tool that can be used by the Army, as well as academia and others who wish to create adaptive computer-based training.

As GIFT developed it became clear that there were multiple groups of individuals who would need to interact with the software. Among those identified groups are: learners, instructors, and researchers. Additionally, there are specific cases where there is overlap between these roles. For instance, graduate students could be learners in specific situations, but instructors in others; or instructors can also be researchers. Since GIFT is a research project, aspects of the interface have been developed over time, generally when there is a need for them or there is a specific use case. There is functionality that has been developed over time that supports the activities that each type of user would use. For instance, the learners are going to primarily interact with GIFT's course tile interface for course selection. In contrast, the instructors and researchers will primarily be using the Course Authoring Tools functions and the interfaces for extracting data from GIFT.

As the features of GIFT continue to grow, in particular toward team learning, there is increased need for interfaces that support specific users and can provide useful information to these different users through data visualizations/output. Again, the different user groups have a need for different types of visualizations: learners will be mostly focused on output of how they are performing in a course, potentially compared against others in the course. Instructors need a real-time easy-to-understand display of the current performance of the individuals in the class. Researchers need a tool that helps them to monitor the current run time of an interaction and potentially influence the software that is currently running. The additional complexity of supporting team learning is that there will need to be ways to distinguish the individual from the team in all of the described visualizations and to clearly and quickly provide opportunities to filter the data in meaningful ways.

User Roles in GIFT

There are three main user groups in GIFT, which all primarily have different goals and interfaces that they generally use within the system.

Learner

The Learner user group primarily interacts with the system through the main GIFT course tile page. Once they login they will have access to view all of the courses that are available to them. Courses can be shared with them by others with a GIFT account. Through the share function they can either have access to view/interact with the course or edit the course. The “share a course” capability is fairly recent, and by having the ability to restrict the learner from editing the course it provides a more realistic teaching scenario (as you would not want a student to be able to change the course itself or the answers to questions within the course). The learner has the ability to view their pass/fail progress on courses that they have taken. However, there is no current gradebook that links from the learner to their course performance and back to the instructor. When moving to team tutoring functionality there are some additional views that the learner may be interested in. From a performance perspective, the comparison of the learner’s grade to the class average for his or her class could be of interest. Further, being able to interface with others that are their teammates in the system is a relevant functionality. During the interaction with the course, either the GIFT interface or the training application will need to provide a way for the learner to communicate with the others in his or her team. If the learner is not logged in with their own GIFT account, then they can be sent a link from “Publish Courses” that allows them to participate in the GIFT course without needing to login. In this case, there will need to be a specific survey question provided that asks for the learner’s name or a student ID number so that it can be matched back up to the instructor’s records.

Instructor

The Instructor user group primarily interacts with the system through the Course Authoring Tool, and data extraction tools. The instructor creates a GIFT course by using the functions within the Course Authoring Tool, and then after it is completed it can either be shared with the learners either through them logging into the main GIFT interface or through a published link from the course. The instructor will also need to be able to extract the data from the learner after the fact. Additionally, instructors will want to have the flexibility to examine learner responses on individual questions and scenarios in order to see what concepts learners are doing well or poorly on. In the current form the primary way that this can be done is by using the *publish course* function, which allows for easy extraction of the data using the web tools. If the preferred way to interact is through the GIFT login, then it is more difficult for the instructor to access the logs. Therefore, it is beneficial to add a function that allows the instructor to have access to the logs for individuals in his or her specific class.

Experimenter

The Experimenter user group will use many of the similar interfaces that the instructor will, however, they may use them in a different way. For instance, the instructor is going to be focused on visualizing data in terms of a gradebook output in a way that is easy to understand and can also be viewed/understood by students. However, the experimenter is not interested in specific grades of individuals in the same way. The experimenter is more interested in collecting overall data that is de-identified and has a specific participant number attached to it for later data analysis. There are now some functions within GIFT, visible through the Game Master interface (Goldberg, Hoffman,

& Graesser, 2020) which allows for an experimenter to monitor performance during the experiment and make notes throughout.

Challenges in the Design of Dashboards in GIFT

Among the challenges that exist for creating Dashboards for different users in GIFT is that GIFT is domain independent and requires consideration to be given to all the different use cases. Just as it is difficult to determine the best design for tools to meet the needs of all the different user types, it is also necessary to keep those tools flexible enough that they can be used for a number of different applications and domains. Creating a tool that allows for intelligent tutors to be constructed in domains as varied as algebra and marksmanship is a difficult challenge. At the same time, there is the need to make the tool user friendly and be able to support the use of instructors, subject matter experts, and content creators who may not have much experience with authoring materials on a computer.

An additional need in the creation of dashboards is that they can be scaled up to support both individuals and teams. One way to approach this on a general scale without knowing the specific academic classes or needs is by providing customizability and filtering functions. One key step is that the most commonly used features and needs are decided on, and they are included in the main interface. The features that are not used as often, but are needed, can be included in sub-menus, or as filtered options (e.g., a check box could be checked to show/hide the selected information). This practice of highlighting key features within interfaces is *progressive disclosure* (Lightbown, 2015, p. 112). One approach that can be used to make these decisions is to do a task analysis for each of the user roles, and determine what features are common among all three of them. Alternatively, dashboard systems within GIFT could adapt views and available visualization templates based on the user's specific role (e.g., Learner, Instructor, Experimenter/Researcher). This may be useful for cases in which little overlap is found between roles in task analyses.

The same design principles that are used throughout the GIFT interface should be followed in any dashboards that are created. This will help ensure that learning one interface will assist in using the other ones as well. The information in the data visualization output should be flexible, and include the ability to examine performance at the individual level, the class level, and any sub-group levels. By doing this, it also allows for the structure to be ported over for use in team exercises and military-related tasks, as this would map to the individual, the overall squad, and the fire team. There should be a way to quickly and efficiently filter the data and switch between these different views. If the instructor is examining the information in real-time, it will be important for them to have easy access to the information in order to make quick decisions based on it to guide the feedback that is provided to the learners.

GIFT can be used in a variety of contexts including academic and military use alike. There will also be challenges in creating a common design language around dashboards that are appropriate for different contexts. For example, the use of colors, shapes, and phrases on a dashboard may have a very specific meaning in a military context but be irrelevant or difficult to understand in an academic context, and vice versa. These style considerations represent yet another layer of challenge sitting on top of user roles, information architecture, flexibility, and consistency with other interfaces within GIFT.

Current Approaches to Displaying Information

GIFT currently has visualizations in the form of an After Action Review (AAR) Course Object, the Event Report Tool (web-based and desktop-based), and the Game Master interface. Each of these tools is specifically aimed at a different user group. For instance, learners will interact with the AAR course object as a part of a course that they are taking. After the learner has completed an assessment, or as part of an adaptive course flow process, they will see an AAR that shows their performance on questions and on specific course concepts. This interface will help the learner understand their specific performance, which they can use to help them adjust their responses going forward. The Event Report tool is available in a full desktop version (to extract local computer data logs), and an abbreviated web-based version to extract data from a published course in GIFT. By using these tools, researchers (and advanced instructors) can extract the data that has been collected from learners in GIFT. There are checkboxes which allow the individual using it to select which data they would like to extract. If an instructor is using the system, they likely would only select survey responses, and merge the file so that each row represented a different learner. While this is not a formal visualization, it provides output in a .csv file which can then be viewed in Excel and the data visualized. Finally, the Game Master interface provides a visualization tool that can be used in real-time by primarily instructors (however, it may have some relevance to researchers). The Game Master interface can be viewed on a tablet and shows the concepts for assessment in real-time. The Game Master interface is the closest to a traditional data visualization interface that currently exists in GIFT.

Designing Dashboards for the Submarine Electronic Warfare Adaptive Trainer: A Use Case

Although the Submarine Electronic Warfare Adaptive Trainer (SEW-AT) was built independently from GIFT, its development offers a helpful use case and lessons learned that apply to the design of dashboards for different user groups in GIFT. SEW-AT was initially developed as a research testbed at NAWCTSD, and it is now a fielded prototype (see Van Buskirk et al., 2019). SEW-AT is a scenario-based trainer that simulates a submarine electronic warfare (EW) operator’s trip to periscope depth. Submarine EW is a very time-sensitive and dynamic task that involves the operator listening to many concurrent radio frequency (RF) signals, examining complicated RF real-time displays, and making reports to Control about this environment at regular and irregular intervals, depending on the events occurring during the scenario. Key metrics include the accuracy of the reports provided by the EW operators in addition to the timeliness of these reports, which are used to calculate overall scenario scores. SEW-AT provides an individualized training experience, such that scenario difficulty is tailored based on a trainee’s performance during training to present trainees with an optimal level of scenario difficulty; therefore SEW-AT includes scenarios of basic, intermediate, and advanced difficulty.

While SEW-AT was under development, the research team sought iterative feedback from stakeholders as often as feasible; stakeholders included operators, instructors, radio chiefs, senior EW SMEs, EW policy writers, and others. Interestingly, a performance dashboard to explore users’ performance over time was one of the most requested features from all parties. To define requirements for performance dashboards, the team held informal focus groups with stakeholders to determine user roles, the type of data they desired for each role, and their goals for using the data in each role. As it turns out, there were differences in the amount of data stakeholders wanted to see; some were interested in scenario-specific details, while others preferred a more holistic, overall score. To this end, the overall approach to the design and selection of these dashboards was to take a top-down perspective, such that overall, high-level performance data is presented first to display overall performance trends (i.e., overall scores by scenario and the level of difficulty). Then users can drill down to get more specific data on report accuracy and timeliness about particular scenarios (and specific trainees, depending on the user’s role) as they wish. Next, the team mocked up a few data dashboards and presented different versions of each user role to solicit specific feedback from the stakeholders. Users preferred to view the data in a tabular format to see performance across the different reports at a glance, rather than a series of line or bar graphs or one “busy” graph with several variables. The users desired as intuitive and easy-to-understand interface as possible, so hyperlinks were utilized to allow users to step through the data as they wished, without requiring them to learn how to use a new tool to build their own reports.

For SEW-AT, there are at least three user roles. The first role is that of Student/Operators who perform scenarios and desire to see feedback on their performance. Some of the SEW-AT users are students currently training to become EW operators, and other users are EW operators onboard a submarine who are using SEW-AT to practice before deployment qualifying events or to maintain proficiency while at sea. Focus groups with these users revealed that they wanted a capability to review their performance over time and to see how their scores broke down by the scenario elements. That is, for the most part, this user group was interested in drilling down to specific scenario elements with the goal of using that information to improve their performance during later scenarios.

The second role is the Instructor/Supervisor. This user role includes instructors at the schoolhouse and radio chiefs or supervisors at sea. In both cases, the Instructor/Supervisor is interested in data from multiple users, such as a whole class or a watch team. During the focus group sessions, users in this role stated that they often coached over the shoulder and would appreciate an easy way to visualize how student/operators are performing. To that end, the team designed a dashboard with red/yellow/green indicators to highlight areas where student/operators might be struggling at a glance, so that the instructor/supervisors can focus on those particular topics for remediation. In this role, instructor/supervisors can select which student/operator user data profiles to include in their dashboard and view overall trends for their students. In addition, this role has the ability to click on individual student/operator records and access the same low-level information about a specific scenario that the individual student/operators can view.

The third role is the Researcher who conducts experiments to explore the benefits of particular instructional strategies on learner performance. In this case, the users were the research team members who developed SEW-AT. They did not develop specific dashboards for this user role, since SEW-AT was intended to be used by the Fleet rather than as a general-use research testbed. Individuals in the researcher role explored data from the lowest level of data granularity (e.g., button clicks) to the highest level (e.g., aggregated score reflecting overall performance on a scenario) to examine a multitude of research questions including: assessing system usability, gauging and adjusting difficulty of scenarios, refining adaptive algorithms, and performing training effectiveness evaluations. The research team is currently exploring data at the button click level to perform advanced diagnosis of performance failures to improve remediation and feedback algorithms.

The key takeaway from discussing this experience was to understand the users and their goals, which are often different, when designing data visualization tools. In the case of SEW-AT, the research team spent considerable time engaging with different users from both ends of the spectrum (and everywhere in between) to get a broad understanding of who typical users are, the ways they would approach the data, and how they would use that data in order to develop effective dashboards for each user type. It is important to consider the key metrics that each user group considers important to ensure that the dashboard design meets their needs. For SEW-AT, some users only wanted a summary of performance over time, while others desired a more fine-grained view with data on how student/operators performed by report for each scenario; therefore the team had to design the dashboards in a way that would satisfy both camps. In addition, it was helpful for the research team to mock up different dashboard prototypes to get feedback from the intended users of these dashboards. As data scientists, researchers are used to designing, presenting, and consuming a variety of data visualizations, but that is not always the case for users of fielded systems. It is critical to remember who the intended audience is and to ensure you are designing dashboards that will be useful for them. The process and lessons learned from the SEW-AT example are highly relevant to GIFT, as the general user groups line up between the two systems.

Dashboard Considerations and Implications for Team Dashboards

First and foremost, users should be at the center of any dashboard solutions and/or designs. Successful dashboard implementations will involve users as early as practical in the design process, and continue to engage users as solutions are designed. For example, interviews and contextual inquiry should be conducted with potential users to understand their needs and goals in their roles. Preliminary designs can be tested with users to determine if they are useful and usable. Continued evaluations should take place as dashboards are being built to ensure that design and engineering effort is still tracking with users' needs. With that in mind, the following sections discuss specific considerations for dashboards based on the user's role in learning and/or training.

Learner Dashboard Considerations

Learners will primarily be data consumers, as opposed to dashboard creators or managers. As such, learners will likely have lower engagement with a team dashboard compared to instructors and researchers. With that in mind, designs for team dashboards for learners should be guided by relevancy, clarity, and action. First, learners will require team dashboards that provide information that is relevant to their actions and outcomes, how their actions ladder up to the team's outcomes, and how interactions between individuals affect one another. Second, team dashboards for learners should provide clarity in the sense that dashboards reveal information that helps to answer specific questions while also deliberately withholding information that is irrelevant or otherwise impeding upon a learner's ability to seek answers to questions about their actions and outcomes. Third, learner dashboards should aspire to help a learner to understand what they need to do to improve, or what to keep doing to maintain a certain level of performance. Additionally, an actionable dashboard might also be one that evokes a useful emotional response in order to benefit the user, such as motivating, praising or encouraging a learner.

Instructor Dashboard Considerations

The instructor dashboard represents a central pillar of the team dashboard user experience. If an instructor cannot figure out how to use or configure a team dashboard, finds it too complicated, etc. that will negatively impact the experiences of learners (and possibly researchers) with team dashboards. That negative impact could result from an instructor using a dashboard incorrectly, or discontinuing use all together in favor of some other solution. Therefore, designs for instructor team dashboards should be guided by learnable, usable, and flexible experiences. It is also recommended that if a team dashboard could only be designed with one type of user in mind, that it be an instructor. Instructors, whether they be educators in the public sector, or training professionals in the military, likely do not have the time to prioritize learning yet *another* tool or system over other responsibilities such as creating lesson plans, interacting with learners, managing assignments, and doing other types of professional development. Consequently, an instructor dashboard needs to be learnable in the sense that the experience has a low barrier to entry with intuitive features requiring little or no training. Of course, the dashboard needs to be usable in the sense that (after learning the system) configuring, sharing, and managing dashboards is a continuous, positive experience, following standard usability guidelines. Finally, instructor dashboards need to be flexible to accommodate different teaching styles, class formats, learner populations, and so on. The design of a team dashboard should also consider the difference between how an instructor uses a dashboard during vs. after a learning/training experience. It is recommended that a team dashboard provide flexibility through a series of templates that can be modified through a configurable, modular user interface. Instructors should be able to save their own team dashboard templates and share them with other instructors.

Researcher Dashboard Considerations

Researchers could be thought of as being like a *power user* group within a productivity application. Researchers often go deeper into the features of a dashboard than the other groups, while also having the most tolerance for tools that are complicated or difficult to use. Considerations for researchers should be secondary to those of instructors and learners; however it is recommended that researcher dashboards be a superset of instructor dashboards for the sake of consistency and continuity across experiences. In fact, it is suggested that the researcher serves as a potential testing population for new features intended for instructors. With that in mind, researcher dashboards should be guided by extendable, and reliable experiences. First, team dashboards for researchers will likely require functionality that is not intended as part of the original dashboard design; it is recommended that team dashboards have the ability to provide new code or plug into other systems in order to extend the functionality of current dashboard solutions to accommodate different study designs. Second, researcher dashboards need to be reliable. Here, reliable refers to the integrity of the data, not necessarily of the team dashboard itself. Because quality data is central to a researcher's work, all assurances should be made that the data is protected against loss and unintended modification through backup systems and semi-automated integrity checks. Finally, researcher team dashboards should represent the latest and greatest development and design efforts of a dashboard system, while also paying special attention to usability. Doing so empowers researchers to pursue new types of research, while also providing a sort of testing of new functionality before it is rolled out to instructors and students.

Suggestions for Improvements to GIFT

As GIFT matures and its audience broadens, additional data visualization features would be appealing for widespread transition into instructors' courses. Learners, instructors, and researchers who may want to use GIFT would greatly benefit from additional data visualization tools. Following a similar process to what was done with the SEW-AT project, and interacting with the user groups would be a good first step for improving the current GIFT Game Master interface. One of the challenges in GIFT is staying flexible to the domain, however, the overall needs of the user groups could potentially be generalized such that the domain independent tools can support some of the most requested features.

Looking ahead to remote training / learning

It is highly recommended that GIFT consider the way in which training and/or learning will take place in a post-pandemic world where *remote* learning and training may be more common. This is a rapidly evolving space, with the immediate future of education still uncertain at the time of this writing (Hobbs & Hawkins, 2020). While it may be too early to know how learning and training will change in various contexts, remote learning may remain a part of our society for the foreseeable future (Dickler, 2020). GIFT has a tremendous opportunity to be part of the remote learning conversation. GIFT is already able to meet some needs of remote learning and training, as it is hosted online. Dashboards specifically may be even more important and urgently needed than before the pandemic.

For example, GIFT may see an increased number of users engaging with the platform remotely, with fewer users engaging with GIFT in a co-located setting (i.e., classroom). Therefore, additional scaffolding and support may be needed to help users understand how to make sense of a dashboard without an instructor or other learner present to discuss the output that is shown on dashboards. Additionally, learners may need additional functions that allow them to highlight specific aspects of a dashboard to discuss with an instructor to follow-up and/or ask questions about their performance. Likewise, researchers may need additional tools to understand how research participants are engaging with GIFT courses in a remote learning/training context (e.g., time spent in a course, hardware used, browser characteristics). Finally, both researchers and instructors may require tools that allow teams of users to co-create custom dashboards and visualizations.

It is not recommended that GIFT *should* have each of those functions included as part of its own platform. Instead, it is recommended that GIFT researchers and developers consider how GIFT usage may change based on societal changes, and then determine the best course of action. In fact, that may result in new or different functionality for GIFT; or, it may result in tighter integration between GIFT and some other collaborative work tool (e.g., Slack or Microsoft Teams). The benefits of enabling remote learning, collaboration, and creation will create a positive impact for GIFT both in the near-term, as well as in the longer term as a generalized platform accommodating many different types of use cases and users.

Conclusions

In summary, interacting with the actual users and determining the common requested features will be very important as data visualization tools are developed in GIFT. By having conversations with learners, instructors, and researchers who intend to work with the system, it will identify what features are currently missing and what is most important to implement. User feedback will be critical to developing the Game Master interface, as well as additional interfaces to display course output and gradebooks to both learners and instructors. These conversations and task analyses will also provide insight into what current features are successful in GIFT, and how to make them even better in the future.

References

Dickler, J. (2020) *Post-pandemic, remote learning could be here to stay*. CNBC.com.https://www.cnbc.com/2020/05/20/post-pandemic-remote-learning-could-be-here-to-stay.html

Hobbs, T.D., Hawkins, L. (2020). *The results are in for remote learning: It didn't work.* The Wall Street Journal. https://www.wsj.com/articles/schools-coronavirus-remote-learning-lockdown-tech-11591375078

Goldberg, B., Hoffman, M., & Graesser, A. (2020). Adding a human to the adaptive instruction system loop: Integrating GIFT and Battle Space Visualization. In *Design Recommendations for Intelligent Tutoring Systems, Volume 8: Data Visualization*.

Lightbown, D. (2015). Designing the user experience of game development tools. CRC Press.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT). *GIFTtutoring. org*.

Van Buskirk, W. L., Fraulini, N. W., Schroeder, B. L., Johnson, C. I., & Marraffino, M. D. (2019, July). Application of Theory to the Development of an Adaptive Training System for a Submarine Electronic Warfare Task. In *International Conference on Human-Computer Interaction* (pp. 352-362). Springer, Cham.

Acknowledgements

The views and conclusions contained in this book chapter are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Combat Capabilities Development Command - Soldier Center, Naval Air Warfare Center Training Systems Division, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright n

Chapter 23 – Adding a Human to the Adaptive Instructional System Loop: Integrating GIFT and Battle Space Visualization

Benjamin Goldberg[1], Michael Hoffman[2], and Arthur C. Graesser[3]
[1]US Army Combat Capabilities Development Command – Soldier Center – Simulation and Training Technology Center, [2]Dignitas Technologies, [3]University of Memphis

Introduction

In this chapter we introduce an extension to the common Adaptive Instructional System (AIS) interaction space by adding an instructor to the tutoring logic chain. While this capability is driven by an Army modernization priority, the utility of this approach can apply in multiple contexts across academic and commercial settings. The objective is to provide visualization tools and User Interfaces (UIs) that allow an instructor to actively monitor an AIS driven interaction, along with tools and methods to drive the adaptive instructional experience. This involves exposing underlying logic and assessment states as an individual or team completes a scenario or problem set, along with available injects and scenario adaptations to influence the learning outcomes. Under this assumption, an AIS must be designed to expose performance to the observer along with recommended strategy injects, but system level actions are managed by the human counterpart.

In the following sections, we present on-going work to develop an initial collaboration AIS capability. The proof of concept is achieved through the integration of two Department of Defense (DoD) open source technologies, the Generalized Intelligent Framework for Tutoring (GIFT) and Battle Space Visualization (BSV). To drive the discussion we will provide: (1) background information on Army training modernization programs and the human-observer role being considered for technology design, (2) an overview of the technologies at-play, (3) the enabled Intelligent Exercise Control concept created through integration, (4) modifications to the GIFT framework and UI to enable the human-on-the-loop collaboration model, and (5) visualization needs and considerations that are based on human factors and cognitive load management strategies.

The Synthetic Training Environment (STE): A Driving Requirement

The U.S. Army is undertaking a large modernization strategy targeting its current simulation-based training technology base with the development of the Synthetic Training Environment (STE; Gervais, 2018). The STE looks to take advantage of advancements in gaming and virtual/augmented reality interfacing to deliver a new training capability that provides high-fidelity interactions with realistic stimuli. The environment is intended to be quite extensible, with configurations to train units at the joint and collective level across dismounted, mounted, and mission command contexts. A critical component of STE are a set of Training Management Tools (TMT), that support proponents in Planning, Preparing, Executing and Assessing an overarching STE training event. Within the TMT portfolio, intelligent tutoring and AIS functions have been defined as system level requirements to support efficient training delivery and promote skill acquisition. From a traditional viewpoint, this requirement targets learning science and Artificial Intelligence (AI) techniques to assist in developing, delivering, and evaluating training scenario interactions that are embedded with automated assessments and instructional supports that are configured to guide the training experience, e.g., give feedback and adjust scenario complexity when appropriate (Woolf, 2009; Goldberg, Brawner, Holden & Sottilare, 2012).

While intelligent tutoring functions are defined as critical capabilities to STE at-large, the nature in which these technologies will be applied vary from the traditional sense considerably. Typical intelligent tutors operate in a closed-loop self-regulated environment with minimal interaction from a coach and/or instructor during execution of a scenario or problem set. In this instance, the computer facilitates all interaction with the trainee, with underlying models to infer domain knowledge and skills that guide pedagogical decisions delivered through the system interface (VanLehn, 2011). In contrast, STE assumes human observer(s) to be a critical component to training delivery and execution. Based on this interaction paradigm, intelligent tutoring and AIS technologies must be implemented in a way to better support that observer, rather than replace them. The interaction space must account for this new variable, with AIS tools and methods being modified and extended to support a human-on-the-loop collaboration model. This creates a unique human-machine team that aims to: (1) balance the assessment space with automated data-driven techniques wherever feasible and leverage human expertise for performance constructs not yet supported by AIS processes, (2) free up human-observer workload to better focus attention on complex phases of a scenario involving team dynamics, and (3) optimize real-time pedagogical decisions by using AIS logic to recommend injects, but leave management decisions to the human teammate. This highlights a requirement to track Observer Controller/Trainer (OC/T) experiences to enable automated instructional components to adapt to the human expert and not just the trainees. A critical component for this new collaboration model is carefully establishing visualization techniques to expose underlying AIS logic for easy interpretation of the training environment and to enable realtime human-driven injects.

Training Facilitation with an Observer Controller/Trainer

As described above, human observers play a critical role in managing collective training events. Traditionally, training delivery is managed by an individual called an OC/T. OC/Ts are often Subject Matter Experts on doctrine and within a specified warfighting function. These functions include, but are not limited to: mission command, movement and maneuver, fires, sustainment, protection, and intelligence. During a training event, OC/Ts play a critical role. They provide informal feedback to a unit during scenario execution; they also provide formal coaching through structured mid and final after action reviews (AARs). It is the job of the OC/T to provide the training audience with discernible actions to consider for performance sustainment and improvement. These are based on doctrine derived assessments that link to task level procedures and team-level interactions. Assessments are based on a set of binary go/no-go determinations across a set of Training and Evaluation Outlines (T&EOs) managed by the Army’s Training and Doctrine Command (Foo, 2019). This provides a structured rubric for determining a unit’s readiness based on a set of established performance thresholds.

For a single unit, multiple OC/Ts are often applied. This helps control the assessment space an OC/T is required to support, with an effort to manage workload and establish consistent scoring procedures. However, we recognize two potential issues that this research aims to address: (1) assessments are highly subjective in nature, and are dependent on the skills and experience of an OC/T, and (2) high-tempo scenarios in dynamic environments create challenges to maintaining situational awareness across all assessment criteria, resulting in post-training assessments being completed several minutes after the conclusion of training. In recognition of these issues, AIS functions are being explored to: (1) automate as much of the assessment space as possible based on available data sources, and (2) reduce the number of required OC/Ts through the extension of AIS capabilities to support robust performance tracking and exercise control. In lieu of this objective, we present a proof of concept in the following sections that creates an OC/T Game Master capability. The Game Master concept combines AIS logic with visualization features to create the AIS collaboration model required for STE.

Training management TOOLS for adaptive instruction

The TMT paradigm for managing adaptive instruction described in this chapter combines technology from two service-oriented research programs: (1) Battle Space Visualization through the Augmented Reality Sand Table (Ares) and (2) adaptive training functions through GIFT. This integration creates a STE capability to enable interactive exercise control with robust data visualization and AIS supports to track performance and recommend coaching interventions. Each technology is described below, with specific attention focused on their applications during the execution of a training event within a simulation environment.

Battle Space Visualization via the Augmented Reality Sand Table (Ares)

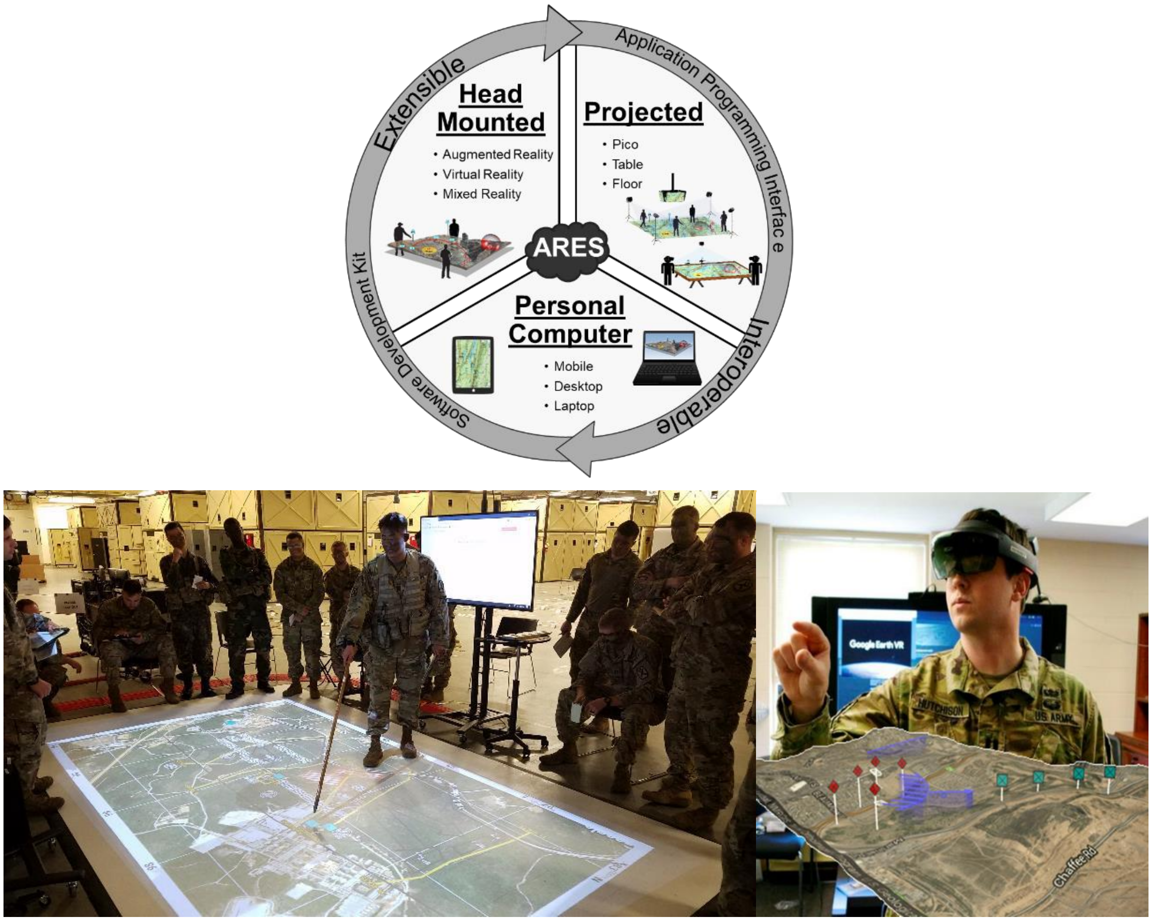


Figure 1. Augmented Reality Sand Table (Ares) visualization modalities, with floor projection and augmented reality examples.

The Ares is a technology research and development testbed comprised of commercial-off-the-shelf hardware components integrated to enable multi-modal data visualization (Garneau, Boyce, Shorter, Vey & Amburn, 2018). An initial objective of the program was to improve battle space visualization to support a user-defined common operating picture for a designated Area of Operation (AO) as represented by specified map coordinates. This enables visualization of multiple data layers, models and services over a specified AO, with a large set of use cases driving its application. To extend the level of interaction with data, Ares renders data in multiple modalities for consumption by a defined user. These modes of interaction include: (1) traditional sandtable using actual sand with a visualized top-down constructive view of the AO with 3-

dimensionsal topography, (2) floor projection system with a top-down constructive view of the AO (see Figure 1), (3) virtual reality headset (e.g., HTC Vive, Oculus Rift) for 3-dimensional rendering of the AO from a designated point of view, and (4) augmented reality headset (e.g., Microsoft Hololens, Magic Leap One) for 3-dimensional visualizations in one's physical environment. It is worth noting that the augmented reality capability can work with sandtable and floor projection modes, further establishing context in a confined two-dimensional constructive view.

For the purpose of this project, Ares is being used to consume and visualize real-time training interaction data being generated during a simulation-based exercise. The type of data in play will be discussed below. We envision OC/Ts as the driving stakeholder in this design. We want to apply Ares visualizations so an OC/T can efficiently track unit movements and behaviors as they relate against a set of defined mission objectives. Another requirement is linking visualization techniques with evolving contextual information being produced by an AIS component. This involves notifying the OC/T that a performance assessment was just observed, rapidly orienting the OC/T to where in the training environment that assessment just occurred, and exposing exercise control functions that can be applied based on an individual's or unit's performance. To support this requirement, we integrated GIFT with Ares.

Simulation Data and the GIFT Learning Effect Chain

In the context of this chapter, the main feature of GIFT that is influencing visualization requirements is the real-time assessment functions applied during execution of a training event. One of GIFT's original design features was its ability to interface with third-party game-engines for the purpose of providing stealth assessment functions (Shute & Ventura, 2013; Shute, Ventura, Small & Goldberg, 2013). Much like Ares, GIFT was designed to interoperate with multiple data sources, but in this instance, the system consumes raw event data to infer performance against a set of tasks, conditions, and standards, and sensor data to classify affective (e.g., fear, anger, frustration) and cognitive (workload, fatigue, etc.) associated states. Ultimately, these data processes are applied to translate raw data into measures of performance and effectiveness. The message flow to support this inference procedure was termed the Adaptive Training Learning Effect Chain (ATLEC, see Figure 2; Sottilare, Ragusa, Hoffman & Goldberg, 2013).

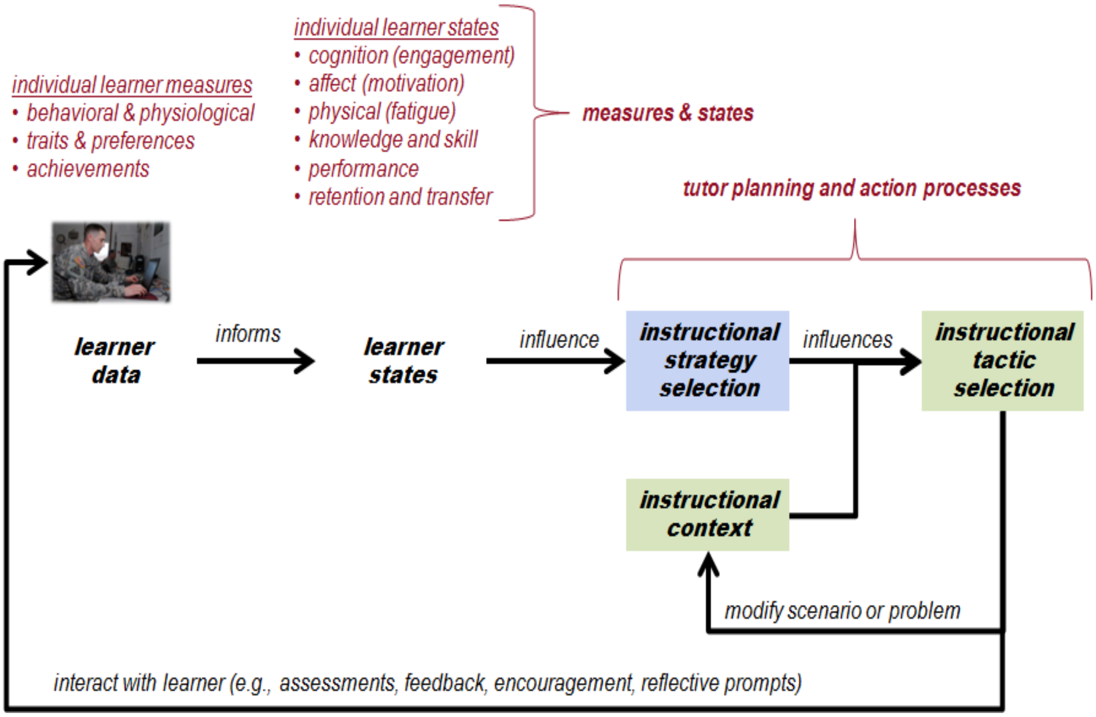

individual learner measures
• behavioral & physiological
• traits & preferences
• achievements
individual learner states
• cognition (engagement)
• affect (motivation)
• physical (fatigue)
• knowledge and skill
• performance
• retention and transfer
measures & states
tutor planning and action processes
learner data
informs
learner states
influence
instructional strategy selection
influences
instructional tactic selection
instructional context
modify scenario or problem
interact with learner (e.g., assessments, feedback, encouragement, reflective prompts)

Figure 2. GIFT's Adaptive Training Learning Effect Chain (Sottilare et al., 2013)

The ATLEC is based on a set of data processes and specifications GIFT adheres to when managing a run-time training interaction. These processes leverage the core GIFT modules (i.e., Domain, Trainee/Unit, Pedagogical, and User Interface modules), where: (1) learner data is consumed and translated into learner states (e.g., performance, affect, cognition), (2) learner states are combined with historical learner data to influence an instructional strategy (e.g., providing feedback, asking a question, adjusting the scenario, etc.), (3) an instructional strategy is converted into a domain contingent instructional tactic (e.g., providing a specified feedback string on a designated concept), and (4) observing if the instructional tactic impacted performance.

This whole experience is configured in a GIFT Domain Knowledge File (DKF; Domain Knowledge File Documentation, 2020). The DKF is a domain-agnostic eXtensible Markup Language (XML) schema used to establish a task model with designated event triggers, assessment logic, and available instructional tactics to influence the training experience. This piece of GIFT is serving as the AIS engine driving the Game Master collaboration model described next.

AIS Collaboration through a Game master concept

To see an architectural breakdown of the Game Master concept in a collective training context, see Figure 3. This diagram shows the conceived data flow across all interacting components, and operates on two assumptions. The first is that all trainees are interacting within a simulated environment that produces granular event data that can be consumed by both GIFT and Ares. Both platforms utilize a Gateway Module to manage these interoperability configurations, with multiple standards currently supported (Distributed Interactive Simulation [DIS], High Level Architecture [HLA], Global Positioning System [GPS], Communication [COMMS], etc.). The second assumption is that other data sources are available via behavioral and physiological sensors to better track training engagements and provide diagnostic information related to performance outcomes. It is important to note that this approach does not rely on sensor technologies, and automated assessments are limited when based on game event information alone.

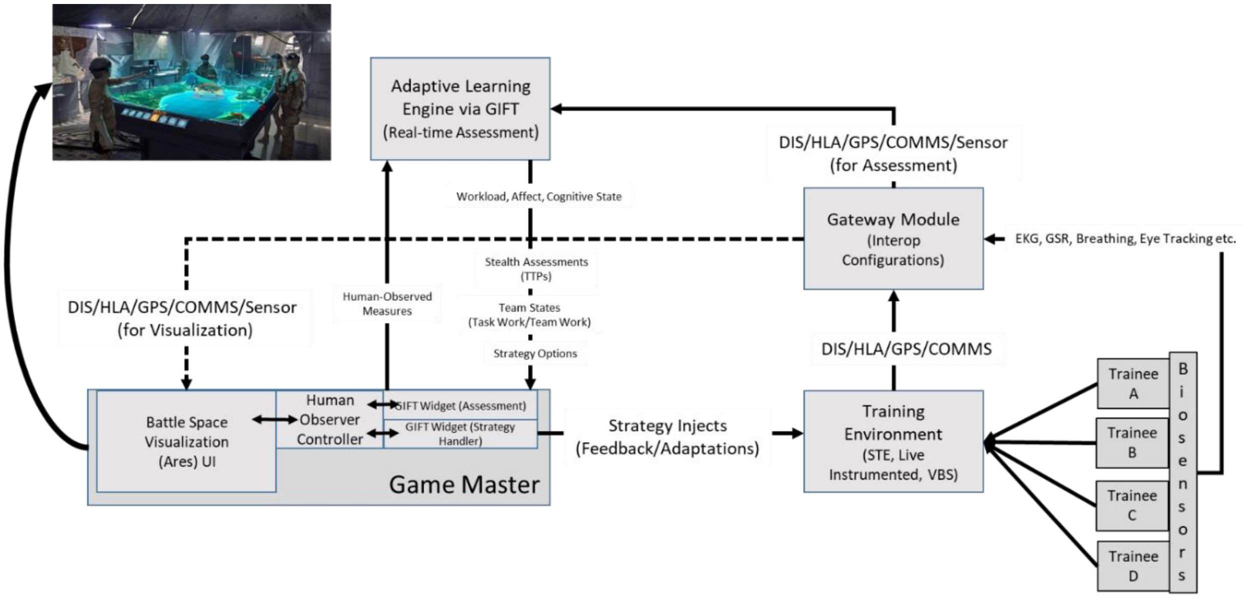

Adaptive Learning Engine via GIFT (Real-time Assessment)
DIS/HLA/GPS/COMMS/Sensor (for Assessment)
Workload, Affect, Cognitive State
Gateway Module (Interop Configurations)
EKG, GSR, Breathing, Eye Tracking etc.
Stealth Assessments (TTPs)
Human-Observed Measures
DIS/HLA/GPS/COMMS/Sensor (for Visualization)
Team States (Task Work/Team Work)
Strategy Options
DIS/HLA/GPS/COMMS
Battle Space Visualization (Ares) UI
Human Observer Controller
GIFT Widget (Assessment)
GIFT Widget (Strategy Handler)
Game Master
Strategy Injects (Feedback/Adaptations)
Training Environment (STE, Live Instrumented, VBS)
Trainee A
Trainee B
Trainee C
Trainee D
Biosensors

Figure 3. Game Master Flow Diagram Concept with GIFT/Ares Integration

During a training exercise, the Game Master Gateway Module consumes real-time raw data and passes it to Ares for visualization across a designated modality and to GIFT for real-time assessment functions. As a group of trainees complete a scenario, a human OC/T monitors their interaction using the Ares Floor Projection System (see Figure 1). While the OC/T visually tracks movement, they hold a tablet with a Game Master UI that exposes two widgets: (1) assessments in real-time based on GIFT DKF rules and (2) available instructional tactics an OC/T can inject into the training environment. During run-time, GIFT is updating the OC/T with data-driven assessment functions that include team performance, individual performance, affect state updates and recommended strategies to execute. In addition, the OC/T has the ability to push human-observed assessments into the task model, further extending the assessment space where data-driven methods lack maturity. To enable the Game Master concept, modifications to the GIFT architecture were required to support a human-on-the-loop AIS collaboration model.

GIFT Modifications

Two main engineering tasks were completed when we implemented the first iteration of the GIFT Game Master concept. This included adding human interaction points within the GIFT Learning Effect Chain and developing the interface to support these new data points. Each of these tasks are described in detail in this section.

Adding the Human to the Learning Effect Chain

Several changes were needed in the GIFT architecture in order to support an OC/T applying custom changes to the learning effect chain of GIFT. The majority of these changes happened in the Domain Module where automated assessments and pedagogical requests are handled. One of the first features exposed to the user was the ability to manipulate the five metrics (Performance Assessment, Priority, Confidence, Competence, Trend) assigned to every Task and Concept, with performance assessment being the most prevalent and useful to date. Now the OC/T can manually provide an assessment value of Below, At or Above expectation at any point during the lesson. However ongoing learner actions such as moving around in the training environment might cause an automated assessment to then over-ride the OC/T's assessment. This situation prompted the need for an OC/T to lock any of the metrics from being changed by automated algorithms in the Domain module. Every metric change is communicated from the Dashboard, where the Game Master is currently located, to the Domain Module through the GIFT ActiveMQ message bus via a new Evaluator Update Request message. The message is then routed to the appropriate domain knowledge representation for that domain session and applied to the appropriate task or concept. This causes a chain reaction that produces a performance assessment message and learner state message, both of which have the updated metric information set by the OC/T through the Game Master application.

A new observed condition was created in the Domain Module to facilitate assessments for which GIFT is not equipped with an appropriate algorithm or the learning environment is not properly instrumented to collect the necessary information. The sole purpose of this condition is to indicate that there will be no automated assessment but rather the OC/T must manually provide an assessment, when appropriate, for the condition's parent concept. This is a powerful addition to GIFT because authors can now eliminate the need to hire software engineers to build condition classes to assess domains GIFT has yet to handle or purchase and configure sensors or other hardware in the training environment because the burden of assessment is placed on the OC/T.

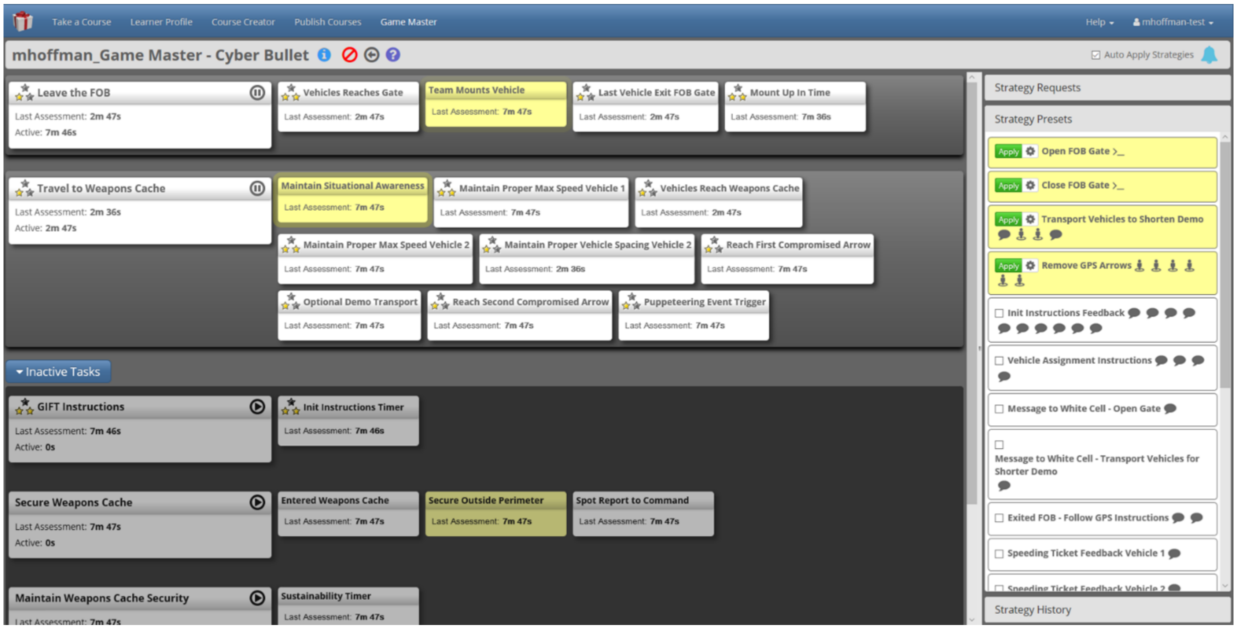
Another feature provided to the OC/T is the ability to control when tasks are active or inactive. Although it is required to define the life cycle of a task for a real time assessment when authoring, the OC/T can now decide when to activate a task for assessment or deactivate a task, thereby stopping any ongoing assessment. To achieve this, we extended the Evaluator Update Request message to also contain an attribute that defines whether a task should change its running state.

The final part of the learning effect chain is pedagogical requests. Before the Game Master application, these would be delivered from the Pedagogical Module to the Domain Module. Then the Domain Module would determine which activity (or tactic) to apply, where to apply it, and how to apply it. Now the automated pedagogical request is delivered to the OC/T instead of the having the Domain Module apply it. Once the request reaches the Game Master, the OC/T is included in the decision to apply the strategy themselves or allow GIFT to automatically apply it. If the request is not automatically applied, the OC/T can determine when to apply the strategy, whether a subset of the strategy should be applied, and if any feedback in the strategy should be changed before being sent to the learner(s). In addition, the OC/T can also choose when to apply any of the predefined strategies (i.e. authored ahead of time) from the strategy preset panel, as well as author new feedback to be used immediately or in the future. Any of these strategy requests, when applied by the OC/T, are delivered from the Dashboard to the Domain Module. This logic establishes where and how to apply the activity as if it were directed from the pedagogical module. It is worth noting the OC/T has the option to disable this strategy oversite, enabling GIFT to operate in a closed-loop capacity.

Game Master Interface

The Game Master shows the OC/T the current learner state and pedagogical strategies for a real time assessment as the lesson unfolds. To deliver unfettered access to this information, several architecture changes had to take place. First, the learner state message, normally sent from the Learner Module to just the Pedagogical Module, needed to also be sent to the Dashboard server to display on the Game Master UI. The pedagogical request message still goes through the Domain Module where it is paired with that strategy's list of activities. This provides more than just a strategy name to the OC/T, enabling a richer context of what the system is acting upon. These two changes provide a real time view into the learning effect chain.

Second, additional user interface enhancements were added to help identify important information about the training, including how long a task has been active, when the last assessment for a task/concept occurred, how many times a strategy has been applied and whether a strategy was applied manually by an OC/T. Third, various parts of the Game Master UI were highlighted in yellow, as seen in Figure 4, to attract the OC/T's attention to portions of the real time assessment and pedagogical requests which will not be triggered by GIFT automatically. In this example the author has defined several concepts that require an OC/T to provide an assessment and several pedagogical requests that can be applied when deemed appropriate.



Leave the FOB
Last Assessment: 2m 47s
Active: 7m 46s

Vehicles Reaches Gate
Last Assessment: 2m 47s

Team Mounts Vehicle
Last Assessment: 7m 47s

Last Vehicle Exit FOB Gate
Last Assessment: 2m 47s

Mount Up In Time
Last Assessment: 7m 36s

Travel to Weapons Cache
Last Assessment: 2m 36s
Active: 2m 47s

Maintain Situational Awareness
Last Assessment: 7m 47s

Maintain Proper Max Speed Vehicle 1
Last Assessment: 7m 47s

Vehicles Reach Weapons Cache
Last Assessment: 2m 47s

Maintain Proper Max Speed Vehicle 2
Last Assessment: 7m 47s

Maintain Proper Vehicle Spacing Vehicle 2
Last Assessment: 2m 36s

Reach First Compromised Arrow
Last Assessment: 7m 47s

Optional Demo Transport
Last Assessment: 7m 47s

Reach Second Compromised Arrow
Last Assessment: 7m 47s

Puppeteering Event Trigger
Last Assessment: 7m 47s

GIFT Instructions
Last Assessment: 7m 46s
Active: 0s

Init Instructions Timer
Last Assessment: 7m 46s

Secure Weapons Cache
Last Assessment: 7m 47s
Active: 0s

Entered Weapons Cache
Last Assessment: 7m 47s

Secure Outside Perimeter
Last Assessment: 7m 47s

Spot Report to Command
Last Assessment: 7m 47s

Maintain Weapons Cache Security
Last Assessment: 7m 47s

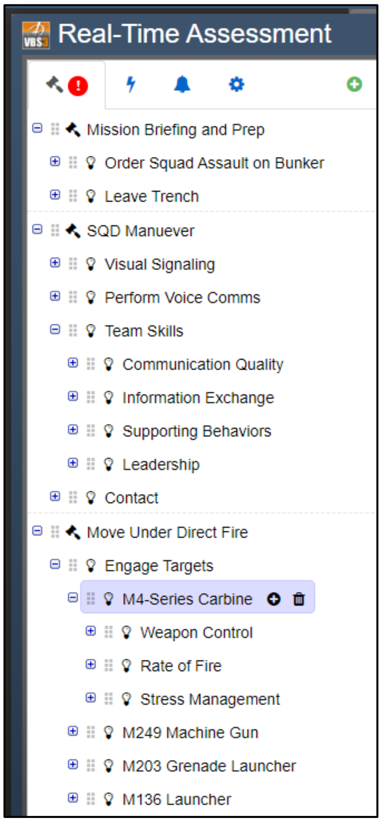
Sustainability Timer
Last Assessment: 7m 47s

Apply Open FOB Gate >_
Apply Close FOB Gate >_
Apply Transport Vehicles to Shorten Demo
Apply Remove GPS Arrows
□ Init Instructions Feedback
□ Vehicle Assignment Instructions
□ Message to White Cell - Open Gate
□ Message to White Cell - Transport Vehicles for Shorter Demo
□ Exited FOB - Follow GPS Instructions
□ Speeding Ticket Feedback Vehicle 1
□ Speeding Ticket Feedback Vehicle 2

Figure 4. Game Master UI with the DKF Task Model showing real-time performance on active tasks (Note: Icons highlighted yellow designate performance concepts that require human-observed assessments).

In the next phase of Game Master development we plan to focus on addressing user workload. One of the first redesigns we plan on making is the ability to support intermediate concepts. Currently the Game Master can only support one nested level of concepts under a task while authors of a GIFT real time assessment (DKF) can create an unrestricted amount of subconcepts as seen in Figure 5. Once intermediate concepts can be rendered, more robust assessment hierarchies can be observed in real time. Beyond merely showing all of the real time assessment information, the Game Master interface needs to be able to filter through the assessment information based on the needs of the observer or instructor. We envision views that are focused around the current standards being assessed, individual learners, teams and observer role assignments. When using a view centered around the current standards being assessed, any intermediate concepts (e.g. "Engage Targets"; see Figure 5) will be collapsed by default, leaving only tasks and leaf concepts being shown on screen. This standards view delivers an immediate list of the concepts being assessed per task which is important because those concepts are the ones being automatically graded or needing observed assessments.

The next two views, individual and team, are variants of each other, and can be used to quickly visualize the activities of specific individuals or groups. In an individual view type, each learner will be shown in a list. For each learner, the tasks/concepts that the learner, or the team the learner is a part of, is being assessed against will appear. This is the same approach used in the team view except each team will be shown in a list. Providing the appropriate, easy to use, filtering and searching capabilities will be crucial in large scale exercises. The final view is the observer role assignment type. In this view we believe observer roles will be something a real time assessment author can define. Each role can have a defined name and can be assigned to individuals, teams, tasks and/or concepts. When the observer assumes a particular assignment uses the Game Master, the corresponding authored role can be selected and the view of the ongoing real time assessment will be filtered so that only the appropriately tagged elements are visible. For example, using Figure 5 for reference, 'observer A' is assigned to assess the 'Move Under Direct Fire' task at Location Alpha while 'observer B' will assess at Location Bravo. To limit the observer's workload required to assess the situation and identify who/what/why/where in the Game Master interface, the view can easily be altered to bring the necessary components to the forefront.



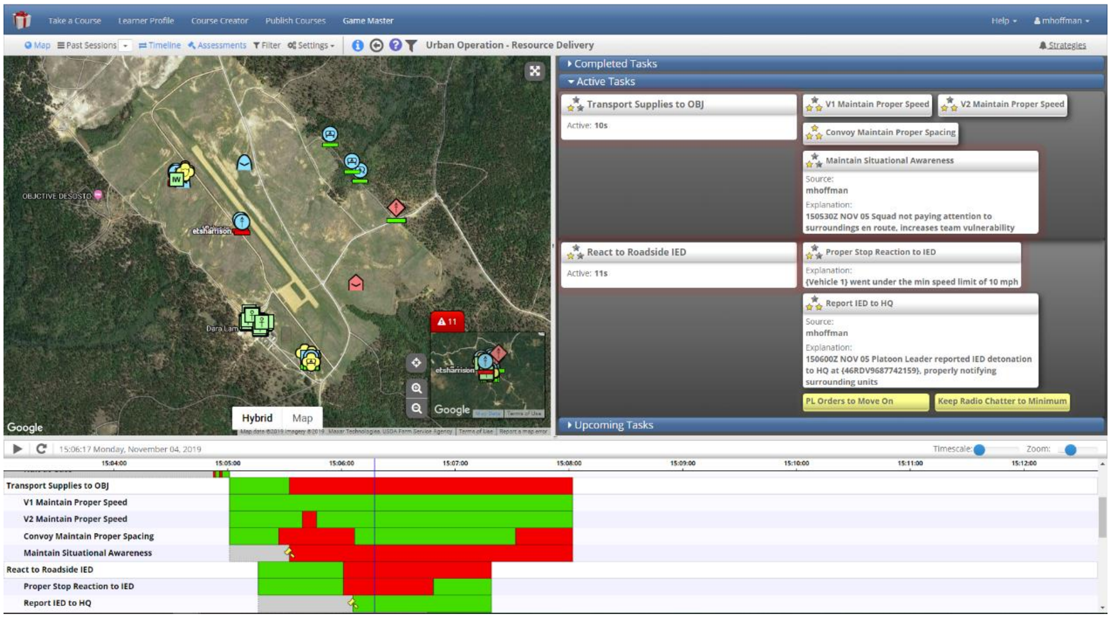
Real-Time Assessment

- Mission Briefing and Prep
  - Order Squad Assault on Bunker
  - Leave Trench
- SQD Manuever
  - Visual Signaling
  - Perform Voice Comms
  - Team Skills
    - Communication Quality
    - Information Exchange
    - Supporting Behaviors
    - Leadership
  - Contact
- Move Under Direct Fire
  - Engage Targets
    - M4-Series Carbine
      - Weapon Control
      - Rate of Fire
      - Stress Management
    - M249 Machine Gun
    - M203 Grenade Launcher
    - M136 Launcher

Figure 5. 'M4-Series Carbine' is an example of a nested concept that was authored and both the parent concept 'Engage Targets' and task 'Move Under Direct Fire' should be shown in Game Master.

Another workload issue we plan on addressing is the need to keep observers from looking away from the environment. We understand the need to maintain visual presence, understand situational awareness and document ongoing observations in a manner that is as natural as possible. Currently the Game Master interface provides a text area for each task/concept where observers can type in a comment or bookmark an important event. The content may be something that was witnessed and/or a reminder to the user to talk about a certain topic in an AAR. By requiring the user to look at a screen while providing text, any observations of the ongoing exercise will be missed. To alleviate this, we are planning to investigate several different mechanisms by which observers can trigger recording of their voice on demand. The recording could start when certain hardware button combinations are pressed like the Volume Up button three times on a tablet, or when saying a specific key phrase like "OK Game Master". In these approaches, the audio files might not be automatically associated with a specific task/concept, but rather a time stamped recording that can be used for preparing for or during an AAR.

Once the real time assessment has completed, Game Master has the ability to playback the entire session. The playback capability can be used as a supporting tool when conducting an AAR. In this interface, seen in Figure6, the user can control playback of the session events by playing, pausing, and looping on the timeline. Changes in assessment for tasks and concepts are seen in color coded bars. Whether the user sees the map, assessment, or timeline panel is customizable. Any bookmarks, text, or audio, are accessible through this interface. Currently the Game Master does not allow users to edit the session details such as changing an existing assessment, adding or over-riding an assessment, adding a bookmark or changing an existing bookmark. To facilitate this need, we plan on developing a patch style framework where users can make their changes and then save them. When loading the session for playback in the future, the user will have the ability to apply any patch files that are found and linked to that session. In this manner, multiple observers can each create their own patch files or build upon other patch files to ultimately build their own version of the session that satisfies the needs of their specific AAR session or audience.



Transport Supplies to OBJ

Active: 10s

V1 Maintain Proper Speed

V2 Maintain Proper Speed

Convoy Maintain Proper Spacing

Maintain Situational Awareness

Source:
mhoffman

Explanation:
150530Z NOV 05 Squad not paying attention to surroundings en route, increases team vulnerability

React to Roadside IED

Active: 11s

Proper Stop Reaction to IED

Explanation:
{Vehicle 1} went under the min speed limit of 10 mph

Report IED to HQ

Source:
mhoffman

Explanation:
150600Z NOV 05 Platoon Leader reported IED detonation to HQ at {46RDV9687742159}, properly notifying surrounding units

PL Orders to Move On

Keep Radio Chatter to Minimum

15:06:17 Monday, November 04, 2019

Transport Supplies to OBJ
V1 Maintain Proper Speed
V2 Maintain Proper Speed
Convoy Maintain Proper Spacing
Maintain Situational Awareness
React to Roadside IED
Proper Stop Reaction to IED
Report IED to HQ

Figure 6. Game Master past session playback user interface that allows playback of a capture session

Visualization Considerations and CONCLUSIONS

An important design consideration with the initial Game Master implementation in GIFT was supporting extensible visualization techniques to address perceptual bottlenecks, and to support evolving requirements. To this end, it is important to document visualization needs based on task dependencies and cognitive limitations of the human user. In Table 1 we present a list of data visualization formats for consideration, with the caveat that they must maintain domain- and environment-independency.

Table 1. Game Master visualization add-ons for consideration

Visualization Technique	Description/Role	Visualization Consideration
Attentional Cueing	Direct attention of OC/T to designated scenario event using established triggers and cueing modalities (e.g., visual, auditory, both).	• Real-time: easy-to-perceive color overlays to reduce time to interpret and act as OC/T adaptive training functions. • AAR: use attentional cueing to highlight task and environment elements to guide discussion and support learning and error-correction process.
Affect Monitoring	Visualize affective condition of individuals and teams (i.e., stress, fatigue, workload) based on available sensors and state classifiers.	• Real-time: OC/Ts can track affect to monitor stress and adapt training. Can use to trigger real-time affect management techniques. • AAR: Provide further diagnostics of performance. Visual overlay of physiological response with task execution (e.g., show rise in stress when casualty incurred) to support discussion on mitigation tactics.
Communication Flow	Visualize information exchange across network of team based scenario events of interest and designated triggers.	• Real-time: n/a • AAR: visualize flow of information around designated scenario events to judge teamwork and better diagnose collective performance (e.g., tracking comms from a specified event until directive issued by unit leader).
Persistent Neurodynamic States (see Stevens et al., 2020 in this book)	Visualize neurodynamic features for detection of collective uncertainty phases within task execution	• Real-time: OC/T tracking of neurodynamic correlations for diagnosing uncertainty for real-time coaching and adaptation. • AAR: Visualization tool to drive AAR discussion and reflection based on neurodynamic signatures tracked against critical scenario events.
Task Progress across Role and Team Objectives	Establish aggregate representation of progress for OC/T situational awareness	• Real-time: Provides up-to-second status on scenario progress as a function of tasks experienced and remaining storyboard triggers. • AAR: Identify situations where decisions could have been made differently to boost progress, maintain objectives and facilitate better cohesion across teams in the same battlespace.
Leader Board Visualizations	Define Measures of Performance and Measures of Effectiveness to serve as leadership board criteria	• Real-time: enable OC/T to track performance against established leaderboard for use as motivation tool. • AAR: Explore leaderboard visualization to see what performance deltas exist between current performance and best recorded.
Remedial Content for Identified Errors	Establish visual reference of performance outcomes with remedial materials to address recorded deficiencies.	• Real-time: n/a • AAR: Build ontological visualization to show interconnected concepts, competencies and available resources for development.

For each visualization technique referenced in Table 1, the manner in which data is processed and presented is critical. In addition, just because a visual technique is supported does not mean it improves the delivery and effectiveness of training. Multiple psychological factors must be considered before implementing a given visualization style. Everything from color, shape, opacity, UI placement, movement features, etc. must be accounted for, along with an understanding of competing information sources over limited cognitive channels. Visualizing information with auditory accompaniments can enhance the efficacy of that visualization by incorporating additional perceptual modes to guide attention.

In this chapter, we presented on-going work in the development of the GIFT Game Master UI to support an AIS collaboration model for team-focused simulation-based training. To support the human AIS counterpart, visualizing processes and components of AIS assessments and decision points was a required starting point. To baseline the capability, we started by integrating GIFT and Ares. With a visualization testbed integrated with a robust AIS, determining how best to visualize the interaction space is an open research question; especially considering the dynamic and high-workload environment the tool is being designed to support. Understanding the data produced and visualizations that can be generated is an important next step. Based on user needs, visualizations can be called up as services to meet specific interaction requirements, whether at real-time during training, or within an AAR focused discussion.

References

Domain Knowledge File Documentation. (2020). Retrieved on April 12, 2020 from https://gifttutoring.org/projects/gift/wiki/Domain_Knowledge_File_2020-1.

Foo, H.S. (2019). Training and Evaluation Outlines (T&EO): Usage and Scoring Method Preference for Task Steps and Sub-steps. Army Research Institute Research Note 2019-01. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/1075577.pdf.

Garneau, C.J., Boyce, M.W., Shorter, P.L., Vey, N.L., & Amburn, C.R. (2017). The augmented reality sandtable (ARES) research strategy, ARL Technical Note ARL-TN-0875. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD (2017)

Gervais, M. (2018). The Synthetic Training Environment Revolutionizes Sustainment Training, https://www.army.mil/article/210105/the_synthetic_ training_environment_revolutionizes_sustainment_training, accessed December 12, 2018.

Goldberg, B., Brawner, K. W., Holden, H., & Sottilare, R. (2012). Adaptive game-based tutoring: Mechanisms for real-time feedback and adaptation. Paper presented at the DHSS 2012, Vienna, Austria

Shute, V. J., Ventura, M., Small, M., & Goldberg, B. (2013). Modeling student competencies in video games using stealth assessment. In Sottilare, R., Hu, X., Graesser, A., & Holden, H. (Eds.), Design recommendations for adaptive intelligent tutoring systems: Learning modeling (pp. 141-152). Washington, DC: Army Research Laboratory.

Shute, V. J., & Ventura, M. (2013). Stealth assessment: Measuring and supporting learning in video games. Cambridge, Massachusetts: The MIT Press

Sottilare, R. A., Ragusa, C., Hoffman, M., & Goldberg, B. (2013). *Characterizing an Adaptive Tutoring Learning Effect Chain for Individual and Team Tutoring.* Paper presented at the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Stevens, R., Mullins, R., Hu, X., Zapata-Rivera, D. & Galloway, T. (2020). Visualizing the Momentary Neurodynamics of Team Uncertainty. In. Sinatra, A.M., Hu, X., Graesser, A., Goldberg, B., & Hampton, A. (eds.) Design Recommendations for Intelligent Tutoring Systems: Data Visualization. Natick, MA: Combat Capability Development Command – Soldier Center.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, *46*(4), 197-221.

Woolf, B. P. (2009). Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning. Burlington, MA: Morgan Kaufmann.

Biographies

Editors

Dr. Anne M. Sinatra is a Research Psychologist in the Learning in Intelligent Tutoring Environments (LITE) Lab within the U.S. Army DEVCOM Soldier Center Simulation and Training Technology Center. The focus of her research is in cognitive psychology, human factors psychology, and adaptive team tutoring. She has specific interest in how information relating to the self and about those that one is familiar with can aid in memory, recall, and tutoring. Her dissertation research evaluated the impact of using degraded speech and a familiar story on attention/recall in a dichotic listening task. Her post-doctoral work examined the self-reference effect and personalization in the context of computer-based tutoring. Her work has been published in journals including the Computers in Human Behavior, Journal of Artificial Intelligence in Education, and Interaction Studies. Her work has also been published in conference proceedings including the Human Factors and Ergonomics Society conference, and the Human Computer Interaction International conference. She additionally has served as an editor on four books (she was lead editor on two of them), and chaired two team tutoring workshops during the Artificial Intelligence in Education conferences in 2018 and 2019. Dr. Sinatra received her Ph.D. and M.A. in Applied Experimental and Human Factors Psychology, as well as her B.S. in Psychology from the University of Central Florida.

Dr. Arthur C. Graesser is a professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis and is a Senior Research Fellow in the Department of Education at the University of Oxford. He received his Ph.D. in psychology from the University of California at San Diego. Dr. Graesser's primary research interests are in cognitive science, discourse processing, and the learning sciences. More specific interests include knowledge representation, question asking and answering, tutoring, text comprehension, inference generation, conversation, reading, education, memory, emotions, computational linguistics, artificial intelligence, human-computer interaction, and learning technologies with animated conversational agents. Dr. Graesser served as editor of the journal Discourse Processes (1996–2005) and Journal of Educational Psychology (2009-2014) and as president of the Empirical Studies of Literature, Art, and Media (1989-1992), the Society for Text and Discourse (2007-2010), International Society for Artificial Intelligence in Education (2007-2009), and the FABBS Foundation (2012-13). He has published over 500 articles in journals, books, and conference proceedings. Dr. Graesser and his colleagues have designed, developed, and tested software that integrates psychological sciences with learning, language, and discourse technologies, including AutoTutor, AutoTutor-Lite, MetaTutor, GuruTutor, DeepTutor, HURA Advisor, SEEK Web Tutor, Operation ARIES!, iSTART, Writing-Pal, AutoCommunicator, Point & Query, Question Understanding Aid (QUAID), QUEST, & Coh-Metrix.

In 2010, Dr. Graesser received the Distinguished Scientific Contribution Award (Society for Text and Discourse) and in 2011 he received the Distinguished Contributions of Applications of Psychology to Education and Training Award (American Psychological Association). In 2012, Dr. Graesser received the first Presidential Award for Lifetime Achievement in Research from the University of Memphis. This award is the University's highest level of research recognition given to its faculty. It was established as part of the University's Centennial fundraising campaign in order to recognize the vital role and impact of research at the University of Memphis. He served as Chair of the Framework group in PISA Collaborative Problem Solving 2015. In 2018 he received the Harold W. McGraw, Jr. Prize in Education.

Dr. Xiangen Hu is a professor in the Department of Psychology, Department of Electrical and Computer Engineering and Computer Science Department at The University of Memphis (UofM) and senior researcher at the Institute for Intelligent Systems (IIS) at the UofM and is professor and Dean of the School of Psychology at Central China Normal University (CCNU). Dr. Hu received his MS in applied mathematics from Huazhong University of Science and Technology, MA in social sciences and Ph.D. in Cognitive Sciences from the University of California, Irvine. Dr. Hu is the Director of Advanced Distributed Learning (ADL) Partnership Laboratory at the UofM, and is a senior researcher in the Chinese Ministry of Education's Key Laboratory of Adolescent Cyberpsychology and Behavior.

Dr. Hu's primary research areas include Mathematical Psychology, Research Design and Statistics, and Cognitive Psychology. More specific research interests include General Processing Tree (GPT) models, categorical data analysis, knowledge representation, computerized tutoring, and advanced distributed learning. Dr. Hu has received funding for the above research from the US National Science Foundation (NSF), US Institute of Education Sciences (IES), ADL of the US Department of Defense (DoD), US Army Medical Research Acquisition Activity (USAMRAA), US Army Research Laboratories (ARL), US Office of Naval Research (ONR), UofM, and CCNU.

Dr. Benjamin Goldberg is a member of the Army Futures Command - Combat Capabilities Development Command Simulation and Training Technology Center in Orlando, FL. He has been conducting research in the Modeling & Simulation community for the past eight years with a focus on adaptive learning in simulation-based environments and how to leverage Artificial Intelligence tools and methods to create personalized learning experiences. Currently, he is the LITE Lab's lead scientist on instructional management research within adaptive training environments and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). Dr. Goldberg is a Ph.D. graduate from the University of Central Florida in the program of Modeling & Simulation. His work has been published across several well-known conferences, with recent contributions to the Human Factors and Ergonomics Society (HFES), Artificial Intelligence in Education and Intelligent Tutoring Systems (ITS) proceedings. Dr. Goldberg has also recently contributed to the journal Computers in Human Behavior and to the Journal of Cognitive Technology.

Dr. Andrew J. Hampton is a research scientist assistant professor at the Institute for Intelligent Systems, within the University of Memphis. His current duties include serving as a grant coordinator and project manager on the pioneering hybrid tutor ElectronixTutor. He is also one of the development leaders on a conversational AI meant to aid in career planning through education and qualification tracking, intelligent recommendation, and mitigation of personal issues. Andrew's research interests include technologically mediated communication, psycholinguistics, semiotics, adaptive educational technology, artificial intelligence, and political psychology. He also serves as the current Chairman for the IEEE Project 2247 working group which is working for the development of standards and recommended practices for AISs.

Authors

Dr. Keith Brawner

Keith Brawner, PhD is a senior researcher for the U. S. Army Combat Capability Development Command Soldier Center at the Simulation and Training Technology Center (CCDC-SC-STTC), and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). He has 13 years of experience within U.S.

Army and Navy acquisition, development, and research agencies. He holds a Masters and PhD degree in Computer Engineering with a focus on Intelligent Systems and Machine Learning from the University of Central Florida. His current efforts are on artificial intelligence for the Synthetic Training Environment Simulation and Network Compression. He manages research in adaptive training, artificial intelligence for training, and architectural programs towards next-generation training.

Lina Brihoum

At the age of 16 in 2015, Lina became an intern for the Department of Defense where she learned all about simulation, testing, game design, and development. While she was an intern, she also studied computer networks and systems and received certifications which enhanced her breath of knowledge in the tech industry. After high school, Lina joined and supported the Synaptic Sparks team as an intern to assist with developing and planning of contracts and other software projects. While an intern, she attended Florida Polytechnic University where she graduated top of her class with a Bachelors of Science in Computer Science in 2019. After graduation, Lina signed on full time to assist with software projects of all kinds as well as providing research and mobile development skills to the team.

Jody L. Cockroft

Jody L. Cockroft, AA, BS, CCRP is a Research Specialist at the University of Memphis (UoM) in the Psychology Department with the Institute for Intelligent Systems where she has been for the past five-plus years. She has over thirty years of experience in scientific research and has worked on both the bench and on various clinical studies. She has been an integral part of the Army Research Laboratory Cooperative agreement, the Advanced Distributed Learning Initiative ADL-A project, the Learner Data Institute (LDI), and the Advanced Learning Theories, Technologies and Impacts (ALTTAI) Consortium for the past several years while at the UoM. She has authored or co-authored over twenty articles in peer-reviewed journals as well as countless abstracts and posters. She is the treasurer of the Adaptive Instructional Systems (AIS) Working Group IEEE P2247.1. Her research interests include the standardization of adaptive instructional systems and improving human learning.

Dr. Cristina Conati

Dr. Conati is a Professor of Computer Science at the University of British Columbia, Vancouver, Canada. She received a M.Sc. and Ph.D. in Intelligent Systems at the University of Pittsburgh. Conati's research is at the intersection of Artificial Intelligence (AI), Human Computer Interaction (HCI) and Cognitive Science, with the goal to create intelligent interactive systems that can capture relevant user's properties (states, skills, needs) and personalize the interaction accordingly. Conati has over 120 peer-reviewed publications in this field and her research has received awards from a variety of venues, including UMUAI, the Journal of User Modeling and User Adapted Interaction (2002), the ACM International Conference on Intelligent User Interfaces (IUI 2007), the International Conference of User Modeling, Adaptation and Personalization (UMAP 2013, 2014), TiiS, ACM Transactions on Intelligent Interactive Systems (2014), and the International Conference on Intelligent Virtual Agents (IVA 2016).

Dr. Conati is an associate editor for UMUAI, ACM TiiS, IEEE Transactions on Affective Computing, and the Journal of Artificial Intelligence in Education. She served as President of AAAC, (Association for the Advancement of Affective Computing), as well as Program or Conference Chair for several international conferences including UMAP, ACM IUI, and AI in Education. She serves on the Executive Committees of AAAI (Association for the Advancement of Artificial Intelligence) and of CAIDA (the UBC Center for Artificial Intelligence, Decision Making and Action).

Darian DeFalco

Darian J. DeFalco is a Site Reliability Engineer for Transfix, Inc.

Dr. Jeanine A. DeFalco, is an Adaptive Training Research Scientist and Post-Doctoral Research Fellow with the Army Futures Command - Combat Capabilities Development Command, Simulation and Training Technology Center, Orlando, Florida, working out of the United States Military Academy at West Point, NY. She received her Ph.D. in Psychology from Columbia University, specializing in Human Development/Cognitive Studies in Education with a concentration in Intelligent Technologies. Jeanine's current research includes developing and testing pedagogical models for the Generalized Intelligent Framework for Tutoring to determine the relationship of creative and analogical reasoning in accelerated expert problem-solving in critical care medical education. Jeanine is currently working on a book on the intersection of psychology, ethics, and AI.

Dr. Jeremiah T. Folsom-Kovarik is a Senior Scientist at Soar Technology, Inc. Dr. Folsom-Kovarik earned his Ph.D. in computer science from the University of Central Florida in 2012, where he added a training overlay to a call-for-fire simulation in use by the U.S. Marine Corps which demonstrated planning under uncertainty, tractable machine learning for policies at real-world scale, and improved outcomes in multiple measures. Dr. Folsom-Kovarik has contributed to the science of intelligent systems in adaptive user interfaces, cognitive ergonomics, human-machine teaming, and other machine learning challenges. One theme of this research has been making machine learning effective in challenging settings which lack large training data or require control interactions with nontechnical personnel. After the advent of deep learning and big data approaches, cognitive and semantic knowledge provides a powerful approach to directing and focusing machine learning algorithms and creating useful and usable artificial intelligence under real-world data constraints.

Trysha Galloway

Trysha Galloway is Director of Cognitive Electrophysiology Research for The Learning Chameleon Inc. Her research interests blend the population based advantages of probabilistic performance modeling with the detection of neurophysiologic signals to help personalize the learning process in complex education and training activities. Applying these ideas to teams suggests the presence of underlying mechanisms of teamwork that have not been exploited for performance measurements, or for the development of training protocols. The development of these quantitative neurodynamic measures of teamwork, provides a first step for building understandable bridges between the theory and practice of social coordination dynamics, and improving team performance in settings as diverse as Submarine Piloting and Navigation by Navy teams, high school problem solving as well as healthcare teams.

Dr. Lane T. Harrison

Lane Harrison is an Assistant Professor in the Department of Computer Science at Worcester Polytechnic Institute. Prior to joining WPI, Lane was a postdoctoral fellow in the Department of Computer Science at Tufts University. He obtained his Bachelor's and PhD degrees in computer science from the University of North Carolina at Charlotte. In 2015, Lane served as general chair for the IEEE Visualization for Cyber Security (VizSec) Symposium, held in conjunction with the IEEE VIS Conference. Lane directs the VIEW group at WPI, where he and his students leverage computational methods to understand and shape how people interpret, use, and create data visualizations and visual analytics tools.

Zachary Heylmun

Zach Heylmun graduated from the University of Florida with a degree in Digital Arts and Science engineering and supports SSI on the GIFT program as an as-needed consultant for the past two years. After graduation, he worked for Lockheed Martin on low-level, high performance graphics as well as virtual reality rendering for flight simulation and training. Since starting his own company, Voidstar Solutions, as well as helping to form Synaptic Sparks, a 501c3 charity dedicated to STEM education, he has worked with a wider variety of technologies. Through a combination of efforts, both for- and non-profit, he has become an expert on web technologies, mobile applications, sensor fusion, and server infrastructures.

Dr. Cheryl Johnson

Dr. Cheryl Johnson is a Senior Research Psychologist at the Naval Air Warfare Center Training Systems Division in Orlando, FL, performing research on emerging training technology and adaptive training systems. She earned her M.A. and Ph.D. in Cognition, Perception, and Cognitive Neuroscience from the University of California Santa Barbara. Dr. Johnson has over 14 years of experience performing technology-based training research and her work has been published in the Journal of Experimental Psychology: Applied, Journal of Educational Psychology, and Computers in Human Behavior. Her research interests include adaptive training, instructional strategies, virtual reality training applications, and multimedia learning.

Dr. Joan Johnston

Dr. Joan Johnston is a Senior Scientist and Research Psychologist with the U.S. Army Combat Capabilities Development Command, Soldier Center, Simulation, Training and Technology Center where she leads research projects on training effectiveness, team training, and training technologies. She began her military research career in 1990 with the U.S. Navy, and for her efforts she was awarded the Office of Naval Research Dr. Arthur E. Bisson Prize for Naval Technology Achievement in 2000, and the Society for Industrial and Organizational Psychology M. Scott Myers Award for Applied Research in the Workplace in 2001. At about that time she was nominated and made a NAVAIR Fellow. In 2012, she became the Orlando Unit Chief of the Army Research Institute, and then joined the Army Research Laboratory, Human Research Engineering Directorate in 2014 as a senior scientist. That same year she was awarded the US Army Civilian Service Achievement Medal for an innovative team training strategy to improve decision making under stress in dismounted Army squads. Dr. Johnston received her M.A. and Ph.D. in Industrial and Organizational Psychology from the University of South Florida.

Mike Kalaf

Mike has over 30 years of Modeling, Simulation and Training leading large scale efforts leveraging cutting edge technology. Mike has worked in the commercial and military aviation, training and simulation business. In his most recent efforts, he has been leading new opportunities applying front end modeling, simulation and analysis. Mike has led several programs integrating “state of the art” technology and delivering highly successful technology and business innovation. Mike has been collaborating with educational organizations and exploring conceptual frameworks, platforms and business models to transform our current system and elevate the performance and quality. He is involved with the University of Central Florida’s College of Education on a unique system of teacher training via classroom simulators. These projects fit well to advance science, technology, engineering and mathematics learning to lay the groundwork for a new generation of engineers and scientists. Mike volunteers his time to numerous education organizations including serving as a board member for the Central Florida STEM council and the Seminole County Public Schools Foundation. Mike’s formal education includes an earned Mechanical Engineering degree from Rochester Institute of Technology, RIT.

Dr. Judy Kay

Judy Kay is Professor of Computer Science. She leads the Human Centred Technology Research Cluster, in the Faculty of Engineering and IT at the University of Sydney. A core focus of her research has been to create infrastructures and interfaces for personalization, especially to support people in lifelong, life-wide learning. This ranges from formal education settings to supporting people in using their long-term ubicomp data to support self-monitoring, reflection and planning. Central to this has been in the design of the Personis user modelling systems and interfaces that enable people to control their own long-term personal information from diverse sensors on devices be they worn, carried, embedded in the environment or conventional desktops. She has integrated this into new forms of interaction including virtual reality, surface computing, wearables and ambient displays. Her research has been commercialized and deployed and she has extensive publications in leading venues for research in user modelling, AIED, human computer interaction and ubicomp. She has had leadership roles in top conferences in these areas and is Editor-in-Chief of the IJAIED, International Journal of Artificial Intelligence in Education (IJAIED) and Editor of IMWUT, Interactive Mobile Wearable and Ubiquitous Technology (IMWUT).

Christopher H. Meyer

Christopher brings a breadth of leadership experience, technical knowledge, and project management to the team. And, most recently, Christopher has supported the GIFT program for two years under the most current contract. He received his Bachelor and Master of Science degrees in Computer Science from Kansas State University, also receiving minors in Economics and Modern Languages, and studied abroad for a year during a tour in Japan at Chukyo University dedicated to the specialized study of Artificial Intelligence. After completing traditional education phases, Chris was employed at Lockheed Martin for 10 years working hand-in-hand with representatives from the Departments of Defense, Health and Human Services, Energy, and Education to assist in the creation of solutions to solve challenges at a national level. Having now co-created his own business segment, Chris enjoys utilizing entrepreneurship, international experience, leadership knowledge, and his own engineering skills alongside his peers to advance world technology, health, and opportunity efficiently and responsibly.

Dr. Morten Misfeldt

Morten Misfeldt is Professor and leader of the Center for Digital Education, Department of Science Education and Department of Computer Science, University of Copenhagen. His research is placed on the intersection between digitalization of teaching and learning and mathematics education. He has published on game-based learning, learning analytics, the role of digital tools in mathematics education and on the influence of tools and representations on mathematical thinking. Recently Misfeldt has worked with the digitalization of infrastructure for teaching, as well as with the interaction between technology education and mathematics education.

Ryan Mullins

Ryan Mullins is the Deputy Director of the Intelligent Analytic Technologies Division, Lead for Interactive Intelligent Systems, and a Senior Research Engineer at Aptima, Inc. As a *Senior Research Engineer*, he leads and contributes to the development of human-AI systems, focusing on the design, implementation, and evaluation of user experiences, human-machine interfaces, and analytics. Mr. Mullins has experience leveraging rapid prototyping methods, such as Design Sprinting, and Agile software development techniques to support a range of customers across the mission planning, cybersecurity, and information analysis domains. As the *Lead for Interactive Intelligent Systems*, he manages a portfolio of projects across the information analysis, command and control, and cyber security domains, and works with stakeholders from

Aptima's Business Engagement Strategy Team and Aptima Ventures to identify core technologies that enable the development of commercial products. As a *Deputy Director*, he oversees the internal operations of a Division of 30 data scientists, software engineers, and artificial intelligence researchers. Mr. Mullins holds an M.S. in Geography and a B.S. in Computer Science from the Pennsylvania State University. He is a member of the North American Cartographic Information Society, the Association for Computing Machinery, and the United States Geospatial Intelligence Foundation.

Dr. Scott Ososky

Dr. Ososky is a User Experience (UX) Researcher at Microsoft. His current work is focused on professional development experiences for teachers, as well as information systems requirements in education. He was also a member of the Xbox research team at Microsoft, working on entertainment platforms and services. His previous UX work with the Army Research Laboratory (ARL) sought to improve the learnability and usability of GIFT's authoring system and interfaces. Scott's original work regarding mental models of human interaction with intelligent robotic teammates has been published in the proceedings of the Human Factors and Ergonomics Society, HCI International, and SPIE Defense & Security annual meetings. Dr. Ososky received his Ph.D. and M.S. in Modeling & Simulation, as well as a B.S. in Management Information Systems, from the University of Central Florida.

Dr. Vasile Rus

Dr. Vasile Rus is a Full Professor of Computer Science at The University of Memphis and the Lead Principal Investigator of the newly NSF-funded Learner Data Institute to lay the foundations of a Data Science Institute for learner data (www.learnerdatainstitute.org). Dr. Rus' research interests lie at the intersection of artificial intelligence, human and machine learning, and natural language processing with an emphasis on developing interactive intelligent systems such as intelligent tutoring systems and care-bots (healthcare bots). Dr. Rus has served in various roles on research projects funded by the National Science Foundation, Department of Defense, Department of Education, and private companies. Many of those projects involved the development of intelligent tutoring systems and medium-size (10-25 people) interdisciplinary teams. For instance, Dr. Rus has led the development of the DeepTutor system (www.deeptutor.org), a project funded by the Department of Education and is currently leading the development of an NSF-funded project to develop an intelligent tutoring system for source code comprehension, called DeepCode. Dr. Rus produced more than 150 peer-reviewed publications and received 6 Best Paper Award nominations of which 3 were Best Paper Awards. His team won the first two Question Answering competitions organized by the National Institute for Science and Technology (NIST) and recently his team won the English Semantic Similarity challenge organized by the leading forum on semantic evaluations – SemEval. Among other accomplishments, Dr. Rus was named Systems Testing Research Fellow of the FedEx Institute of Technology for his pioneering work in the area of software systems testing and is a member of the PI Millionaire club at The University of Memphis for his successful efforts to attract multi-million funds from federal agencies as Principal Investigator (PI).

Dr. David Williamson Shaffer

David Williamson Shaffer is the Vilas Distinguished Professor of Learning Sciences at the University of Wisconsin-Madison in the Department of Educational Psychology and a Data Philosopher at the Wisconsin Center for Education Research. His M.S. and Ph.D. are from the Media Laboratory at the Massachusetts Institute of Technology, and before coming to the University of Wisconsin, he was a teacher, teacher-trainer, curriculum developer, and game designer. Professor Shaffer's current work focuses on merging statistical and qualitative methods to model complex and collaborative thinking skills. He has authored more than 250 publications with over 100 co-authors, including *How Computer Games Help Children Learn* and *Quantitative Ethnography*.

Zachari Swiecki

Zachari Swiecki is a Ph.D. candidate in the learning sciences area of the Department of Educational Psychology at the University of Wisconsin, Madison and a member of the Epistemic Analytics lab. He graduated summa cum laude from the University of Alabama, Tuscaloosa with a B.S. degree in mathematics and physics. In 2016, he received a M.S. degree in Educational Psychology from the University of Wisconsin, Madison. His work focuses on collaboration analytics: modeling collaboration (collaborative learning, collaborative problem solving, and collaborative contexts more generally) as the integration of social and cognitive processes and exploring the implications of such models for teaching, learning, and research.

Dr. Robert A. Sottilare

Dr. Robert A. Sottilare is the Science Director for Intelligent Training at Soar Technology, Inc. He came to SoarTech in 2018 after completing a 35-year federal career in both Army and Navy training science and technology organizations. At the US Army Research Laboratory, he led the adaptive training science and technology program where the focus of his research was automated authoring, instructional management, and analysis tools and methods for intelligent tutoring systems (ITSs) and standards for adaptive instructional systems. He is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT), an open source, AI-based adaptive instructional architecture. GIFT has over 2000 users in 76 countries. Dr. Sottilare has long history as a leader, speaker, and supporter of learning and training sciences forums at the Defense & Homeland Security Simulation, HCII Augmented Cognition, and AI in Education conferences. He is the founding chair of the HCII Adaptive Instructional Systems (AIS) Conference. He is a member of the AI in Education Society, the Florida AI Research Society, the IEEE Computer Society and Standards Association, the National Defense Industry Association (lifetime member), and the National Training Systems Association. He is currently the IEEE Project 2247 working group chair for the development of standards and recommended practices for AISs. He is a faculty scholar and adjunct professor at the University of Central Florida where he teaches a graduate level course in ITS design. Dr. Sottilare has also been a frequent lecturer at the United States Military Academy (USMA) where he taught a senior level colloquium on adaptive training and ITS design. He has a long history of participation in international scientific fora including NATO and the Technical Cooperation Program. He has over 200 technical publications in the learning sciences field with over 1500 citations in the last 5 years. His doctorate is in Modeling & Simulation with a focus in Intelligent Systems from the University of Central Florida. Dr. Sottilare is a recipient of the US Army Meritorious Service Award (2018; 2nd highest civilian award), the US Army Achievement Medal for Civilian Service (2008; 5th highest civilian award), and two lifetime achievement awards in Modeling & Simulation: US Army RDECOM (2012; inaugural recipient) and National Training & Simulation Association (2015). He was also recognized by NTSA in 2019 for his contributions to adaptive instruction and the design and development of GIFT.

Dr. Ron Stevens

Dr. Stevens received his Ph.D. In Microbiology and Molecular Genetics from Harvard University and is Professor (Emeritus) at the UCLA School of Medicine and a member of the UCLA Brain Research Institute. In his position as the CEO of The Learning Chameleon, Inc., his recent research has focused on using EEG-derived measures to model team neurodynamics in the complex and real-world settings of military and healthcare training, including live patient operations. His group first demonstrated the presence of temporally extended (10s to minutes long) neurodynamic organizations in teams, and then extended these studies to show the contributions that the dynamics of each team member makes to the overall dynamics of the team.

Lucy Woodman

Lucy has recently graduated from the University of Central Florida with a Bachelor of Science in Information Technology. Lucy has supported Synaptic Sparks for one year during a successful internship and transitioned to be a major supporter of big data services within SSI since December of 2018. Lucy is a certified Amazon Web Service specialist, and obtained further AWS certifications in System Architecting in January of 2020. Lucy also supports the team by providing research and development support in new fields of network and social technology.

Dr. Diego Zapata-Rivera

Dr. Diego Zapata-Rivera is a Managing Principal Research Scientist in the Cognitive and Technology Sciences Center at Educational Testing Service in Princeton, NJ. His research at ETS has focused on the areas of innovations in score reporting and technology-enhanced assessment including work on adaptive learning and assessment environments, and game-based assessments. His research interests also include Bayesian student modeling, open student models, conversation-based tasks, caring assessment, virtual communities, authoring tools and program evaluation. Dr. Zapata-Rivera has produced over 100 publications including journal articles, book chapters, and technical papers. He has served as a reviewer for several international conferences and journals. He has been a committee member and organizer of international conferences and workshops in his research areas. He is a member of the Editorial Board of User Modeling and User-Adapted Interaction and an Associate Editor of the IEEE Transactions on Learning Technologies Journal. Most recently, Dr. Zapata-Rivera has been invited to contribute his expertise to projects sponsored by the National Research Council, the National Science Foundation, NASA and the US Army Research Laboratory.

Liang Zhang

Liang Zhang is a Ph.D. Candidate of Computer Engineering at the IIS (Institute for Intelligent Systems) at The University of Memphis. His research interests include educational data mining and analytics, learning analytics and programmable matter. He also works as a research assistant following Phil Pavlik in the Optimal Learning Lab.

Design Recommendations for Intelligent Tutoring Systems

Volume 8
Data Visualization

Design Recommendations for Intelligent Tutoring Systems (ITSs) explores the impact of intelligent tutoring system design on education and training. Specifically, this volume examines "Data Visualization". The Design Recommendations book series examines tools and methods to reduce the time and skill required to develop Intelligent Tutoring Systems with the goal of improving the Generalized Intelligent Framework for Tutoring (GIFT). GIFT is a modular, service-oriented architecture developed to capture simplified authoring techniques, promote reuse and standardization of ITSs along with automated instructional techniques and effectiveness evaluation capabilities for adaptive tutoring tools and methods.

About the Editors:

- ***Dr. Anne M. Sinatra*** *is a research psychologist at U.S. Army Combat Capabilities Development Command – Soldier Center – Simulation and Training Technology Center and works on the Generalized Intelligent Framework for Tutoring (GIFT).*
- ***Dr. Arthur C. Graesser*** *is a professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis and is a Senior Research Fellow in the Department of Education at the University of Oxford.*
- ***Dr. Xiangen Hu*** *is a professor in the Department of Psychology at The University of Memphis and visiting professor at Central China Normal University.*
- ***Dr. Benjamin Goldberg*** *is a senior researcher at U.S. Army Combat Capabilities Development Command – Soldier Center – Simulation and Training Technology Center and is a co-creator of GIFT.*
- ***Dr. Andrew J. Hampton*** *is a research assistant professor in the Department of Psychology at the University of Memphis.*

A Volume in the Adaptive Tutoring Series

9 780997 725780