DATA CHALLENGES OF LEVERAGING A SIMULATION TO ASSESS LEARNING

David Gibson
Curtin University
Perth, WA Australia

Peter Jakl
Pragmatic Solutions
West Lake Village, CA USA

ABSTRACT

Among the unique affordances of digital simulations are changes in the possibilities for targets as well as the methods of assessment, most significantly, toward integration of thinking with action, embedding of tasks-as-performance of knowledge-in-action, and unobtrusive observational methods. This paper raises and briefly defines key data challenges of assessing learning in a complex domain of performance within a digital simulation, which at the atomistic level include time and event segmentation, cyclic dynamics, multicausality, intersectionality, and nonlinearity. At the summary level, the key challenge is model building. An example of a simulation designed to develop teachers - simSchool – is integrated with an adaptive content delivery and analytics database – Leverage – which grounds the discussion.

1. INTRODUCTION

Assessing learning in a simulation requires a formalization of familiar everyday reasoning that assumes if someone is observed saying or doing something, the observation can be used to infer what they know and know how to do (Pellegrino, Chudowsky, & Glaser, 2001). The inferences require new methods of analysis because the performance space of simulations is considerably more complex than a traditional psychological test or measurement (Aldrich, 2004; Behrens, Frezzo, Mislevy, Kroopnick, & Wise, 2008; Rupp, Gushta, Mislevy, & Shaffer, 2010). For example, users acting within a digital environment contend with elements such as the interface tools, and purposes that are inherent in both the design of the environment and emergent in its interaction with users. Players also contend with anchored and evolving interaction rules, other players, and their private mental models of the environment as they traverse the available landscape of possibilities of thought and action. Learning what those users know and can do based on their actions as well as the artifacts they create in the digital environment involves three phases of dynamic assessment (Quellmalz et al., 2012): gathering data, applying criteria to make inferences and claims, and undertaking adaptive interactions with the user such as reporting results, offering new digital experiences, or ending the interaction.

This paper briefly defines key data challenges we have been facing and addressing in order to assess the complex higher order skill of classroom teaching based on evidence provided by simSchool as captured, analyzed and reported by Leverage. simSchool is a digital flight simulator for teachers, which provides a performance and assessment platform for the development of teaching skills. Leverage is a user analytics application for data mining interactions in digital environments; these applications work together to form a digital media-learning environment with embedded assessments of higher order knowledge and skills.

2. LEVERAGING A SIMULATION

Across a variety of settings, artificial intelligence and user analytics engines in simulations have been found useful for representing and developing higher order skills such as leadership, responsibility and time management; skills that are displayed when users interact with, influence others and make decisions. These settings include human resource departments, medical training programs, professional development of counsellors, military leadership training, and teacher education programs - anywhere that experts are interested in developing new supervisors, team leaders, and teachers (Aldrich, 2004; Gibson, Aldrich, & Prensky, 2007; Prensky, 2001). In simSchool (www.simschool.org) a user plays the role of a teacher while computer resources play the role of students. An artificial intelligence engine handles user interactions guided by models of teaching and learning; and a user analytics engine, Leverage (www.pr-sol.com), handles the delivery of digital media, the administration of groups of users, and the assessment of learning, including capturing, analyzing and reporting on data.

The data produced during a session has a dual role: it is used to influence the education of, as well as to make inferences about what the user knows and can do *as a teacher*, that is, within the *epistemic frame* (D. Shaffer, 2007) of the profession. The dual role and epistemic positioning illustrates how a simulation can simultaneously fulfil roles in the assessment FOR, OF and AS learning (Bennett, 2010). As a player makes choices in simSchool, a digital trail collected by Leverage provides evidence of the player's teaching expertise that is revealed in how and to what extent the simstudents learn as well as how the user manages available resources including instructional moves and student communications during the simulation.

2.1 Simulations as Complex Systems

The details of how simSchool works - how the simulated students respond to tasks and teacher talk - have been detailed elsewhere (Christensen, Tyler-Wood, Knezek, & Gibson, 2011; Zibit & Gibson, 2005; Zibit, Gibson, & Halverson, 2006) and is only briefly outlined here in order to focus on the data challenges of embedded and automated assessment. In brief, simSchool uses a dynamic modeling approach in which the user is a teacher who is an independent actor that chooses tasks and talking interactions, which in turn act as attractors for the simstudents. The artificial intelligence driving each simulated student is a hill-climbing algorithm; each student will attempt to reach equilibrium by attaining the goals of a given task if the task and setting do not impose too many barriers and the system is not perturbed by any other user actions. The time it takes simstudents to reach equilibrium with a task is determined by how their personality variables (physical, emotional and cognitive variables) interact with the requirements of the tasks and the teacher's talking choices.

The simSchool game mechanic ensures that the difference between any starting condition and any current or ending condition of the game is a result of the decisions made by the player. If a simstudent has learned or failed to learn, it is directly traceable to the user's decisions. While it might seem to oversimplify complex teaching practices, this arrangement actually allows a wide variety of performances by simSchool users, with potential for a number of inferences that can be made based on the digital record as well as by pre and post assessments and concurrent observations of the users.

2.2 Leveraging Data

Leverage software by Pragmatic Solutions, fully integrates with simulations such as simSchool and provides a scalable server application and database backend that is an infrastructure for scaling the application, organizing all users and groups, and creating extensive learner analytics. Server functions that summarize data are based on *atomistic* and *summary events*, which consist of a value associated with *action(s)* or *states* and their time-based contexts. Events identify a discrete user or application activity and exist at a variety of scales (atomic to summary levels) of data creation and collection, recognition of which feeds the assessment process as well as the adaptive digital media learning experience of the simulation. In Figure 1, we show an example of the time sensitive evolution of a cluster of variables in a simSchool teaching situation.

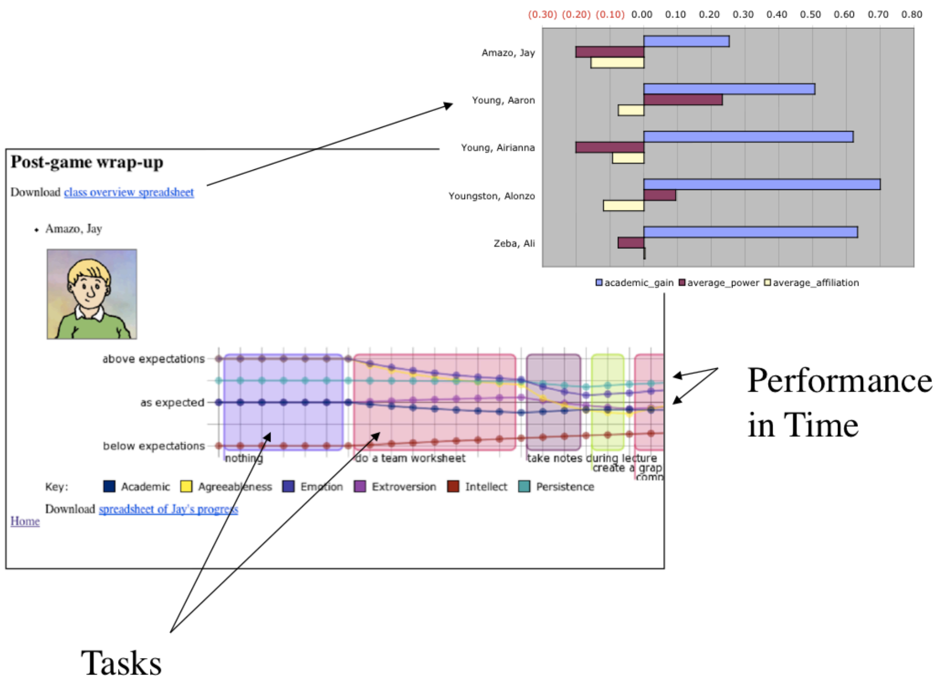

Figure 1. Representations of time-based events and a graph from an overview spreadsheet summary in simSchool display how variables in the simulation change over time in relationship to user actions. Changes in tasks made by the user are represented by colored blocks, time flows from left to right. Changes in simstudent internal states caused by the tasks are represented as points of data forming lines that change over time, displaying whether and to what extent the simstudent adapts and learns.

3. ATOMISTIC DATA CHALLENGES

Data challenges at the atomistic level are numerous, including time and event segmentation, cyclic dynamics, multicausality, intersectionality, and nonlinearity, which we will discuss in this section. Advanced multivariate methods involving serial and canonical correlations for multiple variables combined with automated network analysis methods provide some solutions and methods for these data challenges. We admit here that we have much yet to learn to integrate and place these methods into an automated computational framework with both inductive and deductive capabilities so that near-real time feedback can be provided to users of digital media learning environments and useful summaries can be created analyzing what users know and are able to do.

3.1 Time and event segmentation

The time segmentation problem is illustrated by the fact that since simSchool data captures performance of knowledge-in-action, how should we represent and analyze knowledge *in vivo*? That is, we'd like to be able to say things about what the user knows and can do without killing or masking critical performance information that evolves over time. One possibility we've explored is to provide time-based representations for human and machine analysis and to develop time-sensitive automated analytic methods for making inferences.

In Figure 1, approximately 20 data captures are taken within a span of about 3 minutes. Four user choices of task are evident, as are the impacts of each of those decisions on the learning of one simulated student, represented by changes in 6 dimensions as the agent adapts to the different task requirements. If we conduct a summative assessment at the end of the first task "do a team worksheet," we'd have to conclude that the user's choice of task caused important variables to decline in the agent, most importantly, academic performance. But if we wait until the end of the next task "take notes" the agent begins to recover academically. This illustrates one of the most basic problems of time-sensitive data analysis: *since systems evolve over time, how much time do we need in an analysis and when is the best time to stop gathering and start making sense of the situation?* Is there a form of *continuous interpretation* that should be employed, and should it use all the data or a particular window of time? This first problem includes determining what some have called slices, episodes or segments (Choi, Rupp, Gushta, & Sweet, 2010; Rupp et al., 2010; D. W. Shaffer et al., 2009) and it is not clear yet how to make differently sized slices commensurate with each other when the timing aspects are critical to the analysis.

3.2 Cyclic dynamics

A second problem closely related to time, is the *partially closed loop* or *cyclic dynamics problem of causality*. Complex systems, such as a user making choices and interacting with a digital artifact that in turn responds to those choices over time, have loops or cycles of causes (e.g. chicken and egg dilemmas). By "partially closed loop", we mean that one thing is both a cause and an effect of something else, which might in turn be both the cause and effect of the first thing, in a loop of ongoing relationships. Stopping the cycle is an arbitrary moment in the co-evolving causal network (e.g. imagine measuring the state of a room, cooling engine and thermostat on a day that begins cool, warms up and cools down again). In addition to the intrinsic self-reinforcing loops, external drivers may also be varying, raising the possibility that other things can also be both causes and effects connected to the loop. *The practice of assuming that a cause precedes an effect – a mainstay of linear causality - might be unwarranted*. If we take a snapshot at any point in time (e.g. any summative assessment) then we catch the system at some point in its loop, but we don't get the full loop into that one picture. For cyclic processes, what is the minimum number of measures per cycle and the characteristics of those measures that will produce a particular level of accuracy in the representation and analysis? We need methods that include effects as causes as the complexity of interactions evolve and that represent the differential phases of a loop of relationships when the loop is acting as a cyclical causal factor in a dynamic situation.

An example of data from a simSchool simulation illustrates these and other challenges in building user analytics and interpreting the results of actions in a simulation (Table1) during approximately 5 minutes of a user's performance. The first ten columns represent states of a single agent (a simulated student), with 5 variables for five psychological components (known as OCEAN), 1 variable for academic position, 2 computed variables that aggregate a running average from OCEAN (Power = E+C+O and Affiliation = A+N), 1 variable that relates power and affiliation to an attitudinal position on the Interpersonal Circumplex (Hofstee, de Raad, & Goldberg, 1992; Plutchik & Conte, 1997), 1 variable holding the pose or body position of the student at the classroom desk, and 3 variables that hold the intention of the user when talking to the student, in terms of the content of the talk (e.g. is the comment about student behavior or academic performance), the type of statement (e.g. is the comment an assertion, observation or question) and the attitudinal stance (e.g. one of 16 positions on the Interpersonal Circumplex see (Zibit & Gibson, 2005).

Table 1. Data from five minutes of a simSchool simulation. Trajectories for each field are estimated with data captured every ten seconds; 32 records were captured in 5.33 minutes of real time.

row	E	A	C	N	O	Aca	Power	Affli	C	P	Re	Ty	Att
1	-0.506	-0.197	-0.451	-0.451	0.910	-0.036	-0.016	-0.324	4	E			
2	-0.509	-0.269	-0.428	-0.428	0.873	-0.054	-0.021	-0.348	4	E			
3	-0.511	-0.329	-0.405	-0.405	0.840	-0.071	-0.025	-0.367	4	E			
4	-0.514	-0.378	-0.382	-0.382	0.811	-0.088	-0.028	-0.380	4	E			
5	-0.517	-0.418	-0.360	-0.360	0.786	-0.105	-0.030	-0.389	4	E			
6	-0.519	-0.450	-0.339	-0.339	0.763	-0.120	-0.031	-0.395	4	E			
7	-0.522	-0.477	-0.318	-0.318	0.744	-0.135	-0.032	-0.398	4	E			
8	-0.524	-0.499	-0.297	-0.297	0.727	-0.149	-0.031	-0.398	4	E			
9	-0.526	-0.518	-0.277	-0.277	0.711	-0.162	-0.031	-0.397	4	E			
10	-0.528	-0.532	-0.257	-0.257	0.698	-0.174	-0.029	-0.395	4	E			
11	-0.531	-0.545	-0.238	-0.238	0.686	-0.185	-0.028	-0.391	4	E			
12	-0.381	0.195	-0.067	0.531	0.828	-0.196	-0.127	-0.363	5	E	b	in	3
13	-0.387	0.052	-0.052	0.532	0.801	-0.202	-0.120	-0.292	5	E			
14	-0.394	-0.065	-0.038	0.534	0.777	-0.205	-0.115	-0.234	5	E			
15	-0.400	-0.161	-0.023	0.535	0.755	-0.208	-0.111	-0.187	5	E			
16	-0.406	-0.240	-0.009	0.537	0.737	-0.212	-0.107	-0.148	6	E			
17	0.210	-0.305	0.626	0.538	1.342	-0.218	-0.726	-0.116	8	E	b	ob	0
18	0.185	-0.358	0.625	0.539	1.253	-0.238	-0.688	-0.091	8	E			
19	0.162	-0.402	0.624	0.541	1.174	-0.257	-0.654	-0.070	8	E			
20	0.139	-0.437	0.624	0.542	1.105	-0.275	-0.623	-0.052	8	E			
21	0.117	-0.467	0.623	0.543	1.045	-0.292	-0.595	-0.038	8	E			
22	0.095	-0.491	0.623	0.545	0.991	-0.308	-0.570	-0.027	8	E			
23	0.075	-0.510	0.622	0.546	0.944	-0.323	-0.547	-0.018	8	E			
24	0.054	-0.526	0.622	0.547	0.903	-0.337	-0.526	-0.010	8	E			
25	0.207	-1.290	0.793	-0.202	1.039	-0.350	-0.680	0.746	2	D	b	as	13
26	0.182	-1.166	0.789	-0.184	0.986	-0.378	-0.653	0.675	2	D			
27	0.159	-1.064	0.785	-0.166	0.940	-0.400	-0.628	0.615	2	D			
28	0.136	-0.980	0.781	-0.149	0.899	-0.418	-0.605	0.565	2	D			
29	0.114	-0.912	0.777	-0.132	0.863	-0.433	-0.585	0.522	2	D			
30	0.093	-0.856	0.773	-0.116	0.832	-0.445	-0.566	0.486	2	D			
31	-0.128	-0.241	0.569	0.469	0.604	-0.454	-0.348	-0.114	7	E	a	as	5
32	-0.142	-0.306	0.569	0.472	0.603	-0.453	-0.344	-0.083	7	E			

Note: Field labels = row, extroversion, agreeableness, conscientiousness, neuroticism, openness, academic, power, affiliation, circumplex, talk about, talk type, talk attitude

For simplicity, an additional persistent set of variables that hold the constant targets of the task chosen by the user is not shown in Table 1. Those task data influence all rows in this series as an attractor, the goals toward which the OCEAN and academic variables are striving. The actual values of most variables is in floating point precision, but has been formatted to 3 places for convenience. A simSchool session might have from 1 to 20 agents each on their own multidimensional trajectory. Updating each row is a set of simultaneous equations that integrate the task goals, the previous states and interruptions such as the teacher talking during the task (e.g. rows 12, 17, 25, 31). The equations can be independent of each other during each time slice, but in fact are deterministically related to each other from record to record, so there is an inherent circularity that is only broken by the fact that time moves forward creating the next whole-system state. The user is the independent variable in an ongoing experiment during each simulation, determining which task the user chose and why, and what the user intended by talking to a particular student during the task. These parts of the data are crucial to an analysis of the user, particularly how to think about the meaning of those choices and the resulting data that is generated in order to give feedback to the user both during the simulation and afterward.

3.3 Multicausality

Choosing a time frame for averaging, finding minimums and maximums and performing other calculations, is challenging because significant causes may be present at *multiple time scales* (e.g. the task is present during this entire sequence, but teacher talk events are point-like pulses, and the ongoing variable updates each have their own time frames for calculations). We refer to this as a third data challenge – *multicausality* - which traditional statistical models sometimes approach via multivariate methods. Multivariate correlations are expressed in terms of strengths summed over some period of time or group of data. However, at some scales, short-range dynamics are crucial to long-range causes, but would be obscured by summation over time. How should we represent and analyze multivariate relationships that are changing over time (perhaps rapidly), without oversimplifying them to inert quantities, when their impacts on the system are subtle and time-sensitive? For example, a cause may be building for some time before it exerts its influence on the system. One potential solution we have explored is to capture the network statistics of many performances and use inductively evolved rule sets of a network of relations as a foundation for near-real time assessments relating a current performance to that network.

3.4 Intersectionality

A fourth problem is *intersectionality* at multiple scales. Intersectionality is a form of multicausality in which influences from diverse scales of space and time arrive at *a particular moment in time and space* to cause a joint effect. This problem implies that we need methods appropriate to represent and analyze dynamic geospatial distributions as spaces with probabilities for causing and responding to impacts. Differing amounts and types of intersectionality vary by both the *scale* (e.g. whether the nexus of causes is at the micro, meso or macroscopic level of the system) and also by *positionality* in the network of factors. For example, in Figure 1, relationships at the horizontal plane that are the historical impacts of task 1, are as important at the time task 2 begins, as are the hierarchical relationships of the task 2 requirements. Which exact mixture is determining the evolving context of task 2 and how do these relate to the user's intentions, actions and artifacts?

3.5 Nonlinearity

A fifth data challenge is *nonlinearity*. In addition to familiar relationships expressible by nonlinear functions (e.g. exponential growth), complex systems generally involve multiple interacting relationships expressible by partial differential equations. What are the most appropriate mathematical ideas needed? How much of the arsenal of existing statistical analysis do we have to abandon and which tools and methods can we leverage as we undertake an analysis that is cognizant of the challenges while also remaining relevant to the domain field? We have had success in using *symbolic regression* (Schmidt & Lipson, 2009), an application of genetic algorithms to the discovery of dynamic patterns in complex data.

4. SUMMARY DATA CHALLENGE = MODEL BUILDING

One level up from atomistic time-based events are time-independent *summaries*, which can function as time-dependent atoms for larger and larger summaries, giving both hierarchical organization and time-series power as a representation and analysis system. Hierarchical temporal features are involved in human memory and analysis skills (Hawkins & Blakeslee, 2004), which if understood more clearly, may give future automated assessments the ability to think about and process complex human performance information in complex ways. To do so, the system of summaries will most likely be organized by a conceptual framework that creates chains of evidence from performance information to intermediate and higher levels of representation, some of which are used to report on performance and others as a basis for adapting the digital learning experience. Advanced filtering options and data visualization and analysis allow both human and machine users to dissect and use summary data.

Through the Leverage methodology, the cycles of data collection, reduction, analysis, and reporting occur simultaneously and continuously during user interactions, and in near-real time, facilitating timely, authentic information about user performance. Assessments are created using a scripting language that accesses objects such as events and their attributes, player attributes and a subset of summaries called queries. Each assessment rule is triggered by an event and processed in order of priority assigned by the creator of the rule. Many rules can be processed from a single event, providing an operational platform for *subsumption architectures* (Brooks, 1986, 1999) and *quasi-homomorphisms* (Holland, 1995; Holland, Holyoak, Nisbett, & Thagard, 1986) for decisions. Priorities help in processing rules in an order that may contain dependencies to other rules. The simulation can thus report activity concerning what the user interface allows as well as the internal states of the machine that result from either user interactions or the inherent and emergent behaviors of the system's algorithms.

At the administrative backend of Leverage is an inductive model builder that creates a representation of the aggregated user paths in the network. Assessing proficiency in reaching a goal, as well as the extent and contributions of the constituent direct and indirect influences of actions leading to a goal, lies at the root of the model-building capability. Initially, assessments provide a mechanism to quantify abstract behaviors in the digital space (e.g. integrity or honor in a military simulation, or skill in differentiating instruction and understanding the psychology of learners in simSchool) and to pose hypotheses. Leverage then builds a representation of the aggregated experience of users of the application in the form of an attributes tracking model that reports on the statistical properties of the network.

An example in Figure 2 shows an assessment determination in the online game and simulation "America's Army," specifically modelling the Army's *Every Soldier a Sensor* developmental initiative. Identifying randomly placed target objects in battle impacts the simulation's scoring using the Army's core value system. Correctly identifying objects involves updating someone's record of "Honor," "Duty" and triggered by the situation labeled "es2ObjectReported." An assessment developer can see that Honor is updated 14.4% of the time when the event es2ObjectReported occurs, whereas Duty is impacted 40% of the time. On the other end of the influence line between es2ObjectReported and the two core values, Duty and Honor, the percentage contribution of es2ObjectReported toward their updates is so small that a zero is reported due to the fact that the simulation is so robust that many other events and assessments contribute to Duty and Honor, dwarfing the impact of es2 events. Generally, this particular Leverage analytic tool models the relationship between user action (events), rule-driven assessments (queries) and a digital representation of a target learning behaviour (attribute).

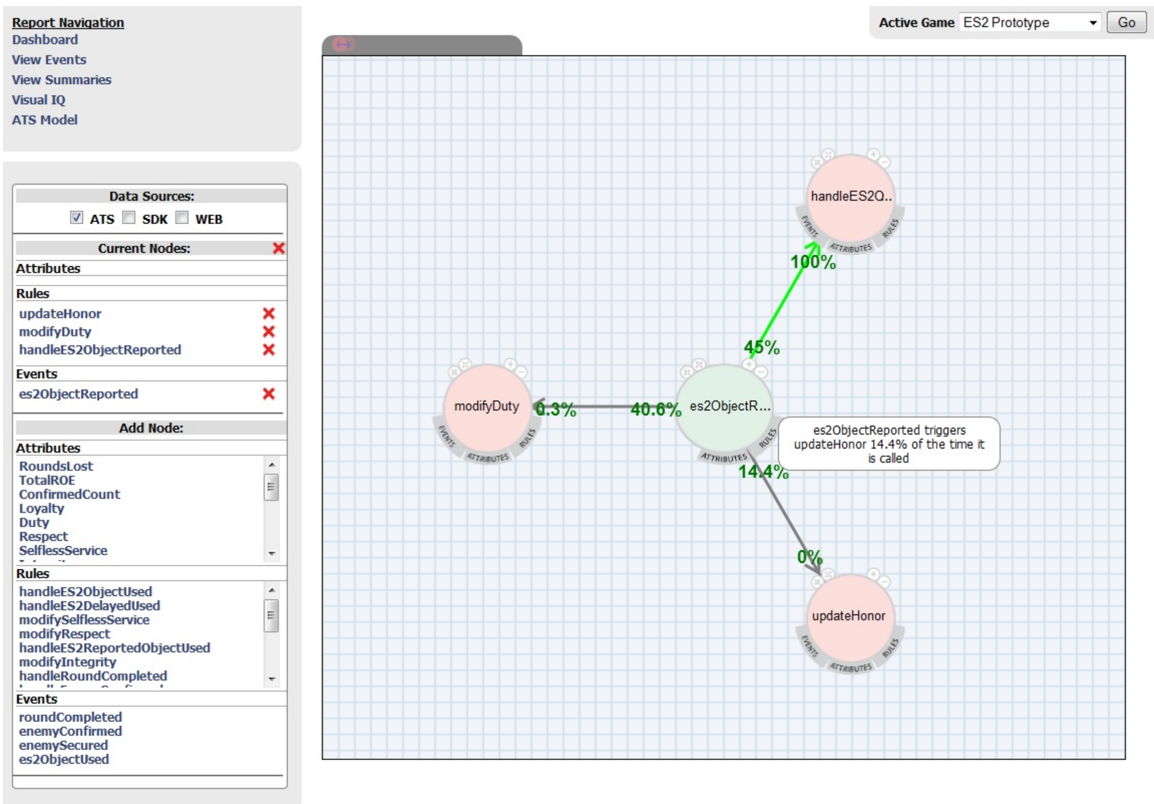

Figure 2. Inductive mapping of the network formed by user paths creates a first-order understanding of the knowledge of users who are performing in the digital space. In this screen, a user's current state of "Honor" is updated 14.4% of the time when an event "es2ObjectReported" is reported.

Network diagraphs (Albert & Albert-Laszio, 2002; Sporns, 2011) such Figure 2 capture important weights among resources in the digital learning environment, and represent states of knowledge of the network. The states can be used in several ways: as benchmarks for performance, for predictions of behavior and performance, and to trigger adaptive responses to guide or tutor the player via rewards and consequences of actions. The states can be stored and recalled in sequences and in relationship to concurrent slices of the user population to computationally represent a chain of evidence that links what a single user does in the application, with what everyone does, or what experts do, or what experts want people to do. The network viewpoint is thus useful for analyzing a user's performance within the social and cultural contexts of learning, teaching, & educational systems (Gibson, 2006).

5. SUMMARY

This paper briefly presented some of the data challenges and analysis approaches in the dynamic performance environment of a digital simulation. At the atomistic level of performance events, the challenges include time and event segmentation, cyclic dynamics, multicausality, intersectionality, and nonlinearity. At the summary level, the key challenge is model building. To ground the discussion, the paper uses an example of a simulation designed to develop teachers - simSchool – that is integrated with an adaptive content delivery and analytics database – Leverage, and briefly outlines how these two applications work together to solve the data challenges.

6. REFERENCES

Albert, R., & Albert-Laszio, B. (2002). Statisical mechanics of complex networks. *Reviews of Modern Physics*, *74*(January), 47-97.

Aldrich, C. (2004). *Simulations and the future of learning: an innovative (and perhaps revolutionary) approach to e-learning.* San Francisco: John Wiley & Sons.

Behrens, J., Frezzo, D., Mislevy, R., Kroopnick, M., & Wise, D. (2008). Structural, functional, and semiotic symmetries in simulation-based games and assessments. In E. Baker, J. Dickieson, & W. Wufleck (Eds.), *Assessment of problem solving using simulations* (pp. 59-80). NY: Lawrence Erlbaum Associates.

Bennett, R. E. (2010). Cognitively Based Assessment of, for, and as Learning (CBAL): A Preliminary Theory of Action for Summative and Formative Assessment. *Measurement: Interdisciplinary Research & Perspective*, *8*(2-3), 70-91. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/15366367.2010.508686

Brooks, R. (1986). A robust layered control system for a mobile robot. *Journal of Robotics and Automation*, *2*(April), 14-23.

Brooks, R. (1999). *Cambrian intelligence: The early history of the new AI*. Cambridge, MA: MIT Press.

Choi, Y., Rupp, A., Gushta, M., & Sweet, S. (2010). Modeling learning trajectories with epistemic network analysis: An investigation of a novel analytic method for learning progressions in epistemic games. *National Council on Measurement in Education* (pp. 1-39).

Christensen, R., Tyler-Wood, T., Knezek, G., & Gibson, D. (2011). SimSchool: An online dynamic simulator for enhancing teacher preparation. *International Journal of Learning Technology*, *6*(2), 201-220.

Gibson, D. (2006). Elements of Network-Based Assessment. In D. Jonson & K. Knogrith (Eds.), *Teaching Teachers to Use Technology* (pp. 131-150). New York: Haworth Press.

Gibson, D., Aldrich, C., & Prensky, M. (2007). *Games and Simulations in Online Learning: Research and Development Frameworks*. (D. Gibson, C. Aldrich, & M. Prensky, Eds.). Information Science Publishing.

Hawkins, J., & Blakeslee, S. (2004). *On intelligence*. New York: Henry Holt and Company.

Hofstee, W., de Raad, B., & Goldberg, L. (1992). Integration of the Big Five and Circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, *63*(1), 146-163. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1494982

Holland, J. (1995). *Hidden order: How adaptation builds complexity*. *Helix Books*. Cambridge, MA: Perseus Books.

Holland, J., Holyoak, K., Nisbett, R., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). Knowing what students know: The science and design of educational assessment. *Committee on the Foundations of Assessment, Board on Testing and Assessment, Center for Education, National Research Council*. Washington, DC: National Academy Press. Retrieved from http://wwwnapedu/books/0309072727/html/

Plutchik, R., & Conte, H. R. (1997). *Circumplex models of personality and emotions* (1st ed., p. xi, 484 p.). Washington, DC: American Psychological Association.

Prensky, M. (2001). Digital Natives, Digital Immigrants. *On the Horizon*, *9*(5).

Quellmalz, E., Timms, M., Buckley, B., Davenport, J., Loveland, M., & Silberglitt, M. (2012). 21st century dynamic assessment. In M. Mayrath, J. Clarke-Midura, & D. Robinson (Eds.), *Technology-based assessments for 21st Century skills: Theoretical and practical implications from modern research* (pp. 55-90). Charlotte, NC: Information Age Publishers.

Rupp, A., Gushta, M., Mislevy, R., & Shaffer, D. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *Journal of Technology, Learning, and Assessment*, *8*(4), 1-45.

Schmidt, M., & Lipson, H. (2009). Symbolic regression of implicit equations. *GENETIC PROGRAMMING THEORY AND PRACTICE*, *7*(Chap 5), 73-85.

Shaffer, D. (2007). Epistemic games. *Innovate*. Retrieved from http://innovateonline.info/pdf/vol1_issue6/Epistemic_Games.pdf

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Frank, K., et al. (2009). Epistemic Network Analysis: A Prototype for 21st-Century Assessment of Learning. *International Journal of Learning and Media*, *1*(2), 33-53. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/ijlm.2009.0013

Sporns, O. (2011). *Networks of the brain*. Cambridge:MA: MIT Press.

Zibit, M., & Gibson, D. (2005). simSchool: The game of teaching. *Innovate*, *1*(6). Retrieved from http://www.innovateonline.info/pdf/vol1_issue6/simSchool_-__The_Game_of_Teaching.pdf

Zibit, M., Gibson, D., & Halverson, B. (2006). SimSchool: Preparing tomorrow's teachers to improve student results. In C. Crawford (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2006* (pp. 3130-3136). Chesapeake, VA: AACE.