AIED Is Splitting Up (Into Services) and the Next Generation Will Be All Right

Benjamin D. Nye

Institute for Intelligent Systems, University of Memphis
365 Innovation Dr. Memphis, TN 38152
benjamin.nye@gmail.com

Abstract. Advanced learning technologies are reaching a new phase of their evolution where they are finally entering mainstream educational contexts, with persistent user bases. However, as AIED scales, it will need to follow recent trends in service-oriented and ubiquitous computing: breaking AIED platforms into distinct services that can be composed for different platforms (web, mobile, etc.) and distributed across multiple systems. This will represent a move from learning platforms to an ecosystem of interacting learning tools. Such tools will enable new opportunities for both user-adaptation and experimentation. Traditional macro-adaptation (problem selection) and step-based adaptation (hints and feedback) will be extended by meta-adaptation (adaptive system selection) and micro-adaptation (event-level optimization). The existence of persistent and widely-used systems will also support new paradigms for experimentation in education, allowing researchers to understand interactions and boundary conditions for learning principles. New central research questions for the field will also need to be answered due to these changes in the AIED landscape.

1 Introduction

Initial efforts to bring learning technology into schools faced hardware hurdles, such as insufficient computing resources. Later efforts encountered serious barriers related to matching technology to teachers' beliefs, pedagogy, and resource constraints. While all of these barriers are still relevant, learning technology is endemic in higher education and has made significant footholds in K-12 schools, with estimates of 25-30% of science classes using technology as early as 2012 (Banilower, Smith, Weiss, Malzahn, Campbell, & Weis, 2013). Correspondingly, an influx of investment into educational technology has occurred, with online learning doubling from a $50b industry to a $107b industry in only three years (Monsalve, 2014).

Future barriers will not be about getting learning technology into schools: they will be about competing, integrating, and collaborating with technologies already in schools. This is not an idle speculation, as it is already occurring. In a recent multi-year efficacy study to evaluate a major adaptive learning system, some teachers started using grant-purchased computers to use other math software as well (Craig, Hu, Graesser, Bargagliotti, Sterbinsky, Cheney, & Okwumabua, 2013). After working for many years to get teachers to use technology, the point may come where they are using so many technologies that it is difficult to evaluate an intervention in isolation.

Some research-based artificial intelligence in education (AIED) technologies have already grown significant user bases, with notable examples that include the Cognitive Tutor (Ritter, Anderson, Koedinger, & Corbett, 2007), ALEKS (Falmagne, Albert, Doble, Eppstein, & Hu, 2013), and ASSISTments (Heffernan, Turner, Lourenco, Macasek, Nuzzo-Jones, & Koedinger, 2006). Traditionally non-adaptive systems with large user bases, such as Khan Academy and EdX, have also started to add basic adaptive learning and other intelligent features (Khan Academy, 2015; Siemens, 2013).

Large-scale online platforms are not just the future of learning, but they are also the future of research. Traditional AIED studies have been limited to dozens to hundreds of participants, sometimes just for a single session. While such studies will remain important for isolating new learning principles and collecting rich subject data (e.g., biometrics), large-scale platforms could be used to run continuously-randomized trials across thousands of participants that vary dozens or even hundreds of parameters (Beck and Mostow 2006; Liu, Mandel, Brunskill, & Popovic, 2014). Even for AIED work not based on such platforms, it is increasingly feasible to "plug in" to another system, with certain systems serving as active testbeds for 3rd-party experiments (e.g., ASSISTments and EdX).

The difference is qualitative: rather than being limited to exploring a handful of factors independently, it will be possible to explore the relative importance of different learning principles in different contexts and combinations. In many respects, this means not just a change to the systems, but to the kinds of scientific questions that can and will be studied. These opportunities raise new research problems for the field of AIED. A few areas related areas will reshape educational research: Distributed and Ubiquitous Intelligent Tutoring Systems (ITS), Four-Loop User Adaptation, AI-Controlled Experimental Sampling, and Semantic Messaging. Some new frontiers in each of these areas will be discussed.

2 Distributed and Ubiquitous AIED

As implied by the title, AIED technologies are approaching a juncture where many systems will be splitting up into an ecosystem of reusable infrastructure and platforms. The next generation of services will be composed of these services, which may be hosted across many different servers or institutions. More specifically, we may be reaching the end of the traditional four-component ITS architecture with four modules: Domain, Pedagogy, Student, and Communication (Woolf, 2010). While the functions of all these modules will still be necessary, there is no reason to think that any given ITS must *contain* all these components, in the sense of building them, controlling them, or owning them. The future for ITS may be to blow them up so that each piece can be used as a web-service for many different learning systems.

With respect to other online technologies, learning technology is already behind. On even a basic blog site, a user can often log in using one of five services (e.g., Google, Facebook), view adaptively-selected ads delivered by cloud-based web services that track users across multiple sites, embed media from anywhere on the internet, and meaningfully interact with the site on almost any device (mobiles, tablets, PC). In short, most web applications integrate and interact with many other web services, allowing them to be rapidly designed with robust functionality and data that no single application would be able to develop and maintain.

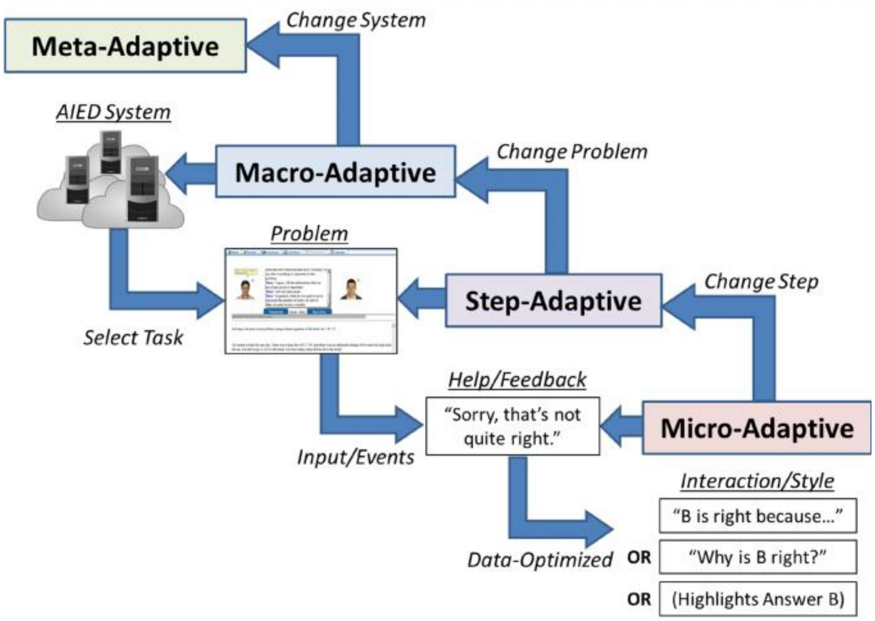
From the standpoint of AIED, moving in this direction is an existential necessity. Without pooling capabilities or sharing components, serious academic research into educational technologies may be boxed out or surpassed by the capabilities of off-the-shelf systems, many of which will have closed architectures. Unfortunately, while industry research can offer powerful results, competing pressures can lead to under-reporting: publishing research is costly, time-consuming, and can risk disclosing trade secrets or unfavorable empirical findings. While some companies make the investment to generalize their research, many others do not. By comparison, academic institutions and research-active commercial systems should be motivated to share and combine technologies to build more effective and widely-used learning technology. This model of collaborative component design stands alone in making platforms that co-exist with major commercial endeavors, such as web-browsers (FireFox), operating systems (Linux), and statistical packages (R; R Core Team, 2013). Moreover, service-oriented computing allows for a mixture of free research development and commercial licensing of the same underlying technologies.

The benefits of moving toward service-oriented AIED will be substantial. First, they should enable AIED research to deeply specialize, while remaining widely applicable due to the ability to plug in to other platforms with large and sustained user bases. In such an ecosystem, user adaptation will be free to expand beyond the canonical inner loop and outer loop model (VanLehn, 2006). Composing and coordinating specialized AIED services will also demand greater standardization and focus on data sharing between systems. While this process may be painful initially, standards for integrating data across multiple systems would enable the development of powerful adaptation, analytics, and reporting functionality that would greatly reduce barriers for developing AIED technology and studying its effects on learners.

3 Four-Loops: Above Outer Loops and Under Inner Loops

One implication of scaling up AIED and moving beyond the standard four-component ITS model is that adaptation to users may become prevalent at grain sizes larger and smaller than traditional ITS. VanLehn (2006) framed the adaption from tutoring systems as consisting of an outer loop (selecting problems) and an inner loop (providing help and feedback on specific problem steps). These are often referred to as “macro-adaptivity” and “step-based adaptivity.” However, recent developments have shown the first steps toward “meta-adaptivity,” where the system adapts to the user by shifting the learner to an entirely different ITS system (which may then adapt to the user differently). Likewise, research on “micro-adaptivity” has looked at the benefits for using data to fine-tune interactions below the problem step level (e.g., keystroke-level inputs, emotion detection, presentation modes or timing of feedback). This implies a four-loop model for user adaptation, as shown in Figure 1.

Fig. 1. Four-Loop User Adaptation



3.1 Meta-Adaptation: Handoffs Between Systems

Meta-adaptation has only become possible recently, due to increasing use and maturity of AIED technology. In the past, learning technologies such as ITS were trapped in sandboxes with no interaction. Due to service-oriented approaches, systems have taken the first steps toward real-time handoffs of users between systems. For example, in the recent Office of Naval Research STEM Grand Challenge, two out of four teams integrated multiple established adaptive learning systems: Wayang Outpost with ASSISTments (Arroyo, Woolf, & Beal, 2006; Heffernan et al., 2006) and AutoTutor with ALEKS (Nye, Windsor, Pavlik, Olney, Hajeer, Graesser, & Hu, in press). Other integration efforts are also underway as part of the Army Research Lab (ARL) Generalized Intelligent Framework for Tutoring (GIFT) architecture, which is built to integrate external systems (Sottilare, Goldberg, Brawner, & Holden, 2012) and version of AutoTutor has also been integrated with GIFT.

These initial integrations represent the first steps toward meta-adaptation: transferring the learner between different systems based on their needs and performance. This type of adaptation would allow learners to benefit from the complementary strengths of multiple systems. For example, learners that benefit most from animated agents might be sent to systems such agents (i.e., trait-based adaptation). Alternatively, different types of learning impasses or knowledge deficiencies may respond best to learning activities in different systems (i.e., state-based adaptation). One problem that this approach might mitigate is the issue of wheel spinning, where an adaptive system detects that it cannot serve the learner's current needs (Beck & Gong, 2013). Meta-adaptation might also mean referring the learner to a human instructor, tutor, or peer. In general, meta-adaptation would focus on passing students and knowledge between different adaptive learning contexts (both AI-based and human).

Meta-adaptation is the maximum possible grain size, which makes it somewhat different from standard adaptation because users are transferred to an entirely different system. This type of adaptation likely requires either distributed adaptation or brokered adaptation. Distributed adaptation would involve individual systems deciding when to refer a learner to a different system and possibly trusting the other system to transfer the student back when appropriate. This would be analogous to doctors in a hospital, who rely on networks of specialists who share charts and know enough to make an appropriate referral, but may use their own judgment about when and how they make referrals. On the converse, brokered adaptation would require a new type of service whose purpose is to monitor student learning across all systems (i.e., a student model integrator) and make suggestions for appropriate handoffs. This service would be consulted by each participating AIED system, probably as part of their outer loop. In the long term, such a broker may be an important service, because it could help optimize handoffs and ensure that students are transferred appropriately. Such brokers might also play a role for learners to manage their data and privacy settings. Other models for coordinating handoffs might also emerge over time.

3.2 Micro-Adaptation: Data-Optimization and Event Streams

In addition to adaptivity at the largest grain size (selecting systems), research on the smallest grain sizes (micro-adaptation) is also an important future area. Micro-adaptation involves optimizing for and responding to the smallest level of interactions, even those that are not associated with a traditional user input on a problem step. For anything but simple experiments, this type of optimization and adaptation is too fine-grained and labor-intensive to perform by hand at scale, meaning that it will need to rely on data-driven optimizations such as reinforcement learning. Chi, Jordan and VanLehn (2014) used reinforcement learning to optimize dialog-based ITS interactions in the Cordillera system for Physics, which showed potential gains of up to 1σ over poorly-optimized dialog or no dialog. Dragon Box has taken a related approach by optimizing for low-level user interface and click-level data, by applying trace-based models to find efficient paths for learning behavior and associated system responses (Andersen, Gulwani, & Popovic, 2013).

These lines of research represent the tip of the iceberg for opportunities for micro-adaptation. A variety of low-level data streams have not yet been leveraged. Continuous sensor data, such as emotion sensors or speech input waveforms, may present rich opportunities for exploring fine-grained user-adaptation based on algorithmic exploration of possible response patterns. Low-level user interface optimization may also help improve learning, such as human-computer interaction design or keystroke-level events or mouse-over actions (i.e., self-optimizing interfaces).

Both the strength and the drawback of micro-optimization is that it is will tightly fit the specific user interface or content (even down to specific words in text descriptions). Optimizing for a particular presentation of a problem can lead to learning efficiency gains by emphasizing parts that are salient to learning from that specific case, while skipping or downplaying other features. However, micro-level optimization will likely suffer from versioning issues (e.g., changes to small problem elements potentially invalidating prior data and policies) and also transferability issues (e.g., an optimized case not transferring well from a desktop to a mobile context). Solutions to weight the relevance of prior data will be required to address issues related to altered problems or new contexts (e.g., mobile devices, classroom vs. home, different cultural contexts).

4 AI-Controlled Experimental Sampling

Techniques for micro-adaptation may also reshape experimental methods. Artificial intelligence can play a major role in the experimental process itself, which is a type of efficient search problem. Educational data mining research has already started looking at dynamically assigning subjects to different learning conditions based on multi-armed bandit models (Liu, Mandel, Brunskill, & Popovic, 2014). Multi-armed bandit models assume that each treatment condition is like a slot machine with different payout distributions (e.g., student learning gains). These models are common in medical research, where it is important to stop treatments that show harms or a consistent lack of benefit. They allow building intelligent systems that explore new strategies, while pruning ineffective ones.

The field is only taking its first baby steps for these types of experimental designs. Fundamental research is needed to frame and solve efficient-search problems present in AIED experiments. Based on varying different parameters and interactions in the learning experience, learning environments can search for interpretable models that predict learning gains. In the long term, models for automated experimentation may even allow comparing the effectiveness of different services or content modules, by randomly selecting them from open repositories of content.

The most difficult aspect of this problem is likely to be the interpretability. While multi-arm bandit models can be calibrated to offer clear statistical significance levels between conditions, models that traverse the pedagogical strategy space are often too granular to allow for much generalization. For example, some popular models for large learning environment focus on efficient paths or traces of learning behavior and associated system responses (Andersen, Gulwani, & Popovic, 2013). Unfortunately, these models are often not easily generalizable: they may capture issues tied to the specific system or may tailor instruction to specific problems so tightly that it is difficult to infer theoretical implications (Chi, Jordan & VanLehn, 2014).

New techniques are needed that can automatically explore the space of pedagogical designs, but that can also output interpretable statistics that are grounded in theories and concepts that can be compared across systems. This is a serious challenge that probably lacks a general algorithmic solution. Instead, such mappings will probably be determined by the constraints of learning and educational processes. A second major challenge is the issue of integrating expert knowledge with statistically-sampled information. Commonly, expert knowledge is used to initially design a system (e.g., human-defined knowledge prerequisites), which is later replaced by a statistically-inferred model after enough data is collected. However, in an ideal world, these types of heterogeneous data would be gracefully integrated (e.g., treating expert knowledge as Bayesian prior weights). Future research in AIED will need to identify where this sort of expert/statistical hybrid modeling is needed, and match these problems with techniques from fields of AI and data modeling that specialize in these issues. Ultimately, a goal of this work should be to blur the lines between theory and practice by building systems that can both report and consume theoretically-relevant findings.

5 Semantic Messaging: Sharing Components and Data

To share technology effectively, AIED must move toward open standards for sharing data both after-the-fact (i.e., repositories) and also in real-time (i.e., plug-in architectures). The first steps in these directions have already been taken. Two notable data repository projects with strong AIED roots exist: the Pittsburg Science for Learning Center (PSLC) DataShop (Koedinger, Baker, Cunningham, Skogsholm, Leber, & Stamper, 2010) and the Advanced Distributed Learning (ADL) xAPI standards for messaging and learning record stores (Murray & Silvers, 2013). The IMS Global Specifications are also a move in this direction (IMS Global, 2015).

Due to solid protocols in messaging technologies, the technical process of exchanging data between systems at runtime is not onerous. The larger issue is for a receiving system to actually apply that data usefully (e.g., understand what it means). Hidden beneath this issue is a complex ontology alignment problem. In short, each learning technology frames its experiences differently. When these experiences and events are sent off to some other system, the designers of each system need to agree about what different semantics mean. For example, one system may say a student has “Completed” an exercise if they viewed it. Another might only mark it as “Completed” if the learner achieved a passing grade on it. These have very different practical implications. Likewise, the subparts of a complex activity may be segmented differently (e.g., different theories about the number of academically-relevant emotions). While efforts have been made to work toward standards, this seldom solves the problem: the issue with standards is that there tends to be so many of them.

So then, ontology development must play a key role for the future of ITS interoperability. There are multiple ways that this might occur. Assuming the number of standards is countable, it would be sufficient to have an occasional up-front investment to develop and update explicit mappings between ontologies by hand. While this is low-tech, it works when the number of terms is fairly small. For larger ontologies of AIED behavior and events, it may be possible to align ontologies by applying both coding systems to a shared task (e.g., build benchmark tasks that are then marked up with messages derived from that ontology).

By collecting data on messages from benchmark tasks, it may be possible to automate much of the alignment between ontologies, particularly for key aspects such as assessment. Research on Semantic Web technologies is also very active, and may offer other effective solutions to issues of ontology matching and alignment (Shvaiko & Euzenat, 2013). The final approach is to simply live without standards and allow the growth of a folksonomy: common terms that are frequently used. These terms can then become suggested labels, with tools that make their use more convenient and prevalent. The one approach that should *not* be taken is to try to develop a super-ontology or new top-down standard for the types of information that learning systems communicate. While there are roles for such ontologies, top-down ontologies have never achieved much support within research or software development communities.

6 Closing Remarks

The future for AIED should be a bright one: expansion of learning software into schools will ultimately result in unprecedented diversity and size of user bases. The areas noted in this paper are only the first wave for new AIED opportunities. In time, it will be possible to explore entirely new classes of questions, such as mapping out continuous, multivariate functional relationships between student factors and pedagogical effectiveness of certain behaviors. Systems such as personal learning lockers for data would allow for longitudinal study of learning over time, either in real-time or retrospectively. A major game-changer for future learning research will probably be data ownership and privacy issues: data will exist, but researchers will need to foster best-practices for data sharing, protection, and archiving.

With this wealth of data, researchers will be able to connect learning to other relationships and patterns from less traditional data sources. In 20 years, the range of commonly-available sensor data will be dizzying: geolocation, haptic/acceleration, camera, microphone, thermal imaging, social ties, and even Internet-of-Things devices such as smart thermostats or refrigerators. Moreover, the ecosystem of applications leveraging this data will likewise be more mature: your phone might be able to tell a student not only that their parents left them a voicemail, but that they sounded angry. This event might then be correlated with a recent report card, and the consequences of the interaction might be analyzed. Learning is a central facet of the human experience, cutting across nearly every part of life. To that end, as life-long learning becomes the norm, the relationship between life and learning will become increasingly important. By consuming and being consumed in a distributed and service-oriented world, AIED will be able to play a major role in shaping both education and society.

References

1. Andersen, E., Gulwani, S., & Popovic, Z. (2013). A trace-based framework for analyzing and synthesizing educational progressions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 773-782). ACM.

2. Arroyo, I., Woolf, B. P., & Beal, C. R. (2006). Addressing cognitive differences and gender during problem solving. *International Journal of Technology, Instruction, Cognition and Learning*, *4*, 31-63.
3. Banilower, E . R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research, Inc.
4. Beck, J. E., & Gong, Y. (2013, January). Wheel-spinning: Students who fail to master a skill. In *Artificial Intelligence in Education* (pp. 431-440). Springer Berlin.
5. Chi, M., Jordan, P., & VanLehn, K. (2014). When Is Tutorial Dialogue More Effective Than Step-Based Tutoring? In *Intelligent Tutoring Systems* (pp. 210-219). Springer: Berlin.
6. Craig, S. D., Hu, X., Graesser, A. C., Bargagliotti, A. E., Sterbinsky, A., Cheney, K. R., & Okwumabua, T. (2013). The impact of a technology-based mathematics after-school program using ALEKS on student's knowledge and behaviors. *Computers & Education*, *68*, 495-504.
7. Falmagne, J. C., Albert, D., Doble, C., Eppstein, D., & Hu, X. (2013). *Knowledge Spaces: Applications in Education*. Springer Science & Business Media.
8. Graesser, A. C. (2011). AutoTutor. In P. M. McCarthy, & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation and resolution*. Hershey, PA: IGI Global.
9. Heffernan, N. T., Turner, T. E., Lourenco, A. L., Macasek, M. A., Nuzzo-Jones, G., & Koedinger, K. R. (2006). The ASSISTment Builder: Towards an Analysis of Cost Effectiveness of ITS Creation. In *FLAIRS Conference* (pp. 515-520).
10. IMS Global (2015). Learning Tools Interoperability. Retrieved from: http://www.imsglobal.org/lti/index.html on Feb 22, 2015.
11. Khan Academy (2015). About Khan Academy. Retrieved www.khanacademy.org/about on Feb. 22, 2015.
12. Koedinger, K. R., Baker, R. S., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S.J.d. Baker (Eds.) *Handbook of educational data mining*, 43-53.
13. Liu, Y. E., Mandel, T., Brunskill, E., & Popovic, Z. Trading Off Scientific Knowledge and User Learning with Multi-Armed Bandits. *Educational Data Mining (EDM) 2014*. 161-168. Springer
14. Monsalve, S. (2014). A venture capitalist's top 5 predictions for 2015. *Fortune*. Retrieved http://fortune.com/2014/12/29/a-venture-capitalists-5-predictions-for-2015/ on Feb. 22, 2015.
15. Mostow, J., & Beck, J. (2006). Some useful tactics to modify, map and mine data from intelligent tutors. *Natural Language Engineering*, *12*(02), 195-208.
16. Murray, K., & Silvers, A. (2013). A learning experience. *Journal of Advanced Distributed Learning Technology*, *1*(3-4), 1-7.
17. Murray, T., Blessing, S., & Ainsworth, S. (2003). *Authoring tools for advanced technology learning environments: Toward cost-effective adaptive, interactive and intelligent educational software*. Springer.
18. Nye, B. D., Windsor, A., Pavlik, P. I., Olney, A., Hajeer, M., Graesser, A. C., & Hu, X. (In Press). Evaluating the Effectiveness of Integrating Natural Language Tutoring into an Existing Adaptive Learning System. In *Artificial Intelligence in Education (AIED) 2015*.
19. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0

20. Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic bulletin & review*, *14*(2), 249-255.
21. Shvaiko, P., & Euzenat, J. (2013). Ontology matching: state of the art and future challenges. *Knowledge and Data Engineering, IEEE Transactions on*, *25*(1), 158-176.
22. Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 0002764213498851.
23. Sottilare, R. A., Goldberg, B. S., Brawner, K. W., & Holden, H. K. (2012). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2012*.
24. VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, *16*(3), 227-265.
25. Woolf, B. P. (2010). Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann.