Educational Data Mining: Current Research and Open Questions

Ma. Mercedes T. Rodrigo
Ateneo Laboratory for the Learning Sciences
Ateneo de Manila University
Loyola Heights, Quezon City, Philippines
+63 (2) 426 6071
mrodrigo@ateneo.edu

ABSTRACT
Educational data mining (EDM) refers to the application of statistical and machine learning methods to educational data in order to achieve one of four typical ends: improvement of student models, improvement of subject matter domain structures, studying pedagogical support and refining educational theories. An interdisciplinary field, EDM draws on mathematics, computer sciences, cognitive psychology, education theory, sociology and others. This paper walks the reader through the EDM process and then discusses recent work and open questions in the first three application areas. The paper hopes to introduce young researchers to the field and suggest problems that are still open for investigation.

1. EDUCATIONAL DATA MINING
Data mining (DM) is defined as the process of extracting interesting, interpretable, useful information from data [37]. It has been used for many years now on large databases in the fields of medicine and business, among others. Educational data mining (EDM), on the other hand, is refers to the application of data mining techniques to educational data in order to answer questions related to learning [37]. EDM is both narrower and broader than traditional DM. It is narrower in focus area, finding theoretical grounding in educational theory and cognitive psychology. It is broader in that it makes use of a wider variety of approaches. Aside from the usual DM methods of classification, clustering, and association rule mining, EDM employs regression, correlation, visualization, and others that are not typically part of the DM repertoire.

The objectives of any given EDM project vary depending on the stakeholders of interest [37]. If learners are the focus, the results of an EDM analysis are often used to personalize learning, to recommend activities or resources, or to provide hints. For educators themselves, EDM provides a means for analyzing student learning, behavior, and attitudes; to assess teachers, teaching techniques, or curriculum; and to detect when students need further support. For course developers, EDM analyses helps evaluate course content and structure, assesses the effectiveness of design decisions, and to automatically construct both teacher and student models. Organizations can use EDM to rationalize technology investments and to find better ways to improve student achievement and retention. Finally, EDM can help educational administrators to find ways to better organize resources and educational offerings, to utilize resources more effectively, and to evaluate teachers and curricula.

There are four key applications of EDM [7]: Improvement of student models, discovering or improving models of subject matter domain structure, studying pedagogical support, and refining educational theories. Student models refer to information about the student that is relevant to learning and may include items such as student knowledge, motivation, attitudes, behaviors, and so on. Educational software uses these features to provide learners (to the extent possible) with learning experiences that best suit their individual needs.

Subject matter domain knowledge, on the other hand, refers to the way the subject matter (e.g. math, history) is internally structured within the educational software [46]. The structure can represent what quiz items test for the same skill, how content should be sequenced, what skills are related, and so on.

Pedagogical support, also known as instructional modeling, is the software's ability to change the mode of teaching based on inferences about the student's learning [46]. It entails providing students with the kind of support they need at the appropriate time. It also entails discovering what types of support are most effective, for different groups and under different circumstances.

Finally, EDM aims to contribute to educational theory by providing empirical evidence to explain educational phenomena. EDM analysis enables us to better understand these phenomena and to refine and extend theory accordingly [7].

The EDM community has been organizing itself since at least 2000 and has been hosting workshops and conferences since then. At these events, researchers share EDM methods as well as results from EDM-related studies.

The purpose of this paper is to discuss some of the ongoing investigations in the area. It hopes to familiarize young researchers with EDM and introduce them to open questions that are still open to investigation. The paper begins with a description of the EDM process. It then uses the three out of the four key application areas specified by [7] as a framework for the discussion of current work. Note that some studies are quite broad and cut across several areas, so some overlap is to be expected. The paper describes current studies and some of the questions that are worth pursuing. Immediately preceding the conclusion, the paper discusses some contemporary issues that affect the wide scale use of educational software and, consequently, EDM. The paper ends with a wrap up of major points.

2. THE EDM PROCESS
Almost every EDM study describes a multistage methodology. It begins with data collection and ends with implementation.

2.1 Data collection
Educational data mining presupposes that there is data for researchers and educators to mine. The data usually comes from educational software or computer-based learning environments that are specifically designed for fine-grained data collection.

Data types can range from student interactions to biometric signals. For example, [42] describes the use of an instrumented version of BlueJ, a Java integrated development environment, to capture all student compilations onto a central server. [38] on the other hand, describes the use of the Scatterplot Tutor, an intelligent tutoring system that teaches students to create and interpret scatterplots. The Scatterplot tutor captured all student interactions with the software, including student answers, hint requests, and so on. Finally, aside from capturing student-tutor interactions, [14]'s AutoTutor environment collects student posture, facial expressions, and gaze. Depending on how many students were involved in the experiment, how long they interacted with the software, and how finely-grained the data collection is, the size of the database can range from a few kilobytes to hundreds of Gigabytes.

2.2 Data cleaning

Once the data has been collected, it has to be cleaned filtered of any junk data that might introduce noise. Examples of junk data include but are not limited to invalid or incomplete records, records from administrator or experiment team test runs, data from pre-tests, and so on. The final data set should consist of valid records only.

2.3 Data summarization and/or clip generation

It is sometimes necessary to collapse data into meaningful chunks. If used to for machine learning without summarization or aggregation, sensor data collected at a rate of 8 samples per second, for example, will probably not yield meaningful analyses. Researchers have to decide on a meaningful time unit, e.g. 1 minute, 3 minutes, etc., and summarize the data accordingly by getting totals, averages, counts, and so on.

Alternatively or in addition, it may be necessary to group student transactions into clips, e.g. subsets of student-tutor interactions defined based on a criterion. For example, intervals may consist of 20 seconds worth of transactions or 5 to 8 actions [35].

2.4 Data labeling

For classification problems, it is necessary for the researchers to provide the classification algorithm with a training set consisting of sample inputs and sample outputs. This training set becomes the basis for constructing a model of the phenomenon under study. In some cases, the expected outputs are collected as part of the data gathering process, e.g. post-test results. In some cases, though, the labels have to be provided by a human judge.

The paper by [6], for example, describes a method for manually labeling playbacks of student activities as either "gaming" (attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material—[4]) or "non-gaming." Another paper by [27] describes the labeling of student actions as "confused" or "not confused." Labels do not have to be binary. The paper by [13] asks the human judges to assign one of seven states to log AutoTutor excerpts: boredom, confusion, flow, frustration, delight, neutral, and surprise.

2.5 Feature distillation or engineering

Feature engineering is the process of identifying or computing for attributes that may imply the existence or non-existence of the phenomenon of interest. In [27] study, the tell-tale signs of confusion among novice programmers included the time between compilations and the number of compilations with errors. [22]'s features, on the other hand, included the average moment-by-moment learning, number of opportunities to learn a specific concept, and the sum of moment-by-moment learning values.

Arriving at a list of meaningful features is still something of an art. Researchers have to draw on past literature about what is known to relate to phenomena of interest (when do classroom teachers say that a student is confused?) and then translate that to some measure that is quantifiable, given the data available. They may use features that past EDM researchers have used, assuming that the logs under study captured the same or similar elements. It is also possible to automate or semi-automated methods to arrive at a list of features. For example, [4] used fast correlation-based filtering and forward selection to arrive at the feature set for gaming the system.

2.6 Analysis

It is only after features have been defined and data has been clipped or summarized that the actual data mining process can begin. At this point, researchers usually make use of tools such as WEKA [19] or RapidMiner [28] to create machine learned models of the data. As mentioned earlier, these models can consist of classifications, clusters, association rules, regressions, correlations, visualizations, and others.

2.7 Validation

Both DM and EDM strive for parsimonious models, that is, the models should embody the complexity of the phenomenon with the fewest possible variables. Once the models have been produced, they have to undergo a process of validation in order to determine well the model fits the data. Validation of classifications can come in the form of computing the level of agreement between the model's output and some known measure of ground truth, e.g. human judgments. Models that produce numerical results can be validated by looking at correlations.

2.8 Implementation

The ultimate test of a model's validity is its implementation in an actual learning environment. To close the loop, and assuming the models were found to be reasonably valid, the learning environment should be modified to take these models into account. Researchers can then see whether the interventions affect learners for the better.

3. EDM TO IMPROVE STUDENT MODELS

What are these different models that EDM researchers try to derive? As mentioned in the introduction, there is a great interest in modeling aspects of learners that are relevant to learning. These may include but are not limited to student knowledge, behavior and affect.

Some of the behaviors and affective states of interest are quite specific. [22]'s paper attempted to use EDM to model students' preparation for future learning (PFL). As the name suggests, PFL refers to the students' ability to learn future, related content quickly and easily [33]. One example of PFL in [11] describes a learning environment in which students teach teachable agents (computer programs that simulate other learners) science by creating concept maps. The software enables the agents to interpret the maps and answer questions. In addition to learning the actual science, concept mapping was a bonus skill, one that they found useful in learning future science lessons. The work of [22] found that the pattern most associated with PFL consists of three substantial moments of learning of comparable magnitudes, spread out over time.

A number of different behaviors have been the subject of study because of their direct relationship with learning. For at least the past decade, EDM researchers have been investigating gaming the system, for example, because of its negative correlation with learning [5]. In this same decade, other behaviors of interest emerged, since the ways in which students deviate from expected or desired patterns of interaction with learning systems, though, can be quite diverse. In [45], researchers point out that students can continue to use a learning system but not for the tasks for which it was intended. They describe an instance in which students were supposed to use a game-like exploratory learning environment to track down the causes of an epidemic. Instead, the students used the environment to put bananas on a toilet. [45]'s group labeled these behaviors as "WTF" or "without thinking fastidiously." They found that it was indeed possible to build automated detectors of WTF behaviors and that further studies were needed to determine how WTF behaviors were related to student achievement and affect.

Another behavior that has captured recent interest is what [9] calls "wheel spinning," the inability of students to master a skill in a timely manner. The study showed that students who are unable to master a skill by the 8th attempt are unlikely to master the skill at all. Giving these students more of the same types of exercise does not help them. They need a different intervention which may come in the form of remediation from within the learning system or from a teacher. The paper was an initial attempt at studying the phenomenon and, like [45]'s WTF behavior, there are a number of open questions related to wheel spinning that are still ripe for investigation: the relationship between wheel-spinning and other off-task behaviors such as gaming the system, the causes of wheel-spinning, and so on.

Aside from student behaviors, researchers are growing increasingly interested in studying student affect because affect relates to achievement, classroom behavior, student retention and attrition, and the students' overall educational experience. For example, [44]studied how boredom and disengagement might manifest themselves among readers. By intuition, reading times should be faster as text becomes easier and should be slower when text becomes harder. Their data showed that readers read at an appropriate pace when text was of average complexity but exhibited a pattern of disengagement with easier or harder portions of the text. Readers tended to spend too much or too little time at these extremes.

Researchers are also interested in broader phenomena such as overall student stress as long-term stress can tax health or lead to maladaptive behaviors [25]. By helping students track their stress levels and understand their stress triggers, researchers may be able to help them live healthier, happier, more academically successful lives.

Recent work by [15] studied mental health among engineering students in a university in Canada. To arrive at their models, they used a survey based on guidelines from the Canadian Mental Health Association. They found that the year of student and the number of hours of homework have the largest effects on student mental health. Second year students, for example, had the highest mental health scores, whereas first year students had the lowest. Self-actualization was negatively affected by a high number of in-class hours, but was positively affected by number of hours of homework.

The work of [30] attempted to determine which students were likely to drop out, given their social behavior. The inputs to their model included but were not limited to exam scores, courses taken, intensity of interpersonal behavior, number of mutually shared files, and publication co-authoring. The authors generated new features from each student's social network. If each student is regarded as a node, they computed for the number the number of arcs emanating from each note as well as the number arcs leading to a node. Characteristics of each students neighbors were also factored in: grade averages, proportion of enrolled and fulfilled course credits, credits per semester and so on. They found that students who had average grades but were in communication with students with good grades were more likely to successfully graduate than students with similar achievement levels who were not communicating with successful students.

4. EDM TO IMPROVE DOMAIN MODELS

Aside from student models, domain models are a topic of interest for EDM researchers. One can think of domain models as check lists of skills or knowledge that a student is supposed to learn. The items on the list may be related—one item may be the prerequisite or co-requisite of another item.

Deriving and representing the concepts or skills that relate to test question are a difficult problem. A single problem may require multiple skills. Skills may overlap. Also, students may answer a question incorrectly even when they already know the skill (slip) or answer a question correctly even when they do not know the skill (guess). Both slips and guesses introduce noise to the data [10]. Furthermore, in complex or open-ended problem solving domains such as computer programming [40] or learning games [20], multiple correct solutions or multiple paths to correct solutions may exist.

There are various methods used to derive domain models. [10] discuss the use of Singular Value Decomposition as well as something they call a wrapper method to find the number of latent skills in a domain. Their success was limited—both algorithms found only 2 skills, whereas expert analysis identified 7. [41] on the other hand described an approach to find the dependencies between test items in by adopting the concept of entropy in information theory. The output of the analysis is a hierarchical structure of test items.

Once domain knowledge is defined, one way in which it is represented is through the use of a q-matrix. A q-matrix is a binary matrix showing the relationship between test items and latent or underlying attributes or concepts (Birenbaum et al. 1993 in [8]). The q-matrix shown in Table 1 shows that question 1 makes use of concepts 1 through 4. Question 2 makes use of concepts 2 and 3. Question 3 makes use of concepts 3 and 4. Questions 4 and 6 only make use of one concept each—concepts 2 and 1 respectively. Question 5 makes use of no concepts within the scope of the system.

Table 1. Sample Q Matrix[8]

	Questions					
	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**
Con1	1	0	0	0	0	1
Con2	1	1	0	1	0	0
Con3	1	1	1	0	0	0
Con4	1	0	1	0	0	0

Q-matrices may be generated manually, by subject matter experts. They can also be generated automatically, by mining student responses to questions.

The work of [26] describes the use of a learning factors analysis algorithm to automatically find better student models by searching through a space of knowledge components (skills that students need to learn). The input to the learning factors analysis is a dataset with the student, the problem step, the order in which the student encounters each step, and the student's success. Input also includes a p-matrix, a binary mapping of candidate features that might influence student performance (see Table 2).

Table 2. Example of Q, P, and Q'-matrices.

Problem Step	Q		P		Q'		
	Mult	**Subt**	**Neg Result**	**Order of Op**	**Mult**	**Pos**	**Neg**
2*8-30 = 16-30	1	0	0	0	1	0	0
16-30 = -14	0	1	1	0	0	0	1
30-2*8=20-16	1	0	0	1	1	0	0
20-16=4	0	1	0	0	0	1	0

The learner factors analysis then outputs a list of q-matrices, q', ranked in order of how well they predict student data. Q'-matrices are examples of new domain models. Table 2 shows that the first step makes use of multiplication only. The second step uses subtraction but yields a negative result. The third step uses both multiplication and subtraction, and requires the student to know the order in which these operations must be performed. Finally, the fourth step makes use of subtraction only. In the q-matrix, only multiplication and subtraction are identified as necessary skills. However, the p-matrix indicates that the subtraction with the negative result and the order of operations may have an impact on student performance. The q'-matrix is the union of the skills represented in both the q- and p-matrices.

Domain models can also be expressed as networks, as seen in the work of [40]. In the contest of computer programming, the authors compute for the probabilistic distance between a student solution and a correct solution. The authors took snapshots of students' code as the student developed their programs. They then aggregated all these paths in to a network where each node is a single snapshot and an arc going from one node of the next represents transitions from one program edit to another. For each of these nodes, the authors computed for the maximum likelihood estimation transition probability from one node to every other node. This representation is useful because, in succeeding runs, of the same exercise, the system can determine whether a student's path will progress to a solution or not and can guide the student accordingly.

5. EDM TO IMPROVE PEDAGOGICAL SUPPORT

Based on their tutees' individual characteristics and needs, good human tutors decide whether they need to scaffold, to motivate, to explain, to adjust difficulty levels. Computer-based tutoring systems attempt to approximate these same skills [46].

One dimension of pedagogical support is how and when the system provides students with help. Being able to provide learners with the right type of help at the right time as well as the ways in which students use help are important research questions because the effects of help are not always intuitive. To this end, a variety of help-seeking behaviors have been under investigation for several years. Research from the early 2000s showed that up to 72% of student actions represented unproductive help-seeking behavior [2]. Students were prone to help abuse or else avoided seeking help even when it was to their advantage to do so. Students with lower abilities were found to be at a particular disadvantage: evidence suggests that they use help less discriminately (Nelson-LeGall, 1990 in [1]), implying that those who need help most are unlikely to receive it in a timely fashion.

Even understanding students' successful use of help needs to be disaggregated because success implies different things depending on the student's proficiency in the subject area [17]. High-knowledge student who succeed after just 1 or 2 hints may be having difficulties identifying salient problem features or mapping problem features to a learning principle. Low-knowledge students who succeed after 2 hints may possess skill in applying principles that they do not know or do not know well. These findings suggest that hints have to be written to help learners recognize principles when they encounter a problem.

Indeed, hints, if written well, can be effective supports for learning. [18] found that when a student makes a hint request the first time they attempt to practice a skill, they succeed at the attempt about half the time. Asking for help on the first attempt, though, correlates with future hint requests, implying that students will continue to use hints as they progress in their work.

Feedback is another type of pedagogical support. This can refer to something as simple as a notification that tells a student whether his/her answer is correct or wrong. It is more effective, though, when it provides students with details about how to improve their performance [46]. Following the principles of reinforcement learning, [24] designed Shufti, a learning environment that helps medical students master medical diagnosis. Shufti's exercises are categorized by difficulty level. Students progress from one level to another, after they accumulate a number of points. Shufti tracks student actions, current and past states, earlier feedback, and student reactions to feedback. It then calibrates the polarity (positive, encouraging, negative corrective) and timing (random, timed, after action, timed after action) of its subsequent feedback.

Still related to reinforcement learning, [32] examines optimal spacing for musical interval training. In traditional music instruction, musical relationships are taught based on their apparent difficulty level. However, the field of musical training lacked empirical data on the effectiveness of these techniques. [32]'s study showed that the typical curriculum was inefficient and that an interleaved approach promoted better learning. "Interleaving multiple pitch relationships and registers during the learning process reinforces the learned label for musical sounds across multiple contexts, promoting greater, more effective learning." [32]

Although the different studies in sections 3, 4, and 5 were categorized according to application area, the truth is that all these works inherently cut across the areas. Pedagogical models use student models and domain knowledge to fire interventions. Student models are assessed based on domain knowledge.

Consider the problem of "test size reduction," i.e. the maximization of estimation accuracy of each learner's knowledge of a concept while minimizing the number of questions a learner must answer [43]. The authors in [43] present two algorithms: a non-adaptive version, and an adaptive version. The former is appropriate when the analysis is done post-hoc and all the learners' responses are already with the instructor. The latter adaptively selects the next best question to present to each learner based on responses to date. This is an example of student models informing pedagogical models.

6. CONTEMPORARY ISSUES

Although the past decade or so has seen a rapid rise in the number and sophistication of student, domain, and pedagogical models, EDM and the artificial intelligence in education community confront a number of issues in the years ahead. In this section, we will focus on two issues in particular: the limitations of developing world context and cultural differences.

6.1 Developing world contexts

Section 2 of this paper discusses the EDM process and notes that the first step is data collection. Data collection implies deployment of software in schools, something that has already been achieved in developed countries but is not prevalent at all in the developing world. Good educational software has the potential to make significant impacts on educational achievement, especially in contexts where students are underserved [31]. However, deploying these technologies in these contexts is awash with issues.

[29] notes learning technologies from the Western world are not designed for these contexts. Learning software is designed for 1:1 use, whereas, in the developing world, it is more common for students to share devices because there is not enough hardware for each student [29] [31]. In the study by [31], the authors observed that students in Latin American countries tended to work more collaboratively than students in the United States.

When designed for mobile phones, learning software requires smart phones or personal digital assistants. Cellular phones among students in the developing world tend to be more basic and cannot support sophisticated applications. Furthermore, many students still lack skills in using mobile phones for learning. Software that relies on Internet connectivity is a poor fit for developing world classrooms. Mobile software that requires a data connection will make using it too expensive for most developing world children. Finally, [29] cites the need for localization of the software. This does not just refer to translation of the content into the local language. It also refers to the selection of appropriate icons, graphics, and other media.

In a paper with a similar theme, [36] identified five factors that made transferring Western field methods and materials to a Philippine context. As with [29], they found that the levels of technology adoption tended to be low. Students had little to no prior experience using laptops and learning software.

One factor that prevented students from gaining experience is the lack of infrastructure in the schools. Schools had few computers and the ones that they did have were in disrepair [36]. In Latin America, children did not have computers in their homes. If they used computers at all, they did so in Internet cafes [31].

In terms of impediments to data gathering, [36] found that it was essential to get the support from school administrators and teachers. If they did not buy into the study, it would be impossible to run it. Teacher support was particularly critical in Latin America [31]. Researchers noticed that, in Brazil, many teachers drove to school from as far as three hours away, hence they did not always arrive in time to teach.

Student culture also affected ease of data collection. Students tended to be reserved around observers, in part because Filipino culture tends to respect authority [36].

Finally, experimentation tends to be disrupted by the Philippine rainy season. Class suspensions because of inclement weather and flooding change not only school schedules but data collection schedules as well [36].

Wide scale use of intelligent tutors and similar software will not be possible until many if not all of these impediments (and others) are addressed. Until then, data collection will be limited to pilot studies or otherwise very limited software usage, thus limiting also the extent to which data can be mined.

6.2 Cultural issues

The second issue that confronts the EDM and artificial intelligence in education community is the growing importance of culture in the design and development of educational software. Literature suggests that culture affects the ways in which people interpret and react to their environment [11]. It is difficult to address culture in a scientific manner because culture is ill-defined and because people tend to be unaware that they are culturally interpreting information [11].

Yet it is important to move towards thinking about how our learning systems can be more culturally sensitive. Researchers have been pointing out that much of the research on human behavior and psychology hails from Western countries, and yet these nations account for only 12% of the world's population [21]. American psychological literature in particular bases its conclusions on about 5% of the world's total population [3]. [21] note that there is considerable variation among different populations in terms of visual perception, analytic reasoning, cooperation, memory, and so on, and these differences may stem from the way people have adapted to their cultures. [3] further argues that demographic contrasts in income, education and health are so stark, e.g.

- Half of the world's population lives on less than US$2 per day.
- 80% of the world's population lives on a family income of less than US$6000 per year.
- In developing countries, 1 in 5 children does not complete primary school; only half enroll in secondary school; and 17% are malnourished.
- In developing countries tertiary education is generally only for the wealthy and the education gap between genders is larger than in developed countries.

that psychological literature reflects very little of the cultural variation worldwide.

This is a concern for the artificial intelligence in education community because this community draws much of its theoretical grounding from Western psychology and educational theory. To quote [16], “We make culturally-charged decision in the design of every aspect of our technologies and these may have significant impacts on users from underrepresented populations.”

Hence, researchers from within this community are actively pushing for the study of culture in learning systems. They advocate, for example, working definitions of culture that can be operationalized in software. Culture, for example can be defined as a cognitive phenomenon that emerges at a group level [11]. Cultural normal refer to a grammar for social interactions. Cultural scripts are prototypical procedures that should be performed in specific contexts, for specific purposes. Finally stereotypes are belief structures that influence information processing.

Other researchers are already investigating the impact of culture on learning and learning-related behaviors. [34] found that students in the Philippines have significantly less off-task behavior than students in the US. However, Filipino students game the system much more. The authors suspect that this difference is culturally-founded. Filipinos put a premium on good interpersonal relationships and are respectful of authority. Hence, they are less likely to perform actions that are directly offensive or defiant of authority. Gaming the system gives the outward appearance of compliance with the learning task whereas overt off-task behavior does not.

Even within a single country, cultural differences have an impact on learning. The group of [16] noted that dialectical differences between children of color and Euro-American children had an impact on test scores. They created two versions of a learning system: one in Mainstream American English (MAE; *The creatures have no claws.*) and another in African American Vernacular English (AAVE; *The creatures don’t have no claws.*) While all participants learned, children of color posted the strongest improvements.

7. CONCLUSIONS

In order to contribute to 21st century society, people must be able to think systematically, critically, and creatively. They have to be able to consider complex solutions to complex problems. A premium is laid on competencies such as persistence, self-efficacy, openness, and teamwork [39]. Computer-based learning, intelligent tutors, and other similar technologies have the potential to foster student achievement as well as many of the other skills that we believe our young people should learn. Educational data mining offers a suite of approaches and methods to make sense of the current and future deluge of data from all these learning systems.

For EDM’s results to be interesting, though, we need to ask interesting research questions. This paper has discussed some of the issues and problems that are still worthy of investigation. Student behaviors such as wheel-spinning, WTF, and PFL may help us understand how effectively our students are using our learning systems. Student affective states and stress levels carry implications about how healthy, happy, and academically successful our students are. Finding latent skills within a domain was and will continue to be a relevant problem, because it identifying a skill is the first step to assessing it accurately. Studies on the timing and content of hints and feedback help us tune pedagogical support according to individual differences.

Indeed, artificial intelligence in education and EDM have much to contribute to education in general and to underserved communities in particular. However, for these technologies to spread and become successful in developing countries, a number of obstacles have to be overcome. Those that are external to the systems themselves include increased access to hardware, software, and connectivity; teacher, student, and administrator preparedness; and greater technology literacy. Factors that are internal to these educational systems include the provision for differences in language and media, and pedagogical methods.

EDM is a rich field, full of open questions. For researchers with an interest in both education and computer science, it is an ideal stomping ground as it draws heavily from both areas. For those of us who practice in the developing world in particular, EDM presents even more meaty challenges that require interdisciplinary approaches as well as deep, creative thinking to address. There is much more work to be done.

8. ACKNOWLEDGMENTS

The author thanks Regina Ira Antonette M. Geli, Gabriel Jose G. Vitug, and the Ateneo de Manila University’s Department of Information Systems and Computer Science.

9. REFERENCES

[1] Aleven V. and Koedinger, K. R. 2001. Investigations into help seeing and learning with a Cognitive Tutor. In *Workshop on Help Provision and Help Seeking in Interactive Learning Environments*, (San Antonio, TX, 19-23 May 2001).

[2] Aleven, V., McLaren, B., Roll, I., and Koedinger, K. 2004. Toward tutoring help seeing: Applying cognitive modeling to meta-cognitive skills. In the *Proceedings of the Seventh International Conference on Intelligent Tutoring Systems*, (Maceio, Brazil, 30 August-3 September, 2004).

[3] Arnett, J. J. 2008. The neglected 95%: Why American psychology needs to become less American. *American Psychologists,* 63, 7, 602-614.

[4] Baker, R.S.J.d. 2007. Modeling and understanding students' off-task behavior in intelligent tutoring systems. In *Proceedings of ACM CHI 2007: Computer-Human Interaction*, (San Jose, CA, 28 April-3 May, 2007), 1059-1068.

[5] Baker, R.S., Corbett, A.T., Koedinger, K.R., and Wagner, A.Z. Off-task behavior in the cognitive tutor classroom: When students "game the system". In *Proceedings of ACM CHI 2004*, (Vienna, Austria, 24-29 April 2004), 383-390.

[6] Baker, R.S.J.d. and de Carvalho, A. M. J. A. 2008. Labeling behavior faster and more precisely with text replays. In *Proceedings of the 1st International Conference on Educational Data Mining*, (Montreal, Quebec, 20-21 June 2008), 38-47.

[7] Baker, R. S. J. D., and Yacef, K. 2009. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1, 1, 3-17.

[8] Barnes, T. 2005. The q-matrix method: Mining student response data for knowledge. In *2005 Educational Data Mining Workshop*, (Pittsburgh, PA, 9-13 July 2005).

[9] Beck, J. E. and Gong, Y. 2013. Wheel-spinning: Students who fail to master a skill. In Lane, H. C., Yacef, K. Mostow, J., and Pavlik P. (eds). *AIED 2013, LNAI 7926*, Springer-Verlag, Berlin Heidelberg, 431-440.

[10] Beheshti, B., Desmarais, M., and Naceur, R. 2012. Methods to find the number of latent skills. In In Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., and Stamper, J. (eds). *Proceedings of the 5th International Conference on Educational Data Mining*, (Chania, Greece, 19-21 June 2012), International Educational Data Mining Society, 81-86.

[11] Blanchard, E. G., Karanasios, S., and Dimitrova, V. 2013. A conceptual model of intercultural communication: Challenges, development methods and achievements. In *The 4th International Workshop on Culturally-Aware Tutoring Systems (CATS 2013)* (Memphis, TN, 10 July 2013), 1-10.

[12] Chin, D. B., Dohmen, I. M., Cheng, B. H., Oppezzo, M. A., Chase, C. C., and Schwartz, D. L. 2010. Preparing students for future learning with Teachable Agents. *Educational Technology Research and Development*, 58, 6, 649-669.

[13] D'Mello, S. K., Chipman, P., and Graesser, A. C. 2007. Posture as a predictor of learner's affective engagement. In McNamara D. S. and Trafton J. G. (eds.). *Proceedings of the 29th Annual Cognitive Science Society* (Nashville, TN, 1-4 August 2007), 905-910

[14] D'Mello, S. K., Jackson, G. T., Craig, S. D., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., el Kaliouby, R., Picard, R., and Graesser, A. C. 2008. AutoTutor detects and responds to learners affective and cognitive states. *Workshop on Emotional and Cognitive Issues in ITS* (Montreal, Canada, 23-27 June 2008).

[15] Deziel, M., Olawo, D., Truchon, L., and Golab, L. 2013. Analyzing the mental health of engineering students using classification and regression. In D'Mello, S. K., Calvo, R. A., and Olney A. (eds). *Proceedings of the 6th International Conference on Educational Data Mining* (Memphis, TN, July 6-9, 2013). International Educational Data Mining Society, 228-231.

[16] Finkelstein, S., Yarzebinski, E., Vaughn, C., Ogan, A., and Cassell, J. 2013. The effects of culturally congruent educational technologies on student achievement. In Lane, H. C., Yacef, K. Mostow, J., and Pavlik P. (eds). *AIED 2013, LNAI 7926*, Springer-Verlag, Berlin Heidelberg, 493-502.

[17] Goldin, I., Koedinger, K. R., & Aleven, V. 2012. Learner differences in hint processing. In Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., and Stamper, J. (eds). *Proceedings of the 5th International Conference on Educational Data Mining*, (Chania, Greece, 19-21 June 2012), International Educational Data Mining Society, 73-80.

[18] Goldin, I., Koedinger, K. R., and Aleven, V. 2013. Hints: You can't just have one. In D'Mello, S. K., Calvo, R. A., and Olney A. (eds). *Proceedings of the 6th International Conference on Educational Data Mining* (Memphis, TN, July 6-9, 2013). International Educational Data Mining Society, 232-235.

[19] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten , I. H. 2009. The WEKA data Mining software: An update. *SIGKDD Explorations*, 11, 1, 10-18.

[20] Harpstead, E., MacLellan, C. J., Koedinger, K. R., Aleven, V., Dow, S. P., and Myers, B. A. 2013. Investigating the solution space of an open-ended educational game using conceptual feature extraction. In D'Mello, S. K., Calvo, R. A., and Olney A. (eds). *Proceedings of the 6th International Conference on Educational Data Mining* (Memphis, TN, July 6-9, 2013). International Educational Data Mining Society, 51-58

[21] Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*, (1 July 2010), 29.

[22] Hershkovitz, A., Baker, R. S. J. d., Gowda, S. M. and Corbett, A. T. 2013. Predicting future learning using quantitative analysis of moment-by-moment learning. In D'Mello, S. K., Calvo, R. A., and Olney A. (eds). *Proceedings of the 6th International Conference on - Educational Data Mining* (Memphis, TN, July 6-9, 2013). International Educational Data Mining Society,74-81.

[23] International Educational Data Mining Society. No date. International Educational Data Mining Society website. Accessed 20 August 2013 from http://www.educationaldatamining.org/

[24] Johnson, S. and Zaiane, O. 2012. Deciding on feedback polarity and timing. In Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., and Stamper, J. (eds). *Proceedings of the 5th International Conference on Educational Data Mining*, (Chania, Greece, 19-21 June 2012), International Educational Data Mining Society, 220-221.

[25] Kocielnick, R., Pechenizkiy, M., and Sidorova, N. 2012. Stress analytics in education. In Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., and Stamper, J. (eds). *Proceedings of the 5th International Conference on Educational Data Mining*, (Chania, Greece, 19-21 June 2012), International Educational Data Mining Society, 236-237.

[26] Koedinger, K. R., McLaughlin, E. A., and Stamper, J. C. 2012. Automated student model improvement. In Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., and Stamper, J. (eds). *Proceedings of the 5th International Conference on Educational Data Mining*, (Chania, Greece, 19-21 June 2012), International Educational Data Mining Society, 17-24

[27] Lee, D. M., Rodrigo, M. M. T. R., Baker, Ryan S. J. D., Sugay, J. O., & Coronel, A. (2011). Exploring the relationship between novice programmer confusion and achievement. In D'Mello, S. and Graesser, A. (eds). *ACII 2011, Part I, LNCS 6974*, Springer-Verlag, Berlin Heidelberg, 175-184,

[28] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M. and Euler, T. 2006. YALE: Rapid prototyping for complex data mining tasks. In *Proc. of the 12th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining (KDD 2006)*, (Philadelphia, PA, 20-23 August 2006), 935-940.

[29] Nye, B. D. 2013. ITS and the digital divide: Trends, challenges, and opportunities. In Lane, H. C., Yacef, K.

Mostow, J., and Pavlik P. (eds). *AIED 2013, LNAI 7926*, Springer-Verlag, Berlin Heidelberg, 503-511.

[30] Obsivac, T., Popelinsky, L., Bayer, J., Geryk, J., and Bydzovska, H. 2012. Predicting drop-out from social behavior of students. In Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., and Stamper, J. (eds). *Proceedings of the 5th International Conference on Educational Data Mining*, (Chania, Greece, 19-21 June 2012), International Educational Data Mining Society, 103-109.

[31] Ogan, A., Walker, E., Baker, R.S.J.d., de Carvalho, A., Laurentino, T., Rebolledo-Mendez, G., and Castro, M.J. 2012. Collaboration in Cognitive Tutor use in Latin America: Field study and design recommendations. *Proceedings of ACM SIGCHI: Computer-Human Interaction*, (Austin, TX, 5-10 May 2012), 1381-1390.

[32] Pavlik, P. I. Jr., Hua, H., Williams, J., and Bidelman, G. 2013. Modeling and optimizing forgetting and spacing effects during musical interval training. In D'Mello, S. K., Calvo, R. A., and Olney A. (eds). *Proceedings of the 6th International Conference on Educational Data Mining* (Memphis, TN, July 6-9, 2013). International Educational Data Mining Society, 145-152.

[33] Pittsburgh Science of Learning Center. (2008). Robust learning. Accessed 19 August 2013 from http://learnlab.org/research/wiki/index.php/Robust_learning

[34] Rodrigo, M.M.T., Baker, R.S.J.d., Agapito, J., Nabos, J., Repalam, M.C., Reyes, S.S. (2010) Comparing disengaged behavior within a Cognitive Tutor in the USA and Philippines. In *Proceedings of the 10th Annual Conference on Intelligent Tutoring Systems* (14-18 June 2010, Pittsburgh, PA), 263-265.

[35] Rodrigo, M. M. T., Baker, R. S. J. d., McLaren, B. M., Jayme, A., & Dy, T. T. (2012) Development of a workbench to address the educational data mining bottleneck. In Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., and Stamper, J. (eds). *Proceedings of the 5th International Conference on Educational Data Mining*, (Chania, Greece, 19-21 June 2012), International Educational Data Mining Society, 152-155.

[36] Rodrigo, M. M. T., Sugay, J. O., Agapito, J., and Reyes, S. Jr. To appear. Challenges to transferring western field research materials and methods to a developing world context. *Research and Practice in Technology Enhanced Learning*.

[37] Romero, C. and Ventura, S. 2010. Educational data mining: A review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40, 6, 601-618.

[38] San Pedro, M. O. C. Z., Rodrigo, M. M. T , and Baker, R. S. J D. (2011). The relationship between carelessness and affect in a cognitive tutor. In D'Mello, S. and Graesser, A. (eds). *ACII 2011, Part I, LNCS 6974*, Springer-Verlag, Berlin Heidelberg, 306-315.

[39] Shute, V. and Ventura, M. 2013. *Stealth Assessment: Measuring and Supporting Learning in Video Games*. Cambridge, MA: MIT Press.

[40] Sudol, L. A., Rivers, K., and Harris, T. K. 2012. Calculating probabilistic distance to solution in a complex problem solving domain. In Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., and Stamper, J. (eds). *Proceedings of the 5th International Conference on Educational Data Mining*, (Chania, Greece, 19-21 June 2012), International Educational Data Mining Society, 144-147.

[41] Sun, X. 2012. Finding dependent test items: An information theory based approach. In Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., and Stamper, J. (eds). *Proceedings of the 5th International Conference on Educational Data Mining*, (Chania, Greece, 19-21 June 2012), International Educational Data Mining Society, 222-223.

[42] Tabanao, E., Rodrigo, M. M. T., and Jadud, M. C. 2008. Identifying at-risk novice programmers through the analysis of online protocols. In *Philippine Computing Society Congress 2008* (UP Diliman Feb 23-24, 2008).

[43] Vats, D., Studer, C., Lan, A. S., Carin, L., and Baraniuk, R. 2013. Test-size reduction for concept estimation. In D'Mello, S. K., Calvo, R. A., and Olney A. (eds). *Proceedings of the 6th International Conference on - Educational Data Mining* (Memphis, TN, July 6-9, 2013). International Educational Data Mining Society, 292-295

[44] Vega, B., Feng, S., Lehman, B., Graesser, A., and D'Mello, S. 2013. Reading into the text: Investigating the influence of text complexity on cognitive engagement. In D'Mello, S. K., Calvo, R. A., and Olney A. (eds). *Proceedings of the 6th International Conference on Educational Data Mining* (Memphis, TN, July 6-9, 2013). International Educational Data Mining Society, 296-299

[45] Wixon, M., d Baker, R. S., Gobert, J. D., Ocumpaugh, J., & Bachmann, M. 2012. WTF? Detecting students who are conducting inquiry without thinking fastidiously. In Masthoff, J., Mobasher, Ba., Desmerais, M. C., and Nkambou, R. (eds). *User Modeling, Adaptation, and Personalization*. Springer Berlin Heidelberg, 286-296.

[46] Woolf, B. 2009. *Building Intelligent and Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning*. Burlington, MA: Morgan-Kaufmann.