Enabling people to harness and control EDM for lifelong, life-wide learning

Judy Kay
Human Centred Technology Research Cluster
The University of Sydney
Australia
judy.kay@sydney.edu.au

ABSTRACT

There has been an explosion of digital *sensors* of learning. Some are in bespoke learning applications. But many more are in every aspect of our lives. This paper takes a *human-centred view of EDM* for lifelong, life-wide learning, where EDM enables people to harness that data. This view includes formal education. But it goes beyond that, to the complex, multi-faceted, ill-defined and long-term learning needed to work towards the most important goals in our broader lives. Some of the most important of these goals are lifelong and are critical for aspects as diverse as health and wellness, responsible citizenship or working effectively with other people.

The paper begins by considering the stakeholders for EDM. It then presents three longitudinal case studies from my research on supporting learning of complex skills. Drawing on these, the analysis that follows presents lessons learnt and a wish list and vision for future EDM directions towards human-centred EDM.

Keywords

Lifelong learning, life-wide learning, educational data mining, user control, student modelling, learner modelling, personal data, privacy, provenance, user control, business analytics, learning analytics, personal informatics.

1. INTRODUCTION

Technology that can support learning is now pervasive. This has created an explosion of opportunities for life-wide and lifelong learning. One important part of that is formal learning and that has dominated decades of AIED and ITS research. Even this happens in diverse contexts, from physical classrooms to nearly every other place. But it goes well beyond formal learning. The pervasiveness of technology means that we have a rich digital ecosystems. This includes devices that we wear and carry as well as those located and embedded in our environments. From an EDM perspective, this provides many streams of digital footprints that might be harvested to support learning. The potential these offer has created the EDM community [9] as well as many others, such as learning analytics [53] business analytics [37] and personal informatics [38].

Figure 1 illustrates one way to think about EDM as the transformer of *sensor data* into the *learner models* that can support learning. The top of the figure shows a person called Mykola[1] who uses many digital tools and devices. For example, Mykola may interact with a collection of maths tutoring applications, perhaps over many months or even years, using various desktops, tablets and smart phones. Viewing each of these applications as sensors that collect data about Mykola, the figure distinguishes two classes of EDM transformations. At the left are the transformations that feed into the learner model for this *individual learner*. At the right are those that aggregate the data from *collections of learners*. The thinner grey lines, numbered 1 to 4, are classic EDM, interpreting raw sensor data to add it to a learner model of the individual learner (1) and then reasoning on it (2) and the corresponding actions for aggregate models for a collection of learners (3 and 4). Of course the whole point of EDM is to create learner models that are to be used. So the lines 1 and 3 are bidirectional. This reflects the times that the learner model serves information to an application. Creating individual and aggregate learner models may draw on diverse methods and infer aspects that include the learner's knowledge, as well as other critical factors, such as motivation [36].

Over the many years of Mykola's maths education, current EDM approaches produce many learner models. This is what we have seen in over 25 years of AIED research. The learner model has typically been just one part of just one application. It has often been short-lived, perhaps just the single session of a research study. That is changing in two important ways. Learner models are becoming *first class citizens* and they are becoming *long-term*.

One driver for learner models to transition to first class citizen-ship follows from our understanding of their direct value for learners when there is a *user interface* onto them. This has long been called an Open Learner Model [12] and, more recently, similar interfaces are called learning analytics and dashboards [53]. At a quite different level, learner mod-

[1] Thanks to Mykola Pechenizkiy, President of International Educational Data Mining Society for agreeing to be everyman in this image.

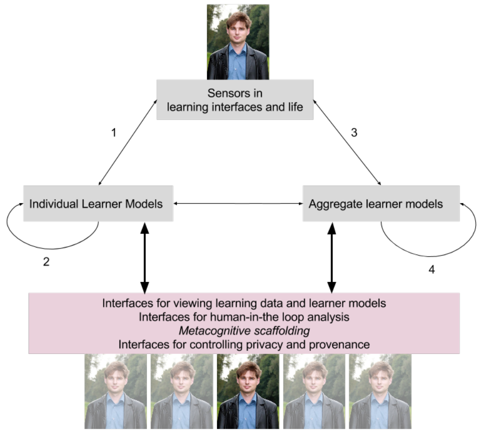

Figure 1: EDM as transformer: from sensor data to learner models as first class citizens

els should have independent standing so that an *application-program interface, API* enables multiple applications to re-use the same learner model. This is a quite different form of openness that is potentially valuable for long term learning. Both these aspects of openness should be designed from a human-centred foundation.

Consider Mykola's broader learning, in areas such as health and wellness. He might use various tracking devices and coaching applications over the long term. The trackers can capture data about aspects like his food intake, physical activity, weight, rest pulse and stress. All of these could serve as sensors for a steadily growing and rich learner model. His interactions with various coaching systems could provide data to model aspects like his metacognitive skills, as he uses each coach to set goals, self-reflect and self-monitor. For life-long goals for good health, Mykola's learner model needs to be kept over the long term, accumulating and mining data from many sensors. Equally, a long-term model representing stable traits, such personality, could be re-used by multiple personalised learning applications [15].

Mykola's sensor data and associated learner model is his *personal data*. People typically want to be able to control such data and its use, as reflected in privacy legislation [34]. User concerns and needs around privacy are important and complex. EDM's learner models may represent aspects that are quite sensitive. These include stable attributes like personality as well ephemeral, sensed attributes such as attention [17]. The aggregated learner model at the right side of Figure 1 has been at the core of EDM research. It is well understood that Mykola's data in these models needs to be treated with care. This is typically achieved with forms of de-identification so that the data actually kept is divorced from the user.

Let us now turn to the lower part of the figure. This depicts several people, in several roles, interacting with suitable interfaces onto these learner models. Some important stakeholders include:

1. builders of ITS/AIED systems;
2. the individual learner;
3. the classroom teacher, parent, mentor, peer learner or other supporter for the learner;
4. administrators, governments and learning "accountants".
5. learning scientists;

In the box above the people, the figure shows *four classes of user interfaces* onto learner models. The first, already mentioned, is the Open Learner Model or learning dashboard. Appropriate forms of these are useful for each stakeholder group. The system builders need them to help understand and debug their systems. Appropriately designed interfaces onto individual and aggregate learner models can be used as part of the *learning process* by the next three groups. The individual learner and their support team are particularly concerned with the individual learner model, although aggregate models can put this in context and help these stakeholders interpret it. The Learning Analytics community has introduced such interfaces for stakeholders at the institutional level with dashboards for administrators. These have much in common with dashboards used in many fields [56, 21]. Learning scientists have a quite different perspective. With their psychology focus, they can use interfaces onto aggregate models to display how people learn, with the exemplar being an increasingly refined understanding of forgetting curves [6].

The figure also highlights other roles. There is an under-explored EDM role for human-in-the-loop systems. Certainly system builders do some of this when debugging. OLMs and dashboards have provided simple interaction to support this role for the learning process stakeholders. The third class of interface shown in the figure is the metacognitive scaffolding that is needed for individual learners because self-reflection, self-monitoring and planning are hard for many people. Finally, the figure show a class of interfaces that enable people to control their own data and provenance. I return to these after the case studies in the next section.

2. CASE STUDIES

This section overviews three strands of my research. The first aimed to help students learn *group work skills*. These are complex and they require many skills. For example, effective group work relies on communication, including listening skills. Long term group work demands considerable self-regulation so that the individual and the group can plan, monitor and reflect on progress. It involves learning leadership skills, which is important for effective teams [51] The second case study is in *computer supported collaborative learning*, a common feature of classrooms. The third concerns design and management of the curriculum for *long term learning* of generic skills across a university degree. All three cases take a human-centred approach to tackling some of the many parts of the complex puzzle of EDM to support lifelong and life-wide learning.

Long term asynchronous group work

Team work skills are important in many contexts. They are so common in the creation of software that computing degrees have a capstone software engineering project. Typically, students work in small teams over a semester to create a substantial software artefact that meets a clients' needs. It is common practice for such teams to use a platform that supports the management of the code and other files as well as the team processes. (Just a few of the many platforms include GitHub[2], trac[3] and BitBucket[4]). These platforms provide rich sensor data about the group behaviour. This case study involved use of trac. We explored how to harness the data about each team member's use of its core tools, a wiki, the ticket system (called issue tracker in other systems) and the version control system (svn in our case). In terms of Figure 1, these three media were the sensors and we wanted to build a model of each individual in the team and make it available to them and their teachers. We wanted to transform the huge amount of data from the sensors (thousands of actions over the three months of the semester) into an OLM.

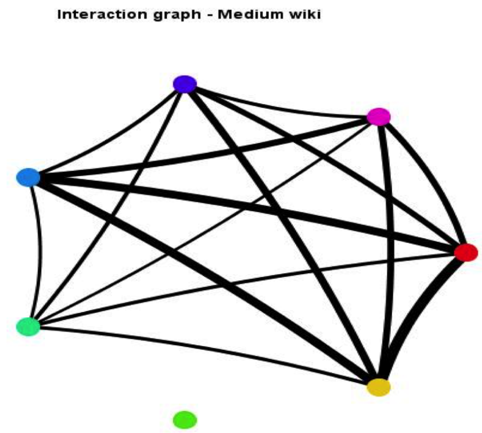

Figure 2: Simple EDM, showing interaction between team members

Figures 2, 3 and 4 illustrate three key stages in our work to create effective OLM interfaces. We began by creating several presentations that were in the spirit of the interaction diagram shown in Figure 2. These were inspired by social translucence visualisations [19]. Each circle represents one team member (the figure has anonymised the display, removing the names that were near each dot). So, for example one user is represented by the pink dot at 2 o'clock. The lines between the dots indicate how much that pair of people interacted on the wiki. This is based on a measure of each person's contributions to the same page. For example, students were advised to create a page for their weekly meeting minutes, with each team member reviewing this and adding a comment, to indicate agreement or identifying problems. If the whole team did this, we would see connections between every pair. In Figures 2, it is striking that the green dot at 7 o'clock is not connected to any others. Worse yet, the green dot represents the person who was supposed to be the team manager. This diagram highlights a potential problem! In practice, these diagrams proved valuable because problems like this became apparent. So long as they were available early enough, and used appropriately in the teaching (not for assessment), they facilitated valuable discussions within the team and with their teachers.

The interaction diagrams and activity diagrams, one for each of the media (wiki, tickets and svn) were our starting point. These were useful for a single point in time. Our next step was to create an OLM showing the long term model. We particularly wanted to be able to see changes. For example, we needed to see what happened after we had identified and tried to remediate a team problem like that of the manager in Figure 2. Our next step was the Wattle Tree visualisation [31], like the example shown in Figure 3[5]. Each green vertical line is one team member. (There are 6 in the figure.) Each day's wiki activity appears as a yellow circle to the left of the person's line. Each svn action is an orange circle to the right of their line. The size of the circle is a logarithmic function of the count of actions or code lines committed. The green lines are ticket actions. Defining a new task is a dark green line on the left. Completing an allocated task gives the light green ones at the right. This figure shows a high functioning team, with very active leadership by the person on the second line. The rightmost user is clearly less active. The almost barren area towards the top was the semester break. This group was doing well enough that they took a break at that time. This OLM proved useful in the hands of a skillful teacher and team leader. But it was fragile in that weak teams did not use it very well without considerable mentoring. It was also flawed as an interface; the idea of the wattle tree was cute, but it was somewhat forced, difficult to extend and the metaphor does not match the learning goals well.

Our next step was the Narcissus interface [54], shown in Figure 4. Now each user is a block (5 of them in them in the figure). As the legend at the upper right indicates, these show the three media as squares, coloured purple (wiki), blue (svn) and green (tickets). The brighter each square, the more that user did that day on that medium. The bottom of each block shows a cumulative picture for that individual, compared to the team average (grey). Narcissus used quite simple EDM measures, pure counts of actions. But it added scrutability, with the user able to configure how the simple sensor counts map to the colour intensity.

Even more valuable, each cell is interactive. In the figure, the user has clicked a blue (svn) cell and the details are available at the right. This lists details of, and links to, all the changesets to the code checked in by that user on that day. So we now transform the OLM into a navigation tool. This was invaluable for tracking down the details as needed.

A parallel stream of this work explored more sophisticated EDM methods [32, 50] for a cohort of 43 students in 7 groups. The sensor data was 1.6 megabytes in mySQL format, with over 15000 events. With a combination of ex-

[2]https://github.com/
[3]https://trac.edgewall.org/
[4]https://bitbucket.org/
[5]The Australian Wattle Tree has small round balls for flowers and they appear in clusters as in the figure.

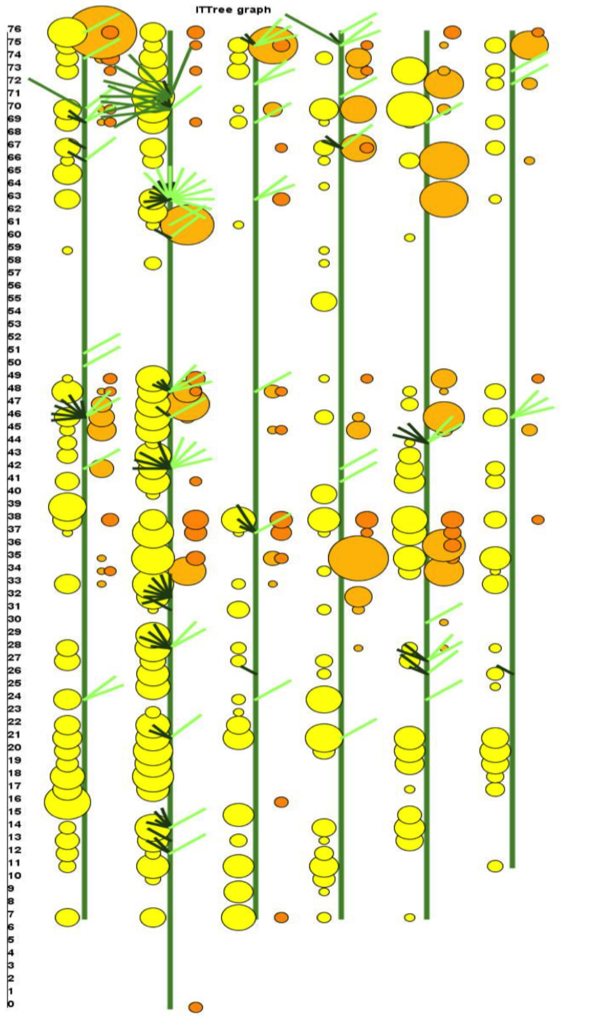

Figure 3: Simple EDM, in a Wattle Tree, showing daily actions by each user on each medium

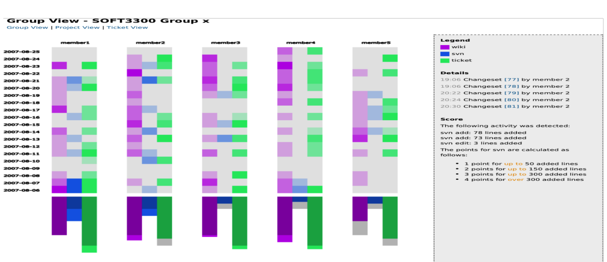

Figure 4: Simple EDM, with Narcissus, a much better and interactive interface onto daily actions by each user on each medium

ploratory and theory-driven approaches, we discovered useful measures of aspects of group and individual behaviour. For example, we identified patterns that are associated with effective leadership. Whether a group's nominated leader exhibited these was predictive of the effectiveness of the leadership. For example, one nominated leader failed to exhibit those patterns, but rather had the pattern of a pure developer, with others having some of the leadership patterns. In fact, the leader was trying to do much of the programming alone, neglecting group management. Other team members tried to fill the leadership void. Promisingly, the group-role patterns were established early in the semester. This is important for remediation. These results also gave the teachers guidance on aspects to teach students and indications of how to better guide them. However, we never translated this more complex approach into an OLM, like Narcissus.

So what did we learn from creating and using the OLMs? They relied on very simple mining of the raw sensor data, just counting actions. And that worked well. In the case of the wiki, actions on the same page were interpreted to model interaction and we used even simpler counts of all ticket actions and lines of code committed. With Narcissus, we transformed this simple data mining into an interactive OLM that supported navigation of the complex information space and this was heavily used in meetings between the tutor and each team and in helping groups with problems. We contributed the Narcissus code to the open source trac project's Track Hacks and used it for several years (until we stopped using trac). We know anecdotally that our students asked for Narcissus to be installed for use in other classes using trac. But there ends it deployment. We are not aware of any team programming environment that provides comparable facilities to model group health. We fell short of any deployment of the complex data mining approaches. It would be excellent to see them integrated into a future version of Narcissus-like tools. That will require design of suitable interfaces. It will be challenging to do this and maintain the Narcissus philosophy of user control. In summary, this sequence of work explored how to harness the digital footprints of teams using trac-like tools and demonstrated the power of simple measures in OLMs and the promise of more sophisticated mining.

Collocated collaborative learning

This case study continues the exploration of how to harness data about groups to inform learning. But we now move to a collocated context (rather than the above asynchronous long term collaboration). This work was inspired by the potential of interactive tabletops to provide a new way to support small group work. This is because small groups commonly work at a table, and tabletops offer a shared interactive canvas. From an EDM perspective, tabletops offer new sensors of collocated collaboration.

Figure 5 shows users in one of our lab studies. This group of three students is doing a collaborative learning task. Each of them had previously worked individually to create a concept map summarising their understanding of provided learning material. Then they came together to create a joint map. We had considerable sensor data: the tabletop touches, so we could determine which learner did each action; the speech captured by the microphone that is visible at the front mid-

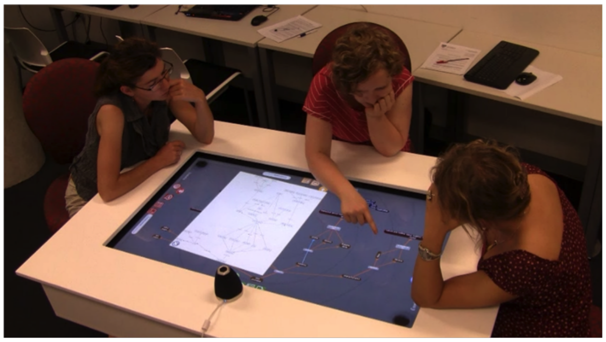

Figure 5: Lab study of small group creating a concept map collaboratively

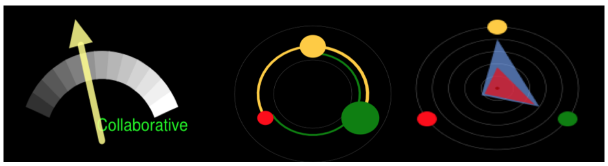

Figure 6: Group health sophisticated dashboard

dle of the table; the initial individual maps, the final group map and all the intermediate states.

To harness all that data, we used data mining to build a model of the quality of the collaboration [45, 44]. Figure 6 shows a set of visualisations for these. The dial at the left summarises how well the group is collaborating [41]. This is left of centre, indicating somewhat poor collaboration. The transformation of sensor data to produce this single measure are complex. They use a model learnt from measures derived from video analysis and the automatically collected sensor data. The other two figures are versions of the interaction and activity diagrams from our earlier work. The middle one shows interactions, such as one user working with interface elements created by another user. The right one shows the level of activity for each user in terms of tabletop touches and speech with both of these highest for the yellow user near 12 o'clock. The left collaboration dial is inscrutable; we never attempted to create an explanation or justification of it for use by a teacher. However, the two simpler diagrams do help as they show elements that use some of the same sensors as that measure.

The tabletop lab work seemed promising. So we then explored how to move it to an authentic classroom, such as shown in Figure 7 [43, 42]. Under the real time and curriculum pressures of a classroom, the teacher wanted to harness the tabletops and an OLM to track each class group's progress on the *content learning objectives* for the class, rather than our inscrutable models of collaboration. This led to the design of the dashboard shown in Figure 8 [41]. The tabletop activity data was used to create models of each group. The OLM for these can be seen at the right of the dashboard. This has a set of bar chars, each group in a different colour, with one bar for each student. This shows a count of the number of propositions that student has created. The darker part of the bar shows propositions matching ones the teacher had defined. Other propositions that students created are indicated in the light parts of each bar. Under the pressure of classroom teaching, even such simple OLMs were more effective when there were alerts to help the teacher see which group seemed to most need attention [42].

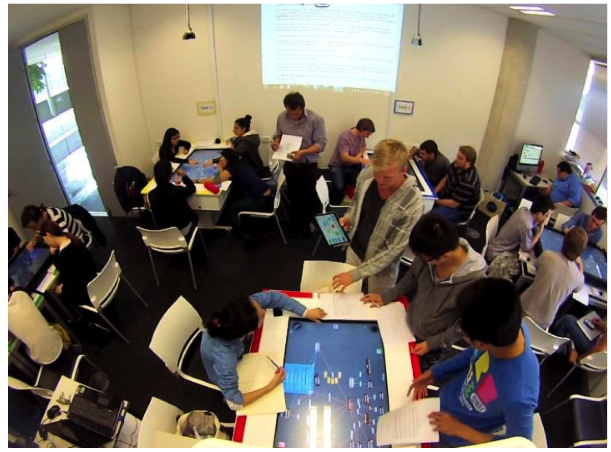

Figure 7: Classroom of interactive tabletops

Figure 8: Simple dashboard

So what did we learn? We showed that sophisticated EDM resulted in a model of how collaborative a group was. This black box result has the potential to be valuable information for a classroom teacher. When they need to spend time with one group, they cannot be closely following all of others. So this measure might be helpful when they finish working with one group and need to decide which group most needs their attention. In practice, when we moved to a real classroom, the teachers saw a greater need for simpler measures about the learning activity. The EDM could both provide these in an OLM and drive the alerts to help them quickly decide how to split their time between the groups.

Modelling generic skills over 3-5 years

This case study was driven by two key demands. The first was to improve the design of a university curriculum, with

Figure 9: Aggregate long term learning model of generic skills and potential sensors

a particular focus on systematic building of generic skills, such as group work and communication. A second driver was the needs of accreditation of engineering and IT degree programmes. A human-centred perspective was essential. This is because a whole university degree involves hundreds of academics. We needed all of them to contribute to the project by using our system effectively. More importantly, they control the actual classroom teaching.

We created CUSP [25]. This supports definition of the ontology for the generic skills in the whole of a university degree that runs from 3 to 5 years. CUSP also supports mapping the core ontology, from the university (and faculty) learning outcomes, to each of multiple accreditation standards. It provides views of the curriculum, in terms of these ontologies. CUSP has been in use for several years at the University of Sydney. Indeed, although it was initially designed for use by academics to manage the curriculum design and accreditation processes, it has been repurposed for use by students. This means it used daily, with thousands of students relying on it in Sydney[6].

From the perspective of Figure 1, CUSP provides the *ontology* for an aggregate learner model. The interfaces enable an individual teacher to see their subject in terms of this ontology. The academics responsible for a whole degree have interfaces that show the big picture. For example, Figure 9 shows the seven broad classes of generic skills that are assessed at each of the five levels. This defines the *intended* learning outcomes and level of each across the degree.

Key to the success of CUSP was its mapping of the generic skill ontologies, from the institutional one, as in the figure, to several used by external accreditation bodies. This was based on a very simple approach. There is a mapping only between these ontologies. Each subject co-ordinator maps their detailed learning objectives against the institutional ones. The approach assumes this will correctly translate to the various accreditation ontologies. In practice, our evaluations indicated this worked well [22, 25]. When we examined the small proportion of errors, most were due to the lecturers incorrectly coding against the institutional ontology.

CUSP's ontology is hand crafted centrally with individual teachers mapping their subject against it. This worked well for generic attributes. We explored how to take this approach further, to deal with subject matter content. We did this in the context of the Programming Fundamentals sequence of subjects in Computer Science. These are intended to build and develop skills and knowledge, with students reaching higher levels of mastery over several subjects and several years. The result was ProGoSs [23]. This enabled teachers to map either their subject description of their exam against a standard curriculum, such as ACM 2013. Our educators found exam mapping easier. ProGoSs provided a framework, or ontology, for a learner model. Notably, teachers needed to augment the standard concepts. This is partly because standard curricula need to be framed in general terms. By contrast, an actual subject learning objectives are linked to aspects like the particular programming language, as well as very fine-grained or new concepts. The actual exam became the sensor. The marks for each question were added to the individual learner models. This detailed learner modelling has considerable potential for EDM or learner analytics if combined with other information about the individual learner. ProGoss was tested over multiple subjects in multiple institutions. But it has not been in broader use.

A similar approach is being incorporated into tools like Gradescope[7]. The demands of accreditation, linked to the process of grading exams, seem potential drivers. The human factors will be critical for real world deployment. These include interfaces that make it easy to create the learner model and it will rely of the value of the learner models for the stakeholders.

In both CUSP and ProGoSs, the main stakeholders were the custodians of the curriculum, both at the level of the degree and the individual subject. CUSP has been repurposed for student use because its curriculum role meant that it encoded considerable detail of each subject and how it fits into the degree programme. ProGoSs, as a research prototype, foreshadows the potential for tracking fine-grained learning progression.

3. LESSONS LEARNT, FEARS, VISION

The stakeholders identified in the introduction all share the need for EDM to provide evidence-based insights about learners. But there are important differences. In terms of Figure 1, the individual learner model and the aggregate models have different roles. By contrast, the learning scientists are concerned with the aggregate models only. For all the stakeholders aiding the individual learner, that learner's individual learner model is key. But these stakeholders also need an aggregate model, to make sense of the individual one. For example, for a learner to judge their progress, they may want to compare the learning model against those of successful students – where some students want to compare against bare pass students and others against the high achievers.

The builders of ITS/AIED systems need EDM to drive personalised learning. Learning scientists want to understand learning more broadly. Both these roles reflect core goals of AIED/ITS communities, from their foundations. Both are well represented in current EDM research and in deployed AIED systems.

[6]and thousands more use a commercial version called U-Improve http://www.u-planner.com/products/u-improve

[7]https://gradescope.com/

But EDM can also give a quite different level of benefit. One of these is the OLM or dashboard interface into a learner model. As in Figure 1, EDM can be seen as any process that transforms data from sensor of learning into a learner model, where this is any representation of learners that is intended to support learning. Then a suitable OLM (or dashboard) has the potential to serve many purposes [12], listed below an illustrated in terms of Narcisses.

- Improving the accuracy of the model − students could set the counts that triggered colour changes;
- Supporting metacognitive processes of planning, monitoring and reflection, with evidence supporting self-awareness [52] − the group's tutor played a key role in useing the OLM to support individuals and the group.
- Facilitating collaboration or competition − the main teachers used the OLM in meetings with the managers from each group so each manager could use the OLM to share their current challenges and actions;
- Facilitating navigation − with the OLM linking to each sensor element;
- Respecting the learner's right to access and control their personal data, and their trust in the learner model − all sensor data was accessible and controlled by the students;
- Using the learner model as an assessment of the learner − this was purely formative and was far too simplistic to use directly to assess the learning objectives.

A human-centred view frames EDM in terms of the needs of stakeholders. One such need is for interfaces onto the learner models. This should also impact the design of sensors, the way that they are processed, the design of learner models, the ways that sensors contribute to them and the ways that people can control and harness their own data. Taking this perspective, what are the key lessons from the case studies?

Embrace simplicity, with care

Baker has suggested that we need "stupid tutoring systems, and intelligent humans" [8] where this is possible because of rich data-driven teaching system. This matches our experiences. A human-centred perspective favours at least taking serious account of the simplest approaches to the design of each element of the EDM processes, as in Figure 1. All three case studies indicated simple models were valuable and could be deployed.

EDM can point to the need create new sensors. For example, we concluded that we needed to link the tabletop touches to the individual who did the touch. Tabletop hardware generally does not support this (one exception being DiamondTouch). To create this sensor, we integrated a Kinect with the tabletop to provide a hardware independent way. When EDM uses systems that were not designed for it, we may need new sensors to support downstream simplicity of the EDM. There is a need for this in MOOCs. These appear to have been created without EDM as a core design driver. So it is hard to link each sensor, for example video activity and MCQ responses, to learning objectives for learner modelling [33].

CUSP has proved useful for curriculum management and its very simple mapping of learning "ontologies" has worked well in practice. ProGoSs has potential that has yet to be demonstrated in a deployed system. For long term learner models, our work with CUSP, ProGoSs and infrastructures highlight the power of exploring simple approaches. Even these demand effort to create effective interfaces and to carefully take account for the pragmatics that will mean that people will see it as worth investing the time needed to make use of them.

In terms of Figure 1, *simple interfaces* onto *simple learner models* have huge but under-exploited potential. The human, context, cultural and interface challenges are critical and should underpin EDM design. There is a risk in taking too narrow a focus or too simplistic approaches that miss these [40]. Standardised tests and arbitrary data within administrative systems are simple sensors to drive EDM to produce aggregate models. But they also pose potential risks for misuse, for example, creating pressure for teachers and learners to focus on improving arbitrary measures on whatever is readily tested even if these relate only weakly to important learning outcomes.

Gaming-aware design

Gaming educational systems occurs when people subvert or violate the use that was intended and is required to support effecyive learning. We can expect gaming [3]. If we intend EDM for real world use, beyond the lab, we need to consider the drivers and opportunities for gaming by any and all of the stakeholders. The direct sensor data and the learner models of EDM are supremely "gamable" at many levels.

High stakes, simplistic use of learning data invites gaming. Even quite low stakes system, such as an ITS with formative assessment can have gaming [7] and there may be a fine line between gaming and "help" [1]. ITS/AIED/EDM communities have amassed considerable experience of gaming both in recognising it and using that to tackle it. This could be a foundation for a checklist to help system builders inform design, by considering potential gaming throughout the EDM process.

We detected gaming with Narcissus. For example, a student was called to account in one class for their lack of recent visible activity. In the following week, they appeared to have considerable activity. The design of Narcissus, its use, meant the group mentor simply clicked on each link to each action, to navigate through the activity. When that revealed trivial actions to create the *appearance* of activity, it created a teachable moment! Narcissus's simple measures, and its scrutability, meant that students see how to game it if its use is simple-minded. At the same time, its direct mapping to navigation of the complex space of the trac site made it easy for a teacher to scrutinise the actual activity.

One might expect that gaming would not occur in systems for personal use alone. There would seem to be no point in gaming the system. The learner would simply be fooling themselves, reflected in OLMs that are incorrect. Our earliest work aimed to provide a personal learning tool about a text editor [14]. Even so, we discovered one (and only one) user who tediously used the OLM interface to indicate they knew a great deal. Their log of actual activity told a very different story.

Our Personis learner model representation was robust against this form of gaming. It kept all evidence from its sensors. These were: log data use of the editor, with inferences translating this into a model of demonstrated knowledge; data from the use of the OLM. The OLM used a *resolvers* to interpret the evidence from the learner as more reliable than the usage analysis. Other resolvers treated behaviour as more reliable. Perhaps the OLM interface should have helped the learner appreciate this.

Personal data-mining

It is currently difficult for people to mine their control data. This is true on every level. It is often challenging to *access* the data, be it at the sensor level or within a learner model. Beyond that, it is difficult for a typical use to combine sources of data and to analyse collections in useful ways. Figure 1 shows interfaces for human-in-the-loop analyses such as advocated for machine learning generally [2]. These approaches seem particularly important for personal data mining so that the individual can annotate and preprocess their own data, its interpretation and processing.

Reflecting its roots, EDM has substantial work in sophisticated tutoring systems. Issues of personal data mining may be less pressing in these. But for the many sources of simple sensor data, the Quantified Self community is already exploring how to mine many forms of their own data, typically to gain self-awareness and often to tackle important long term goals. The famous example of Nicholas Felton [18] points to the huge amounts of diverse data that an individual can amass and then actually harness for their own needs. Another example is lifelogging, for example based on worn cameras [39] to collect rich personal data, requiring sophisticated image processing to support personal data mining.

When personal data mining is conducted by a learner, it has the potential to play a key role in the metacognitive and self-regulation processes. These are already central to the use of data by the Quantified Selfers. We know that many learners benefit from metacognitive scaffolding for such activities [5]. We need to explore how to create these.

It is unsurprising that mainstream data mining deals with big data, and aggregate model for many people. For example, Microsoft has a patent on personal data mining [48] but this is actually about mining of personal-data. EDM researchers are perfectly placed to lead initiatives in the quite different task of personal data-mining. It calls for methods that can deliver useful insights with the relatively small amounts of data of an individual.

Infrastructures for personal data-mining

Figure 1 hints at the infrastructures needed for the EDM processes of lifelong and life-wide learning. The sensors it shows are already embedded in the many technologies we use across our lives. These include formal learning with its plethora of devices and applications, ranging from the LMSs to the thousands to specialised apps, videos and other learning researches that can produce digital footprints.

Currently, the sensor data is splattered across many devices that we own as well as in the cloud and managed by diverse services that we do not control. When we explored how people wanted such data managed [11], we learnt that most people want to have control over it, even when they cannot yet establish a use for it. If we are to make use of such findings, we need to explore how to create infrastuctures that can support people in bringing together their diverse personal data. We will also need to tackle the HCI challenges of creating interfaces that enable people to actually manage their collected data and use it effectively.

One strand of my research has explored how to create an infrastructure for a lifelong learner model [30]. This is similar to the vision recently proposed by Nye [47] where a learner model becomes a web-service. My Personis family of learner models can make flexible use of ubicomp sensors [13, 4]. A central design goal was to support user control. So its representation has hooks that interface designers need to create interfaces that enable people to control data *from the sensors*, *in the learner model*, including the reasoning *within the model* and in all *uses of the model* by applications [29].

The infrastructure needed for lifelong EDM is similar to the notion of a personal data vault [46]. The key difference is that a learner model is more than a unified collection of sensor data. It needs to be designed to answer questions that matter to learners, either with direct access to the data, via an OLM interface, or indirectly because it can be used effectively in one or more applications. This was a driver for work, like CUSP and ProGoSs to help define the learner model ontology, and mappings between multiple ontologies [47]. In the case of CUSP, the ontologies were handcrafted. But our ProGoSs work explored use of standard and widely used resources, such as international and professional curriculum specifications for the ontology of learning elements. For the levels of learning, ProGoSs imported the definition and associated tutoring elements to help classroom teachers understand them. Our experiments with Bloom and Neo-Piagetian learning taxonomies make it clear that teachers can to learn to use these effectively [24]. This will enrich progressive modelling of lifelong learning.

EDM has made real headway on this problem of infrastructures for *aggregate models*. For example, the DataShop [35] has tackled infrastructure issues, providing standards for representation, tools for analysis and interfaces. These are valuable for the builders of ITS/AIED systems and for learning scientists. Learners may well be willing to contribute their learner models to similar aggregate data stores in a form of citizen science, so long as they have the assurances they need about the management of their data, including provenance metadata and privacy [20].

Interfaces for user control

System builders could consider how their design decision impact user control. This is partly a matter of taking the perspective of the learner, or other users, when building elements for EDM. For example, how can a learner define the structure of their long term learner model? And then link in the sensors? For the case of personal informatics sensors, such as physical trackers, we have demonstrated that a promising approach is to create interfaces for people to define goals [10] and link various sensors to these. In our user study, people could readily think about this data and its use in terms of their goals. They could then link the goals to various sensors, such as a FitBit, mobile phone app or a smart cushion. This is one human-centred approach to the design of control interfaces for infrastructures for EDM.

At a quite different level, *scrutability* could become a criterion for the design of the EDM, as well as software architectures [28]. These could then flow into a test-driven approach for building EDM systems. For this, we need to identify success measures that include interpretability of the EDM processes and learner models [49].

Stealth assessment [55] is appealing since it enables a learner to focus on learning, and getting assessment measures for free. These approaches could be made compatible with user control if the user is able to define what the systems that can log, as has been done for computer use [27].

Conclusions

Figure 1 presented a view of EDM with sensors associated with each learner, and EDM processes transform the sensor data into learner models. This view highlights the individual learner model, which holds only the data of one learner. This is increasingly becoming a long term *first class citizen*, independent of any one application, especially for reuse by multiple applications [16]. The data within the EDM system belongs to the learner (even if they may barely be aware of that). An OLM can enable a learner to see and, perhaps, also make good use of it. This personal EDM is catching on in the Quantified Self community.

A human-centred view of EDM will raise the profile of all the interfaces in the figure. One of these will scaffold metacognitive processes, to help the learner make effective use of their OLM. A quite different class of interfaces is needed to ensure learners can manage their learner models. It will be far harder to graft these on, as an after-thought, to the EDM processes. As we design and build each element of the EDM processes, we need to consider this goal. For example, this calls for consideration of how intelligible the processes are. We may begin to measure the trade-off between the performance of an EDM algorithm and how easy it is to explain it in ways different stakeholders find satisfactory. It will require capture of provenance and support for people to use this to define how they want their learner model used.

EDM and ITS/AIED have built strong foundations for aggregate learner modelling. These fit well into our historic core business of building personalised teaching systems, that combine aggregate learner models with the individual learner's model. Aggregate models could be key EDM contributions to learning science. They are also core for Learning Analytics, especially for use by teachers and the administrators.

Returning to the three cases studies, all relate to learning complex skills that we develop over years. All involved *lab* and *deployed* systems. All explored both *simple* and *sophisticated* EDM. The longitudinal group work had rather *conventional sensors*, based on interactions with trac. The tabletops involved more *diverse sensors*. The effectiveness of that deployed EDM in the classroom relied heavily on the ways that *students* and *teachers* used the OLMs. The tabletop work was driven by the needs of *classroom teachers*, with student interfaces still on the future work slate. CUSP and PRoGoS explored *infrastructures* for long term learning, with key stakeholders being the *curriculum caretakers*, both *administrators* at the level of the whole degree and the *teachers* of individual subjects.

This paper has presented a reflection on three strands of my research, with lesson learnt and how they might contribute to a vision for EDM research. In terms of the sensors, the learner models and the stakeholders, these case studies are outside the mainstream of EDM. This seems set to change. For example, Baker has comments that: "there is a disconnect between the vision of what intelligent tutoring systems could be, and what they are ... between the most impressive examples of what intelligent tutors can do, and what current systems used at scale do". He highlights the power achieved by a human-centred approach, with extensive EDM informing the design and refinement of the ASSISTments system [26]. One of the key lessons of the three case studies is that we have much to gain from *simplicity*. It is important for practical and useful deployments of systems. It should help in the design of interfaces for scrutability. This paper has argued for the need for a set of evidence-based guidelines for EDM design that considers the *many levels of gaming*. These must particularly help account for potential misuse of learner models when a stakeholder group repurposes them. I have proposed that we explore *personal data mining*, potentially building links with the Quantified Self movement. Finally, it calls for EDM research into practicalities of creating *infrastuctures for EDM* with associated interfaces so *people can control their data and associated EDM processes.* These are parts of a broad vision for EDM that supports lifelong and life-wide learning.

4. ACKNOWLEDGMENTS

The research discussed in this paper was supported by several sources: Australian Research Council; Smart Internet Co-operative Research Centre; Smart Services Co-operative Research Centre; the Faculty Research Cluster Program, Faculty of Engineering and Information Technologies, The University of Sydney.

5. REFERENCES

[1] V. Aleven, I. Roll, B. M. McLaren, and K. R. Koedinger. Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, pages 1–19, 2016.

[2] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza. Power to the people: The role of humans in interactive machine learning. American Association for Artificial Intelligence, 2014.

[3] D. Ariely and S. Jones. *Predictably irrational.* HarperCollins New York, 2008.

[4] M. Assad, D. Carmichael, J. Kay, and B. Kummerfeld.

Personisad: Distributed, active, scrutable model framework for context-aware services. *Pervasive Computing*, pages 55–72, 2007.

[5] R. Azevedo and A. F. Hadwin. Scaffolding self-regulated learning and metacognition–implications for the design of computer-based scaffolds. *Instructional Science*, 33(5):367–379, 2005.

[6] A. D. Baddeley. *Human memory: Theory and practice*. Psychology Press, 1997.

[7] R. Baker, J. Walonoski, N. Heffernan, I. Roll, A. Corbett, and K. Koedinger. Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research*, 19(2):185, 2008.

[8] R. S. Baker. Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education*, 26(2):600–614, 2016.

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

[10] D. Barua, J. Kay, B. Kummerfeld, and C. Paris. Modelling long term goals. In *User Modeling, Adaptation, and Personalization*, pages 1–12. Springer, 2014.

[11] D. Barua, J. Kay, and C. Paris. Viewing and controlling personal sensor data: what do users want? In *Persuasive Technology*, pages 15–26. Springer, 2013.

[12] S. Bull and J. Kay. Smili: a framework for interfaces to learning data in open learner models, learning analytics and related fields. *International Journal of Artificial Intelligence in Education*, pages 1–39, 2016.

[13] D. J. Carmichael, J. Kay, and B. Kummerfeld. Consistent modelling of users, devices and sensors in a ubiquitous computing environment. *User Modeling and User-Adapted Interaction*, 15(3-4):197–234, 2005.

[14] R. Cook and J. Kay. The justified user model: a viewable, explained user model. In *Proceedings of the Fourth International Conference on User Modeling UM94*, pages 145 – 150, 1994.

[15] M. Dennis, J. Masthoff, and C. Mellish. Adapting progress feedback and emotional support to learner personality. *International Journal of Artificial Intelligence in Education*, pages 1–55, 2015.

[16] P. Dillenbourg. The evolution of research on digital education. *International Journal of Artificial Intelligence in Education*, 26(2):544–560, 2016.

[17] S. K. D'Mello. Giving eyesight to the blind: Towards attention-aware AIED. *International Journal of Artificial Intelligence in Education*, 26(2):645–659, 2016.

[18] C. Elsden, D. S. Kirk, and A. C. Durrant. A quantified past: Toward design for remembering with personal informatics. *Human–Computer Interaction*, pages 1–40, 2015.

[19] T. Erickson and W. A. Kellogg. Social translucence: an approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)*, 7(1):59–83, 2000.

[20] A. Fekete, J. Kay, M. Franklin, D. Barua, and B. Kummerfeld. Managing information for personal goals (vision). In *Data Engineering Workshops (ICDEW), 2015 31st IEEE International Conference on*, pages 30–33. IEEE, 2015.

[21] S. Few. *Information dashboard design*. O'Reilly, 2006.

[22] R. Gluga, J. Kay, and T. Lever. Modeling long term learning of generic skills. In *Intelligent Tutoring Systems*, pages 85–94. Springer, 2010.

[23] R. Gluga, J. Kay, R. Lister, M. Charleston, J. Harland, D. Teague, et al. A conceptual model for reflecting on expected learning vs. demonstrated student performance. In *Proceedings of the Fifteenth Australasian Computing Education Conference-Volume 136*, pages 77–86. Australian Computer Society, Inc., 2013.

[24] R. Gluga, J. Kay, R. Lister, and S. Kleitman. Mastering cognitive development theory in computer science education. *Computer Science Education*, 23(1):24–57, 2013.

[25] R. Gluga, J. A. Kay, and T. Lever. Foundations for modeling university curricula in terms of multiple learning goal sets. *Learning Technologies, IEEE Transactions on*, 6(1):25–37, 2013.

[26] N. T. Heffernan and C. L. Heffernan. The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4):470–497, 2014.

[27] Z. Hinbarji, R. Albatal, N. O'Connor, and C. Gurrin. Loggerman, a comprehensive logging and visualization tool to capture computer usage. In *MultiMedia Modeling*, pages 342–347. Springer, 2016.

[28] J. Kay. Scrutable adaptation: Because we can and must. In *Adaptive hypermedia and adaptive web-based systems*, pages 11–19. Springer, 2006.

[29] J. Kay and B. Kummerfeld. Creating personalized systems that people can scrutinize and control: Drivers, principles and experience. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(4):24, 2012.

[30] J. Kay and B. Kummerfeld. Lifelong learner modeling. *Adaptive Technologies for Training and Education*, pages 140–164, 2012.

[31] J. Kay, N. Maisonneuve, K. Yacef, and P. Reimann. The big five and visualisations of team work activity. In *Intelligent tutoring systems*, pages 197–206. Springer, 2006.

[32] J. Kay, N. Maisonneuve, K. Yacef, and O. Zaïane. Mining patterns of events in students' teamwork data. In *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems (ITS 2006)*, pages 45–52, 2006.

[33] J. Kay, P. Reimann, E. Diebold, and B. Kummerfeld. Moocs: So many learners, so much potential... *IEEE Intelligent Systems*, (3):70–77, 2013.

[34] A. Kobsa. Privacy-enhanced web personalization. In *The adaptive web*, pages 628–670. Springer, 2007.

[35] K. R. Koedinger, R. S. Baker, K. Cunningham, A. Skogsholm, B. Leber, and J. Stamper. A data repository for the edm community: The pslc datashop. *Handbook of educational data mining*, 43, 2010.

[36] K. R. Koedinger, E. Brunskill, R. S. Baker, E. A. McLaughlin, and J. Stamper. New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine*, 34(3):27–41, 2013.

[37] R. Kohavi, N. J. Rothleder, and E. Simoudis. Emerging trends in business analytics. *Communications of the ACM*, 45(8):45–48, 2002.

[38] I. Li, Y. Medynskiy, J. Froehlich, and J. Larsen. Personal informatics in practice: improving quality of life through data. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 2799–2802. ACM, 2012.

[39] N. Li, M. Crane, C. Gurrin, and H. J. Ruskin. Finding motifs in large personal lifelogs. In *Proceedings of the 7th Augmented Human International Conference 2016*, AH '16, pages 9:1–9:8, New York, NY, USA, 2016. ACM.

[40] D. Y.-T. Liu, T. Rogers, and A. Pardo. Learning analytics - are we at risk of missing the point? In *Proceedings of the 32nd Ascilite Conference*, 2015.

[41] R. M. Maldonado, J. Kay, K. Yacef, and B. Schwendimann. An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In *Intelligent Tutoring Systems*, pages 482–492. Springer, 2012.

[42] R. Martinez-Maldonado, A. Clayphan, and J. Kay. Deploying and visualising teacher's scripts of small group activities in a multi-surface classroom ecology: a study in-the-wild. *Computer Supported Cooperative Work (CSCW)*, 24(2-3):177–221, 2015.

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

[44] R. Martinez-Maldonado, Y. Dimitriadis, A. Martinez-Monés, J. Kay, and K. Yacef. Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. *International Journal of Computer-Supported Collaborative Learning*, 8(4):455–485, 2013.

[45] R. Martinez-Maldonado, K. Yacef, and J. Kay. Data mining in the classroom: Discovering groupsâĂŹ strategies at a multi-tabletop environment. In *Proceedings of the International Conference on Educational Data Mining*, pages 121–128, 2013.

[46] M. Y. Mun, D. H. Kim, K. Shilton, D. Estrin, M. Hansen, and R. Govindan. Pdvloc: A personal data vault for controlled location data sharing. *ACM Transactions on Sensor Networks (TOSN)*, 10(4):58, 2014.

[47] B. D. Nye. Its, the end of the world as we know it: Transitioning aied into a service-oriented ecosystem. *International Journal of Artificial Intelligence in Education*, 26(2):756–770, 2016.

[48] R. Ozzie, W. Gates, G. Flake, T. Bergstraesser, A. Blinn, C. Brumme, L. Cheng, M. Connolly, N. Dani, D. Glasgow, et al. Personal data mining, 2011. US Patent 7,930,197.

[49] R. Pelánek. Metrics for evaluation of student models. *JEDM-Journal of Educational Data Mining*, 7(2):1–19, 2015.

[50] D. Perera, J. Kay, I. Koprinska, K. Yacef, and O. R. Zaïane. Clustering and sequential pattern mining of online collaborative learning data. *Knowledge and Data Engineering, IEEE Transactions on*, 21(6):759–772, 2009.

[51] E. Salas, D. E. Sims, and C. S. Burke. Is there a "Big Five" in Teamwork? *Small Group Research*, 36(5):555–599, Oct. 2005.

[52] D. A. Schön. *The reflective practitioner: How professionals think in action*, volume 5126. Basic books, 1983.

[53] G. Siemens and R. S. d Baker. Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 252–254. ACM, 2012.

[54] K. Upton and J. Kay. Narcissus: group and individual models to support small group work. In *User modeling, adaptation, and personalization*, pages 54–65. Springer, 2009.

[55] L. Wang, V. Shute, and G. R. Moore. Lessons learned and best practices of stealth assessment. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 7(4):66–87, 2015.

[56] O. M. Yigitbasioglu and O. Velcu. A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, 13(1):41–59, 2012.