JOURNAL OF LEARNING ANALYTICS

Volume 2 - Issue 1

SOLAR
SOCIETY for LEARNING ANALYTICS RESEARCH

(2015). Learning Analytics – A Growing Field and Community Engagement. *Journal of Learning Analytics, 2*(1), 1-6.

Learning Analytics – A Growing Field and Community Engagement

Dragan Gašević, Shane Dawson, Negin Mirriahi, and Phillip D. Long
Editors, Journal of Learning Analytics

jla.editorial@gmail.com

Abstract. This editorial discusses events that marked the period since the publication of the previous issue – the 5th International Conference on Learning Analytics and Knowledge (LAK 2015), Learning Analytics Summer Institutes (LASIs 2015), and Learning Analytics Policy Briefing in the European Parliament. This period saw releases of two important publications for system-wide implementation of learning analytics in higher education published by Jisc and the Australian Government's Office for Learning and Teaching. An important recognition of the maturation of the field of learning analytics is the recent publication of the 2015 Google Scholar Metrics identifying the LAK proceedings as the only conference proceedings among the 20 most cited publication venues in educational technology. Building bridges for enhancing impact is another important activity for the field maturation through developing linkages of learning analytics with educational data mining, user modeling, the learning sciences, technology enhanced learning, cyber-learning, and learning at scale. This editorial also introduces a special section published in this issue dedicated to the exploration of connections between self-regulated learning and learning analytics, introduces two regular research papers featured in this issue and describes several special sections that will be published in future issues of the journal.

Keywords: learning analytics, self-regulated learning, Society for Learning Analytics Research, bridge building, Google Scholar

1 WHAT HAS HAPPENED IN THE MEANTIME?

Welcome to the first issue of the second volume of the Journal of Learning Analytics. The issue comes on an exciting wave of continuous growth for the field. Numerous events and publications have marked the steady maturation of the field among a wide international community of learning analytics researchers and practitioners. For example, the centrepiece annual event, organized by the Society for Learning Analytics Research (SoLAR), is the International Conference on Learning Analytics & Knowledge (LAK 2015). This year LAK15 reached its highest number (320) of participants leading to a fully-capped conference. The event was impeccably organized by Marist College with the outstanding leadership of Josh Baron and his co-leads Grace Lynch (SoLAR) and Nicole Maziarz (Marist College). The conference also featured an exciting program arranged by the conference program co-chairs (Paulo Blikstein, Agathe Merceron, and George Siemens), practitioner track chair (Alan Berg), doctoral consortium chairs (Simon Buckingham Shum, Katherine Maillet, Dan Suthers, and Stephanie Teasley), workshop and tutorial chairs (Christopher Brooks and Negin Mirriahi), and demonstrations and posters chairs (Gregory

(2015). Learning Analytics – A Growing Field and Community Engagement. *Journal of Learning Analytics, 2*(1), 1-6.

Dyke and Marcelo Worsley). The community owes many thanks to Marist College and a small army of individuals and sponsoring organizations/partners who invested much time, hard work and donations to further promote the field. While academic recognition is often attributed to research publications and awards, the role of those who serve the community through the organization of field development and growth events needs to receive at least as equal recognition. As per our tradition established with the LAK 2013 and LAK 2014 editions, a special section of the journal showcasing some of the LAK 2015 papers is in preparation and scheduled for publication in issue three of this volume.

The hard work and commitment to high quality research and practice in learning analytics has already payed off. We were very excited to learn that that the ACM proceedings from the LAK conference are the only proceedings in the top 20 most cited publication venues in the field of educational technology according to the 2015 Google Scholar Metrics[1]. Everyone who has supported the conference through their organization, paper authoring, reviews, sponsorship, workshops and promotions deserve our many thanks. The rapid rise and quality of the LAK proceedings reflects very positively on the efforts of the entire community - congratulations. We are confident that this achievement is merely the start of a long and fruitful future for learning analytics and is a good motivator for all us to stay committed to the core values of the field – high quality, inclusiveness, openness, transparency, and multidisciplinarity.

The recognition of contributions, open and inclusive have been core values that SoLAR has tried to promote and follow. We have previously discussed there is a need for ongoing reflections and evaluations of the emerging research networks forming in the learning analytics literature (Dawson, Gašević, Siemens, & Joksimovic, 2014; Mirriahi, Gasevic, Dawson, & Long, 2014). These reflections help to identify clusters of emerging research and also raise awareness of individuals developing their careers and interest in the field. The need for community growth and leadership turnover is critical if we are to maintain a strong and fertile presence. To date many individuals have contributed to the development of SoLAR and the field. A bulk of that has been completed through activating ties of informal networks established around our emerging field. Presently, SoLAR has consolidated its structure and governing procedures and there a many ways individuals can be involved in supporting and leading the direction of the field via the governing body of SoLAR. For the first time, SoLAR held elections that are open to all its members for the first four seats available on the Executive Committee (EC). We welcome Abelardo Pardo, Hendrik Drachsler, Srecko Joksimovic (student representative), and Leah Macfadyen to the SoLAR EC. The term on the EC ended for George Siemens (the founding President of SoLAR), Erik Duval and Nancy Law. George, Erik and Nancy have made great contributions to the field in these early critical years and deserve many thanks and much appreciation for their hard work and commitment. As part of this process, Dragan Gasevic was elected as President for a two year term (2015-2017). Other office bearers are Simon Buckingham Shum (Vice-President), Philip Long (Vice-President), Caroline Haythornhwaite (Secretary), and Shane Dawson (Treasurer). We strongly encourage all members of the community to be involved and run for seats on the SoLAR EC or general committee for the 2016 elections.

[1] https://scholar.google.ca/citations?view_op=top_venues&hl=en&vq=soc_educationaltechnology

(2015). Learning Analytics – A Growing Field and Community Engagement. *Journal of Learning Analytics, 2*(1), 1-6.

The Learning Analytics Summer Institutes (LASIs) have been well established as "an awareness-raising, capacity-building, international network of events, where researchers and practitioners meet to share work, ideas and opportunities for future collaborations (Buckingham Shum & Ochoa, 2015). The major news for LASIs 2015 is that they are now organized as a network of distributed events happening across the world. This is unlike LASIs 2013 and LASIs 2014 , which were designed and organised as a central hub (at Stanford University in 2013 and at Harvard University in 2014), with several localised yet connected events. This development recognises the broader learning analytics community and supports those members who are interested in organizing more local events that serve the needs of their geographic regions. The LASI events that were or are planned to be held during 2015 include: Aalborg (Denmark), Bilbao (Spain), Boston (USA), Washington DC (USA), Wellington (New Zealand), Pretoria (South Africa), Wuhan (China), and Sydney (Australia). The coordinators of the LASI network Simon Buckingham Shum and Xavier Ochoa together with the numerous organizers of the regional LASI events deserve many thanks for their hard work and contributions.

Several other events and publications of high impact for policy and strategy of learning analytics marked the period between the publication of this and the past issue of the journal. The EU-funded LACE project organized a learning analytics policy briefing in the European Parliament in Brussels, Belgium (LACE Project, 2015). This event was of strategic importance for the field building as it increased awareness of the learning analytics for the many policy makers in spaces of education and information and communication technologies in Europe. Jisc – a non-departmental public body responsible for promoting the use of information and communications technology (ICT) in higher education in the United Kingdom – published its code of practice for learning analytics for guiding higher education institutions in their implementation activities (Sclater & Bailey, 2015). The code is a result of a very thorough process engaging a board range of stakeholders and through publication of a comprehensive literature review (Sclater, 2015). The Australian Government's Office for Learning and Teaching (OLT) also released a report (Colvin et al., 2015), which offers guidelines for the critical tasks higher education institutions need to undertake to aid learning analytics implementations. Both the Jisc and OLT reports fulfill a significant gap in the existing literature by providing practical guidelines in high demand by many institutions that are exploring the ways of how to engage in system-wide implementation of learning analytics.

2 LOOKING AHEAD

Since its early days, recognizing the multi-disciplinary nature of our field, SoLAR aimed to build bridges with other fields related to learning analytics. We have been fortunate enough to establish strong partnership links with educational data mining researchers and the International Educational Data Mining Society (IEDMS). George Siemens and Ryan Baker – presidents of SoLAR and IEDMS at the time – wrote an important piece (Siemens & Baker, 2012, p. 252) arguing for "increased and formal communication and collaboration between these communities in order to share research, methods, and tools for data mining and analysis in the service of developing both fields". Examples of this collaboration are well demonstrated through numerous formal and informal occasions – e.g., SoLAR organized LASI 2013 and LASI 2014 in assistance with IEDMS (in addition to several other partners), while numerous individuals from both communities participated in the organization of or presentation of papers at their main annual conferences (EDM and LAK) and published in their journals (Journal of Educational Data Mining and Journal of Learning Analytics). This exchange between the two communities is presently a very healthy state and many new exciting collaborations are underway. We would also like to take an opportunity to express our great appreciation for the work done in the development of these bridges to both George Siemens and Ryan Baker – whose presidency terms ended earlier this year. George and Ryan have well demonstrated visionary leadership, outstanding work ethic for their respective communities and an unconditional commitment to forging relationships among the two communities of researchers and practitioners. We would also like to take this opportunity to congratulate Mykola Pechenizkiy on his election as the President of IEDMS and look forward to working with him and other educational data mining researchers on new joint projects.

Although the link with educational data mining was quite natural from the very beginning, learning analytics has drawn on, builds on, and connects to many other fields such as the learning sciences, computer supported collaborative learning, technology enhanced learning, cyber-learning, learning at scale, and user modelling. The connections with other fields are obvious even through the examination of the proceedings of the main conferences of some of these related fields. For example, the recent 11th International Conference on Computer Supported Collaborative Learning (CSCL 2015)[2] had nine papers, three posters, and one invited session with the phrase learning analytics in their title[3]. This is a remarkable change compared to CSCL 2013 where no paper used learning analytics in their title. This does not include quite a bit of work, potentially presented at both CSCL 2013 and CSCL 2015, which did learning analytics without labeling it as such. Initiated by Carolyn Penstein Rosé – the new President of the International Society of the Learning Sciences (ISLS) – regular meetings have been scheduled for coordination of the organizations behind some of these related fields. In addition to SoLAR, the organizations represented on these coordination meetings are ISLS, EDM, International Artificial Intelligence in Education Society (IAIEDS), European Association for Technology Enhanced Learning (EATEL), User Modeling Inc., the steering committee of the ACM Learning at Scale conference, and the Center for Innovative Research in Cyber-learning (CIRCL). This has already resulted in joint panels on grand research challenges organized at the CSCL 2015 and EDM 2015 conferences. We can expect this trend to continue and further accelerate in the future. An obvious indicator of this trend is the recently announced LAK 2016 conference – hosted by the University of Edinburgh in April 2016. The conference specifically calls for papers that look at enhancing impact through convergence of communities for grounding, implementation and validation[4]. In this spirt of bridge building, it is exciting to see that the LAK conference will be collocated with the third Annual ACM Conference on Learning at Scale (L@S 2016)[5].

[2] http://www.isls.org/cscl2015/

[3] Many thanks to Alyssa Wise for compiling and sharing this information.

[4] http://lak16.solaresearch.org/?page_id=6

[5] http://learningatscale.acm.org/las2016/

(2015). Learning Analytics – A Growing Field and Community Engagement. *Journal of Learning Analytics, 2*(1), 1-6.

3 IN THIS ISSUE

This issue features a special section that builds an important bridge between learning analytics and educational psychology by featuring the special section dedicated to self-regulated learning and learning analytics. Learning analytics brings the potential to understand and optimize learning and the environments in which learning happens by using data about learners and learning situations. For learning analytics to realise its potential, it is necessary to understand what cognitive, metacognitive, affective and motivational states influence learners' decisions when studying and when analytics are fed back to learners to reflect on their learning. We have been fortunate to have Ido Roll and Philip H. Winne – leading experts in self-regulated learning and learning analytics – to take a lead on this topic and guest edit this special section. Roll and Winne compiled a special issue that features five exciting articles that look at different angles of the linkage between learning analytics and self-regulated learning.

This issue also features two regular research papers. Very much in sync with highly represented interests in using different methods for analysis and interpretation of discourse and text expressed in several papers presented at LAK 2015 (Baron et al., 2015), Justin Reich, Dustin Tingley, Jetson Leder-Luis, Margaret E. Roberts, and Brandon Stewart introduce the Structural Topic Model – a computational method for language analysis – and showcase its use for computationally-aided discovery and reading for student generated text in massive open online courses. In line with the topic, Simon Knight and Karen Littleton map the lay of the land of discourse-centered learning analytics (DCLA) and discus specific contributions that DCLA brings to the table in contrast to computational methods for discourse analysis.

4 BEYOND THIS ISSUE

The process of setting up a new journal is highly demanding and requires involvement and contributions of numerous parties and individuals. As learning analytics connects with many well-known and emerging themes related to learning and education, we felt that arranging different special sections of the journal dedicated to some specific topics is a good way to stimulate engagement of a broad group of interested authors. We have been fortunate to work with many outstanding researchers who were willing to contribute their time and expertise to lead and guest edit several special sections that will be published in the coming issues. At the moment, the special sections on the following topics are being prepared:

- Learning analytics and learning theory guest edited by Alyssa Wise and David Williamson Shaffer
- Multimodal learning analytics guest edited by Xavier Ochoa and Marcelo Worsley
- Ethics and privacy in learning analytics guest edited by Hendrik Drachsler, Rebecca Ferguson, Tore Hoel, and Maren Scheffel
- Learning analytics for 21st century competencies guest edited by Simon Buckingham Shum and Ruth Deakin Crick
- Tutorials in learning analytics methods and techniques guest edited by Dragan Gašević and Mykola Pechenizkiy

- Dataset descriptions for learning analytics guest edited by Stefan Dietze, Hendrik Drachsler, George Siemens, and Davide Taibi
- Selected and revised best papers presented at LAK'15 guest edited by Paulo Blikstein, Agathe Merceron, and George Siemens

We ask you stay tuned and watch out for the publication of these exciting special sections in the future issues of the journal. In the meantime, enjoy the papers published in this issue. And, do not forget to submit your manuscripts reporting on your recent research and practice results in the near future. We would also very much like to learn about your ideas for the new special sections and other ideas that you think could make the journal an even a better place for the exchange of the state of the art results and ideas in learning analytics.

REFERENCES

Baron, J., Lynch, G., Maziarz, N., Blikstein, P., Merceron, A., & Siemens, G. (Eds.). (2015). *LAK '15: Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*. New York, NY, USA: ACM.

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gašević, D., Buckingham Shum, S., … Fisher, J. (2015). *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement* (Research Report). Canberra, Australia: Office of Learning and Teaching, Australian Government.

Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014). Current State and Future Trends: A Citation Network Analysis of the Learning Analytics Field. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 231–240). New York, NY, USA: ACM. http://doi.org/10.1145/2567574.2567585

LACE Project. (2015). Policies for Educational Data Mining and Learning Analytics Briefing. Retrieved July 14, 2015, from http://www.laceproject.eu/spring-briefing/

Mirriahi, N., Gasevic, D., Dawson, S., & Long, P. D. (2014). Scientometrics as an Important Tool for the Growth of the Field of Learning Analytics. *Journal of Learning Analytics*, *1*(2), 1–4.

Sclater, N. (2015). *Code of practice for learning analytics: A literature review of the ethical and legal issues*. Jisc. Retrieved from http://repository.jisc.ac.uk/5661/1/Learning_Analytics_A-_Literature_Review.pdf

Sclater, N., & Bailey, P. (2015). *Code of practice for learning analytics*. Jisc. Retrieved from https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

Siemens, G., & Baker, R. S. J. d. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). New York, NY, USA: ACM. http://doi.org/10.1145/2330601.2330661

Buckingham Shum, S., & Ochoa, X. (2015). LASI 2015 International Network. Retrieved July 14, 2015, from http://solaresearch.org/events/lasi/lasi2015/

(2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12.

Understanding, evaluating, and supporting self-regulated learning using learning analytics

Ido Roll
University of British Columbia, Canada
ido.roll@ubc.ca

Philip H. Winne
Simon Fraser University, Canada

ABSTRACT: Self-regulated learning is an ongoing process rather than a single snapshot in time. Naturally, the field of learning analytics, focusing on interactions and learning trajectories, offers exciting opportunities for analyzing and supporting self-regulated learning. This special section highlights the current state of research at the intersection of self-regulated learning and learning analytics, bridging communities, disciplines, and schools of thought. In this opening article, we introduce the papers and identify themes and challenges in understanding and supporting self-regulated learning in interactive learning environments.

KEYWORDS: self-regulated learning, learning analytics

1 INTRODUCTION

In line with many models of self-regulated learning and, more generally, constructivism, we adopt the view that learners are agents (Jonassen, 1991). Agents seek out, choose and carry out options to achieve goals. However, whether by design or not, a learner's environment constrains the possibility of achieving particular goals and of taking particular steps to approach a goal. Experimenters and instructors may interpret that their interventions directly change processes learners use in learning, what learners learn, and learners' motivations and affective stances. But, they are mistaken. Interventions do not determine how learners engage with tasks. Rather, interventions are affordances that agentic learners meld with other elements in the context as they perceive it to regulate learning.

In terms of one widely cited model of self-regulated learning proposed by Winne and Hadwin (1998; Winne, 2011; Greene & Azevedo, 2007), learners exercise agency across four loosely sequenced phases: (1) They scan their environment to identify internal factors (cognitive, motivational, affective) and external features that may influence a task. (2) They frame goals and design plans to approach them. (3) They implement actions to animate their plans, monitor the match between a plan and its actualization and modestly adjust actions as they judge appropriate. And, (4) they re-examine aspects across these three prior phases to consider major, strategic revisions to understanding and action if progress toward goals is blocked, too slow or in some other way unsatisfactory. From this perspective, agentic learners are inherently self-regulating learners. Learners regulate their learning constantly, even if not always successfully as judged by the learner or by another, e.g., a teacher or researcher.

While the importance of self-regulated learning (SRL) to learning is obvious (Bransford, 2000), understanding how SRL is applied in context is not a simple task. We know how to interpret and evaluate a learner's knowledge of domains in terms of correctness or the quality of solutions to problems. Yet interpreting and evaluating qualities of actions, strategies, goals and, more broadly, regulation is a much more challenging task (Roll et al., 2014). Data gathered in online learning environments are instrumental to addressing this challenge. However, without appropriate tools, theories, and frameworks, data alone are insufficient to shape theory and guide practice. In this context, learning analytics can play an immensely important role for learners and for the learning sciences.

Learning analytics are reports of analyses of data that describe features of, and factors that influence, SRL. At the intersection of learning analytics and SRL, researchers can realize significant opportunity to better research and understand learning in authentic learning settings. Because learners are agentic, learning analytics can inform them about options that may bear on the phases of their self-regulated learning as described in Winne and Hadwin's model. Finally, expanding our understanding how SRL unfolds in context spurs developing learning environments that are concurrently better tools for gathering data for learning analytics and better at helping learners learn.

To realize the potential learning analytics for advancing the learning sciences, we collectively need to undertake three main kinds of research. The first relates to developing instrumentation to record traces of self-regulated learning across all its phases. Without data, neither learners nor learning scientists can progress toward their respective goals. Second, methods need to be fashioned that identify structures within data. Correspondingly, methods need to be tested for the extent and conditions they contribute to valid inferences about the constructs that the structures of data represent (Roll, Baker, Aleven, & Koedinger, 2014). Thirdly, standing on these pillars of instrumentation (data collection) and methodologies that operate on data, learners and learning scientists alike are positioned to explore the effects of interventions. We construe interventions broadly as changes in a learning environment that range over internal cognitive, motivational and affective states of a learner as well as information externally introduced to learners by experimenters instructors or peers. In keeping with our axiom that learners are agents, we characterize interventions as affording opportunities for learners to more productively regulate learning in the service of achieving outcomes they value.

2 ABOUT THE SPECIAL SECTION

This special section showcases current exciting opportunities for learning analytics and learning sciences. Data and computing power are no longer a bottleneck. It is now time to re-examine theories, methods, and practice, hand-in-hand, promoting a synergy among these three pillars. The research reported in this special section illustrates important steps made in each of these areas of work. The first paper by Segedy, Kinnebrew, and Biswas (this special section) introduce Coherence Analysis – an analytical framework for interpreting students' actions in an exploratory learning environment.

(2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12.

Coherence analysis interprets sequences of actions, and evaluates whether learners make use of information they collect. Thus, coherence analysis offers a tool to evaluate learning behaviours that are highly contextualized. The second paper, by Cutumisu, Blair, Chin, and Schwartz (this special section) also focus on assessing SRL, albeit using a different approach. Cutumisu and colleagues introduce Posterlet, a learning environment designed to evaluate students' tendencies to seek actionable feedback and act upon it. In addition to offering this environment, Cutumisu and colleagues shed some light on skills comprising SRL and other learner attributes. Whereas their paper evaluates the impact of feedback at the domain level, Sonnenberg and Bannert, the third paper in this special section, take a different approach by focusing on the effect of metacognitive prompting. They offer a broad analysis of their intervention using verbal protocol analysis, trace data analysis, and assessment of learning.

Two additional papers are included in a Work in Progress part of this special section. Nussbaumer, Hillemann, Gütl, and Albert take a theory-driven approach to design and implement an architecture for supporting students' SRL in a variety of activities. Their architecture is interesting in that it builds on multiple approaches to SRL, and combines intelligent models with open learner model. Colthorpe, Zimbardi, Ainscough, and Anderson look for traces of SRL and metacognition in students' use of a common environment, the Learning Management System. By comparing students' traces, self-reports, and academic performance, their work adds to our understanding of SRL in academic online settings.

The diversity of papers in this special section offers an interesting snapshot of the current state of research into SRL and learning analytics. Several themes emerge:

- **Choices**. Rather than looking at the content of learners' actions, we can look at the choices that learners make. SRL theories describe how students manage their learning, and open ended environments that gather trace data about these processes allow us to evaluate the types of actions students choose to perform, rather than only their content (Cutumisu; Colthorpe)
- **Relativity**. Actions are neither inherently good nor bad. Rather, choices students make reflect their perceived challenges, knowledge, prior experiences, and habits. Thus, actions should be interpreted in relation to a learner's context (Biswas; Nussbaumer).
- **Reusable analytics.** The volume of learning data gathered in online settings affords new opportunities to explore complex research questions. New methods and analytical techniques that provide insight into the complexity of SRL represented in these spaces are required. Specifically, we see a shift from looking at individual events to looking at sequences of actions (Biswas; Bannert). Developing new tools that can be reused across contexts and settings can contribute to the generality of theories and to support and evaluate transfer of SRL patterns (Biswas; Colthorpe).
- **Challenge: operationalization.** Many of the papers address similar constructs (e.g., planning, revising) and operationalize them in different ways, contextualized in their respective systems. Learning analytics allows us to take abstract constructs and instantiate them in different learning environments. Comparing and contrasting these instantiations can help us evaluate the effect of contextual and environmental factors on learners' SRL.

(2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12.

3 FUTURE DIRECTIONS FOR RESEARCH

The sum of papers in this special section also highlights opportunities for future research. In a general sense, learning to self-regulate involves identifying patterns for later use. Patterns can offer greater information value that may improve the predictability of goal attainment and enhance a learner's understanding (theory) about which actions relate to goal attainment. The world, however, is stochastic; so methods for detecting patterns – human and machine – must wrestle with noise and indeterminacy.

One approach to this problem requires two distinct moves. The first is gathering an increased variety of types of data such that each type of data is relatively statistically independent of other types. The second move is to massively increase the volume of data. Elevating dimensionality of data in these two ways can mitigate two problems. First, adding dimensionality to data combats randomness expressed in the form of ignorance about factors that play causal roles. Second, worries about overfitting models can be lessened because methods that cluster or partition vastly dimensional, big data can be applied to form relatively homogeneous subsets in which variance is less prevalent than it is in the complete pool. Together, these two tacks are merely a post hoc operationalization of the lab experimenter's everyday stance toward controlling a potentially confounding factor in an empirical study.

This two-pronged approach leads to four implications for learning analytics and research on self-regulated learning. First, instrumentation needs to be designed that learners can use nearly ubiquitously. This helps to increase the volume of data. Winne and his colleagues' nStudy platform (e.g., Winne, 2015) and content-agnostic learning management systems illustrate this step. However, as learning technologies evolve, the footprint of any single system shrinks. Thus, systems need to support better interoperability. New technologies and standards (such as TinCan xAPI) attempt to bridge this gap. Second, algorithms capable of detecting partitions or clusters in vastly dimensional, big data need to be improved or invented. Presently, analytics that are supposedly general are often tested only in one setting. Researchers should evaluate the generalizability not only of theories, but also of tools and methodologies. Third, as data are collected more ubiquitously, we need to do a better job of attending to and capturing features of context. While systems collect information about students' actions, it is rare that enough data are available to describe both static and modulating external factors such as learner and task characteristics. Interpreting analytics about SRL needs to take account of the context in which data were collected. Last, new research is needed that investigates how to meaningfully convey analytics to learners when the data analyzed originates in vastly dimensional data. The challenge here is to explore how options for self-regulating learning are usefully described given the likely complexity of patterns of learning activities. In other words, how can information overload be avoided while not stripping out necessary information about how learning activities unfold over time?

Future research on learning analytics in service of understanding and promoting productive self-regulated learning should also tackle challenges presented by the very probable, and sometimes actively encouraged, condition that learning is social. One attempt to sidestep this challenge is to entertain a view that interactions among learners are no different than interactions a focal learner has with a static text. On a turn-by-turn basis, this is reasonable. Like Zeno's apparent paradox of Achilles and the tortoise (wherein Achilles never catches a tortoise allowed a head-start because, in each successive interval of time, Achilles covers only a portion of distance the has travelled in that time interval), the turn-by-turn view is inadequate. A genuinely social exchange consists of successive turns are not independent (not Markovian).

Wait – isn't this temporal structure fundamentally the same as temporally unfolding SRL wherein an agentic learner reviews her learning history to judge and adjust learning activities? Yes, it is. We surmise a bridge to join the solo to the social context for regulated learning activities lies in using the topic a learner studies as a means to follow how "threads" of content unfold across time in conjunction with patterns of self-regulated learning. To our knowledge, methods are not yet available that can simultaneously track patterns in which elements simultaneously represent the topic (subject matter, features of group process, a learner's affect) in the same unit as a representation of the cognitive operations applied to that topic. We recommend this as a topic worthy of future work. Furthermore, as learning is often a social process, designs for learning analytics need to take account of how a group regulates its learning, i.e., forms of co-regulated and socially-shared regulation of learning (Hadwin, Jarvella, and Miller, 2011). The field needs to make progress toward mapping and developing learning analytics for nested models of regulated learning.

Finally, many of the challenges faced by learning analysts who explore self-regulated learning are shared with colleagues working in the field of data mining (see Winne & Baker, 2013). Strengthening bridges that join the learning analytics and data mining communities will benefit and, perhaps, even merge these two fields of work. We recommend scholars and funders find ways to promote increased exchanges amongst members of these communities.

Overall, the intersect of learning analytics and SRL offers a grand challenge. Grand in its magnitude; grand in its potential impact; and grand in that opportunities for meaningful progress are within reach. This special section takes several significant steps along the path to address this challenge.

REFERENCES

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school.* Washington, D.C.: National Academy Press.

Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research*, *77*(3), 334-372.

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In D. H. Schunk & B. Zimmerman (Eds.), *Handbook of self-regulation of learning and performance* (pp. 65-84). New York: Routledge.

Jonassen, D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm?. *Educational Technology Research and Development*, *39*(3), 5-14.

Roll, I., Wiese, E., Long, Y., Aleven, V., & Koedinger, K. R. (2014). Tutoring self- and co-regulation with intelligent tutoring systems to help students acquire better learning skills. In R*. Sottilare, A. Graesser, X. Hu, & B. Goldberg (Eds.), Design Recommendations for Adaptive Intelligent Tutoring*

Systems: Volume 2 - Adaptive Instructional Strategies (pp. 169-182). Orlando, FL: U.S. Army Research Laboratory.

Roll, I., Baker, R.S.J.d., Aleven, V., & Koedinger, K. R. (2014). On the benefits of seeking (and avoiding) help in online problem solving environment. *Journal of the Learning Sciences,* 23:4, 537-560, doi: 10.1080/10508406.2014.883977

Winne, P. H. (2011). A cognitive and metacognitive analysis of self-regulated learning. In B. J. Zimmerman and D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 15-32). New York: Routledge.

Winne, P. H. (in press, 2015). Leveraging big data to help each learner upgrade learning and accelerate learning science. *Teachers College Record, 117*(9).

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Mahwah, NJ: Lawrence Erlbaum Associates.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments

James R. Segedy, John S. Kinnebrew, & Gautam Biswas
Institute for Software Integrated Systems
Department of Electrical Engineering and Computer Science
Vanderbilt University
Nashville, Tennessee, USA
james.r.segedy@vanderbilt.edu

ABSTRACT: Researchers have long recognized the potential benefits of open-ended computer-based learning environments (OELEs) to help students develop self-regulated learning (SRL) behaviours. However, measuring self-regulation in these environments is a difficult task. In this paper, we present our work in developing and evaluating *coherence analysis* (CA), a novel approach to interpreting students' learning behaviours in OELEs. CA focuses on the learner's ability to seek out, interpret, and apply information encountered while working in the OELE. By characterizing behaviours in this manner, CA provides insight into students' open-ended problem-solving strategies as well as the extent to which they understand the nuances of their current learning task. To validate our approach, we applied CA to data from a recent classroom study with *Betty's Brain*. Results demonstrated relationships between CA-derived metrics, prior skill levels, task performance, and learning. Taken together, these results provide insight into students' SRL processes and suggest targets for adaptive scaffolds to support students' development of science understanding and open-ended problem-solving skills.

KEYWORDS: open-ended learning environments, self-regulated learning, learning analytics, coherence analysis, metacognition, computer-based learning environments

1 INTRODUCTION

For several years, researchers have sought to leverage the power of computer-based learning environments (CBLEs) to study aspects of students' self-regulated learning (SRL) behaviours (Sabourin, Shores, Mott, & Lester, 2013). SRL is an active theory of learning that describes how learners are able to set goals, create plans for achieving those goals, continually monitor their progress, and revise their plans when necessary. SRL is a multi-faceted construct: it involves emotional and behavioural control, management of one's learning environment and cognitive resources, perseverance in the face of difficulties, and social interactions to promote effective learning (Zimmerman & Schunk, 2011). For decades, researchers have recognized academic advantages for learners who are self-regulated (e.g., Bransford, Brown, & Cocking, 2000; Butler & Winne, 1995; Zimmerman, 1990), and devising techniques for automatically detecting and supporting students' development of self-regulation in CBLEs is an active area of research (Winters, Greene, & Costich, 2008).

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

SRL is particularly important for students working in *open-ended computer-based learning environments* (OELEs; Clarebout & Elen, 2008; Land, Hannafin, & Oliver, 2012), which provide a learning context and a set of tools for exploring, hypothesizing, and building solutions to authentic and complex problems. Such environments are demanding; they require students to wrestle simultaneously with their emerging understanding of a complex topic (e.g., ecosystems or macroeconomics), develop and utilize problem-solving skills to support their learning, and employ SRL processes for managing the complexity and open-ended nature of the learning task. As such, OELEs can *prepare students for future learning* (Bransford & Schwartz, 1999) by developing their ability to investigate and solve open-ended problems independently.

However, research with OELEs has produced mixed results. While some students with higher levels of prior knowledge and SRL skills show large learning gains as a result of using OELEs, many of their less capable counterparts experience significant confusion and frustration (Azevedo & Witherspoon, 2009; Hacker, Dunlosky, & Graesser, 2009; Kinnebrew, Loretz, & Biswas, 2013). Research examining the activity patterns of those students indicates that they typically make ineffective, suboptimal learning choices when they work independently toward completing open-ended tasks (Kinnebrew et al., 2013; Land, 2000; Mayer, 2004; Sabourin, Mott, & Lester, 2013).

Thus, an important goal of learning analytics research is to develop techniques for studying aspects of SRL and their manifestations in OELEs. These environments can provide a wealth of fine-grained process data, and the inferences made from such data necessarily depend on the analytic lens applied. In this paper, we present our work in developing and evaluating *coherence analysis* (CA), a novel approach to analyzing and interpreting student behaviour in OELEs. CA, an extension of our model-based approach to analyzing learner behaviour in OELEs (Segedy, Biswas, & Sulcer, 2014), focuses on the learner's ability to seek out, interpret, and apply information encountered while working in the OELE. By characterizing behaviours in this manner, CA provides insight into students' open-ended problem-solving strategies, as well as the extent to which they understand the nuances of the learning task they are currently completing.

To validate our approach, we applied CA to data from a recent classroom study with the *Betty's Brain* OELE (Kinnebrew, Segedy, & Biswas, 2014; Leelawong & Biswas, 2008). Results demonstrate the effectiveness of CA in 1) predicting students' task performance and learning gains and 2) identifying common problem-solving approaches among the students in the study. Further, the results demonstrate relationships between CA-derived metrics and students' prior skill levels, offering a potential explanation for students' problem-solving behaviours. Taken together, these results provide insight into students' SRL processes and suggest targets for adaptive scaffolds to support students' development of open-ended problem-solving skills.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

2 OPEN-ENDED LEARNING ENVIRONMENTS AND SELF-REGULATED LEARNING

OELEs focus on supporting learners' development of strategies for independently completing open-ended problem-solving tasks. They are typically designed "to support thinking-intensive interactions with limited external direction" (Land, 2000, p. 62) by providing a learning context and a set of tools for learning and problem solving. Some OELEs provide explicit goals, while others allow learners to define their own goals. Examples include hypermedia environments (e.g., Bouchet, Harley, Trevors, & Azevedo, 2013), modelling and simulation environments (e.g., Barab, Hay, Barnett, & Keating, 2000; van Joolingen, de Jong, Lazonder, Savelsbergh, & Manlove, 2005; Sengupta, Kinnebrew, Basu, Biswas, & Clark, 2013), and immersive narrative-centred environments (e.g., Clark et al., 2011; Spires, Rowe, Mott, & Lester, 2011). While OELEs may vary in the particular sets of tools they provide, they often include tools for 1) *seeking and acquiring* information, 2) *applying* that information to a problem-solving task, and 3) *assessing* the quality of the constructed solution. For example, students may be given the following task:

> *Use the provided simulation software to investigate which properties relate to the distance that a ball will travel when allowed to roll down a ramp, and then use what you learn to design a ramp suitable for wheelchairs at a local community centre. To test a solution, enter the details of your ramp into the system and press "Test."*

OELEs place students in a *self-regulatory context* in which they must utilize both cognitive and metacognitive processes to achieve success (Kinnebrew, Segedy, & Biswas,2014; Segedy et al., 2014). To accomplish this wheelchair task, for example, students must manage their own learning processes in order to 1) use the system's resources to learn about factors important to the design of ramps; 2) apply their knowledge to a problem-solving context by designing a wheelchair ramp; and 3) assess their developing understanding by testing their designs. As part of managing their learning processes, students need to plan their interactions with the system, monitor their progress toward completing their goals, and, when necessary, modify their problem-solving strategies.

2.1 Metacognition and Self-Regulated Learning

Metacognition (Brown, 1975; Flavell, 1976), when applied to learning, is a key component of SRL that describes the ability to reason about, manage, and redirect one's own approach to learning (Whitebread & Cárdenas, 2012). It is often broken down into two sub-components: knowledge and regulation (Schraw, Crippen, & Hartley, 2006; Young & Fry, 2008). Metacognitive knowledge refers to an individual's understanding of her own cognition and strategies for managing that cognition. Metacognitive regulation refers to how metacognitive knowledge is used for creating plans, monitoring and managing the effectiveness of those plans, and then reflecting on the outcome of plan execution in order to refine metacognitive knowledge (Veenman, 2011).

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

Metacognitive regulation is often considered a subset of SRL that deals directly with *cognition* without explicitly considering its interactions with emotional or motivational constructs (Whitebread & Cárdenas, 2012). Despite this, models of self-regulation are valuable in depicting key metacognitive processes. For example, Roscoe, Segedy, Sulcer, Jeong, & Biswas describe SRL as containing “multiple and recursive stages incorporating cognitive and metacognitive strategies” (2013, p. 286). Their description of SRL is summarized in Figure 1; it presents SRL as involving phases of orientation and planning, enactment and learning, and reflection and self-assessment.

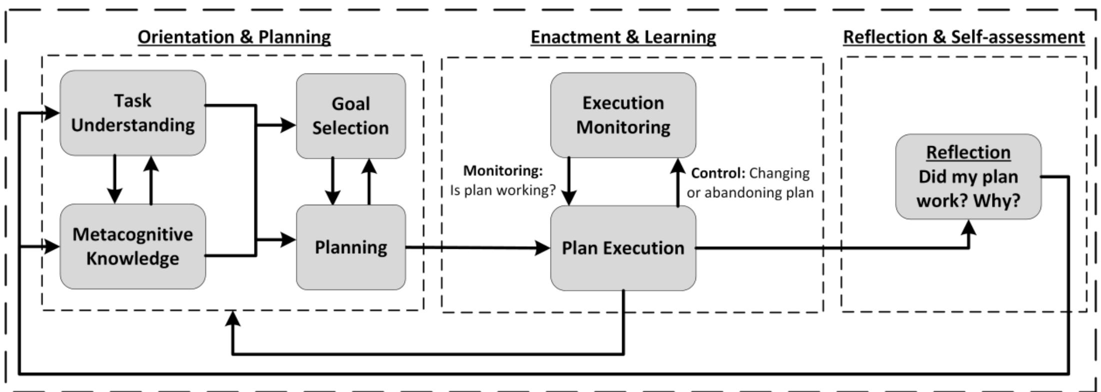


Figure 1. A model of SRL according to the description in Roscoe et al. (2013)

Students may start by orienting themselves to the task and formulating *task understanding* (i.e., an understanding of what the task is). A student’s task understanding is necessarily influenced by her metacognitive knowledge about her own abilities and available strategies for completing the task (Veenman, 2013). Together, these two sources of information, task understanding and metacognitive knowledge, provide a foundation that, in conjunction with other student attributes such as self-efficacy, governs students’ subsequent goal-setting and planning processes. Once a plan has been formulated, students begin executing it. As they carry out the activities specified in their plans, students may exercise metacognitive monitoring as they consciously evaluate the effectiveness of their plans and the success of the activities they are engaging in. The result of these monitoring processes may lead students to exercise metacognitive control by modifying or abandoning their plan as they execute it. Once a plan has been completed or abandoned, students may engage in reflection as they analyze the effectiveness of their plans and their planning processes. Such reflection may lead students to revise their metacognitive knowledge and task understanding.

2.2 Real-Time Measurement of Metacognition and Self-Regulated Learning

Measuring students’ self-regulation and metacognitive behaviour in real time is a difficult task; it requires developing systematic analysis techniques for detecting aspects of goal setting, planning, monitoring, and reflection in the context of the learning environment. In OELEs, such diagnoses involve identifying and assessing learners’ cognitive skill proficiency, interpreting their action sequences in

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

terms of their goals and the learning strategies they apply to achieve those goals, and evaluating their success in accomplishing their current tasks. The open-ended nature of OELEs further exacerbates the measurement problem; since the environments are learner-centred, they typically do not restrict the approaches that learners take to solving their problems. Thus, interpreting and assessing students' learning behaviours is inherently complex; they may simultaneously pursue, modify, and abandon any of a large number of both short-term and long-term approaches to completing their tasks.

Despite this complexity, researchers have developed several approaches to measuring aspects of self-regulation and metacognition in OELEs. For example, *MetaTutor* (Bouchet et al., 2013) adopts a very direct approach; it provides interface features through which students can externalize their SRL processes. By selecting an option from the *SRL Palette*, students indicate that they would like to, for example, judge their learning or activate their prior knowledge. To ensure that these features are used regularly, the system includes pedagogical agents that prompt students to engage in SRL processes through these features. This allows the system to capture students' SRL processes directly without having to make inferences based solely on their activities in the system.

Another approach employed in several OELEs involves developing a predictive, data-driven model for diagnosing constructs related to SRL in real time (e.g., engagement, frustration, or confusion). In some OELEs, such as *Crystal Island* (Sabourin, Shores, Mott, & Lester [2013]) and *EcoMUVE* (Baker, Ocumpaugh, Gowda, Kamarainen, & Metcalf, 2014), models have been created by first employing human coding to label students' log data with aspects of SRL and then using that labelled data to construct predictive models. For example, Sabourin, Mott, & Lester (2013) asked students to author "status updates" at regular intervals while using *Crystal Island*. These updates were later coded according to whether or not they included evaluations of the student's progress toward a goal, and this coded data, along with other features related to the student and her behaviours in the system, was used to build a predictive model of good versus poor self-regulation.

In other OELEs, researchers have developed theory-driven models of SRL and embedded those models into learning environments. For example, *EcoLab* (Luckin & du Boulay, 1999; Luckin & Hammerton, 2002) measures students' metacognitive awareness of their own ability by comparing the system's assessment of students' ability levels with the difficulty of the activities they choose to pursue. Should students choose activities that are too easy or too difficult, the system decreases its confidence in the student's metacognitive awareness and then prompts them to reconsider their choice (e.g., "*You should try a more difficult activity*"). As another example, Snow, Jackson, & McNamara (2014) measured the order and stability of students' behaviour patterns as they used iSTART-ME, a computer-based learning environment for helping students improve their science comprehension. In their model, lower levels of the information theoretic measure, *Shannon Entropy* (Coifman& Wickerhauser, 1992) were interpreted as indicative of ordered and self-regulated behaviours.

The approach presented in this paper is similar to this second set of environments: we have developed a

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

novel theory-driven approach to modelling learning behaviours in OELEs, called *coherence analysis* (CA), and applied that approach to the interpretation of log data from *Betty's Brain*. However, rather than focusing on a specific aspect of SRL (e.g., awareness of one's own ability), CA focuses on students' ability to seek for, interpret, and apply information encountered while working in the OELE. In doing so, CA models aspects of students' problem-solving skills and metacognitive abilities simultaneously. The approach is designed to be general, and should apply to OELEs beyond *Betty's Brain*, allowing researchers to study the coherence aspect of SRL in multiple contexts. The next section presents a brief overview of *Betty's Brain*, and the one following presents our coherence analysis approach.

3 BETTY'S BRAIN

Betty's Brain (Kinnebrew, Segedy, & Biswas, 2014; Leelawong & Biswas, 2008), shown in Figure 2, presents the task of teaching a virtual agent, Betty, about a science phenomenon (e.g., climate change) by constructing a causal map that represents that phenomenon as a set of entities connected by directed links representing causal relationships. Once taught, Betty can use the map to answer causal questions and explain those answers by reasoning through chains of links (Leelawong & Biswas, 2008). The goal for students using *Betty's Brain* is to construct a causal map that matches an expert model of the domain.

As an OELE, *Betty's Brain* includes tools for acquiring information, applying that information to a problem-solving context, and assessing the quality of the constructed solution. Students acquire domain knowledge by reading hypertext resources that include descriptions of scientific processes (e.g., shivering) and information pertaining to each concept that appears in the expert map (e.g., friction). As students read, they need to identify causal relations, such as "*skeletal muscle contractions create friction in the body.*" Students can then apply the learned information by adding the two entities to the causal map and creating the causal link between them (which "teaches" the information to Betty). In *Betty's Brain*, learners are provided with the list of concepts, and link definitions are limited to the qualitative options of increase (+) and decrease (-). Students can also add textual descriptions to each link.

Learners can assess their causal map by asking Betty to answer questions (using a causal question template) and explain her answers. To answer questions, Betty applies qualitative reasoning methods to the causal map (e.g.*, the question said that the hypothalamus response increases. This causes skin contraction to increase. The increase in skin contraction causes...*). After Betty answers a question, learners can ask Mr. Davis, another pedagogical agent that serves as the student's mentor, to evaluate her answer. If Betty's answer and explanation match the expert model (i.e., in answering the question, both maps utilize the same causal links), then Betty's answer is correct. Note that a link's textual description is not considered during this comparison.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

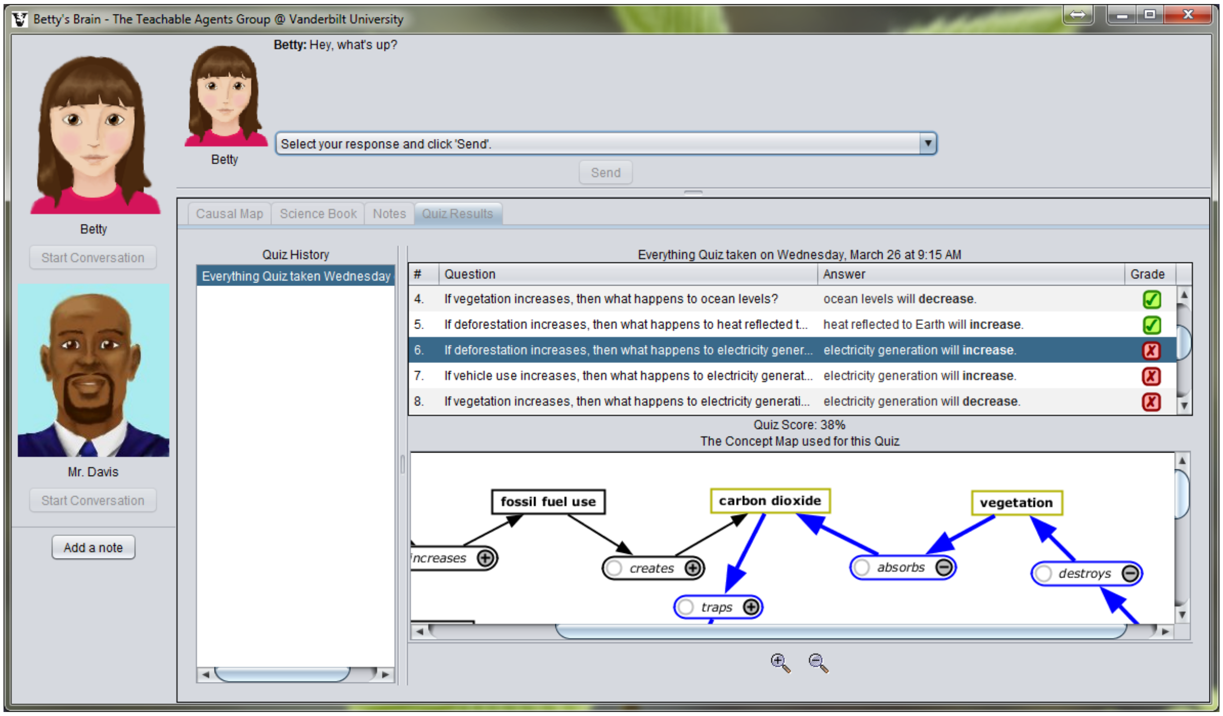


Figure 2. ***Betty's Brain*** **showing the quiz interface**

Learners can also have Betty take *quizzes*, which are a set of questions that can be answered by chaining together causal links in the map. An example quiz question is "*If vehicle use increases what happens to heat reflected back to earth*?" The question can be answered by following a chain of links from the concept *vehicle use* to the concept *heat reflected back to earth*, to derive the answer "*heat reflected back to earth will increase*." Quiz questions are selected dynamically by comparing Betty's current causal map to the expert map such that a portion of the chosen questions, in proportion to the completeness of the current map, will be answered correctly by Betty. The rest of her quiz answers will be incorrect or incomplete, helping the student identify areas for correction or further exploration. When Betty answers a quiz question correctly, students know that the links she used to answer that question are correct. Similarly, when Betty answers a question incorrectly, students know that at least one of the links she used to answer the question is incorrect. To help students keep track of correct links, the system allows students to annotate causal links as being correct.

4 COHERENCE ANALYSIS

This section describes our novel *Coherence Analysis* (CA) approach to learner modelling in OELEs. To develop this approach, we first performed a task-driven analysis of *Betty's Brain* (similar to cognitive task analysis; Chipman, Schraagen, & Shalin, 2000) to derive 1) the primary tasks that students should be able to complete to succeed in an OELE, and 2) the processes students must execute to complete those tasks. The result of this analysis is presented in the following section on "Cognitive and Metacognitive Problem-solving Processes in OELEs"; the CA approach is presented in "Modelling Learner Behaviours

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

with Coherence Analysis" below.

4.1 Cognitive and Metacognitive Problem-solving Processes in OELEs

As discussed above, learners working in OELEs need to access and interpret information, apply that information to constructing their problem solutions, and assess the quality of the constructed solutions using assessments provided in the system (Clarebout & Elen, 2008; Land et al., 2012; Land, 2000). These tasks and their more specific implementations in *Betty's Brain* have been incorporated into a task model (Figure 3) that specifies the tasks important for achieving success in *Betty's Brain*. The highest level of the model identifies the three broad classes of OELE tasks related to 1) *information seeking and acquisition*, 2) *solution construction and refinement*, and 3) *solution assessment*. Each of these task categories is further broken down into three levels that represent 1) general task descriptions common across all OELEs (according to the definition of OELE discussed above); 2) *Betty's Brain* specific instantiations of these tasks; and 3) interface features in *Betty's Brain* through which students can accomplish their tasks.

The directed links in the task model represent dependency relations. Information seeking and acquisition depends on one's ability to identify, evaluate the relevance of, and interpret information in the context of the overall task. Solution construction and refinement tasks depend on one's ability to apply information gained both by conducting information seeking tasks and by analyzing the solution assessment results. Finally, solution assessment tasks depend on one's ability to interpret the results of solution assessments as actionable information that can be used to refine the solution in progress. In order to accomplish these general tasks in *Betty's Brain*, students must understand how to perform the related *Betty's Brain* specific tasks by utilizing the system's interface features.

Identifying and evaluating the relevance of information describes the processes students employ as they observe, operate on, and make sense of the information presented in an OELE's information acquisition tools (Land, 2000; Quintana et al., 2004). Productively employing these processes requires an understanding of how to identify critical information and interpret it correctly. While learning in *Betty's Brain*, students need to identify sections of the hypertext resources that describe causal relations between entities in the problem domain. They must then correctly interpret those relations in order to create an accurate mental model of the science phenomena involved. Such processes can be difficult for learners; they may not have a firm grasp of causal reasoning mechanisms and the corresponding representational structures, or they may have difficulty in extracting the correct causal relations from the nuanced, technical writing style typical of science texts (McNamara, 2004). Further complications exist when the information contained in the resources conflicts with or challenges learners' prior inaccurate understandings of the problem domain. Land (2000) explains that in such situations, learners are resistant to restructuring their knowledge; instead, they often misinterpret the information in a way that supports their original conceptions.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

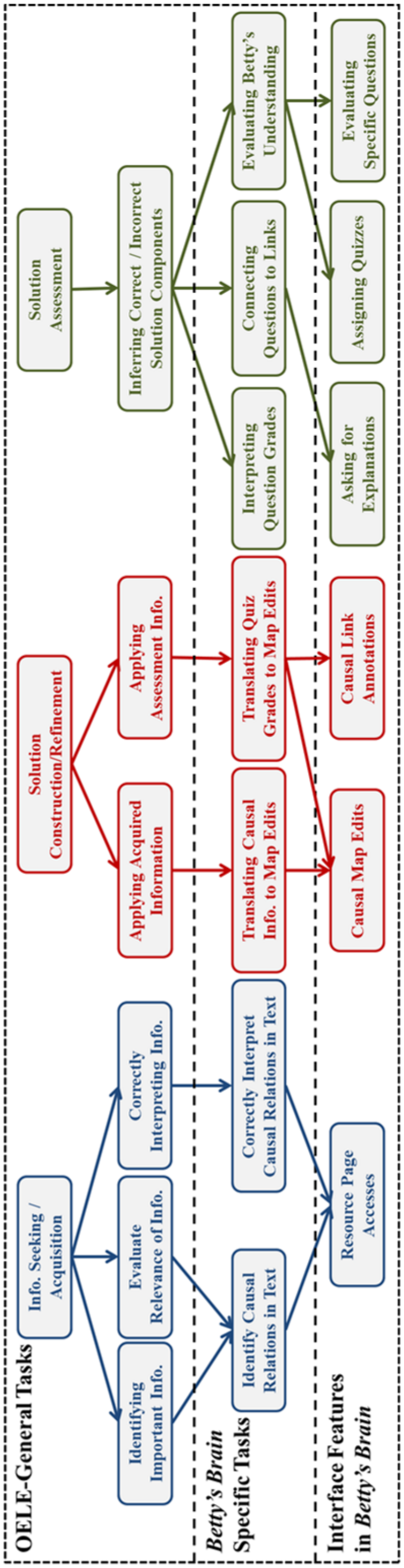


Figure 3. Task model for Betty's Brain OELE

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

When constructing problem solutions, learners utilize their developing understanding of the problem domain to make decisions about how to construct solutions. Productively employing these processes requires an understanding of 1) the structure of problem solutions; 2) the tools available for constructing solutions; and 3) methods for translating one's own understanding of the problem domain and solution requirements into explicit plans for solution construction in the OELE. In *Betty's Brain*, solutions take the form of a visual causal map, and accurately constructing such a map requires *representational fluency* (Suh & Moyer, 2007). Students must be able to convert causal information between and among the system's hypertext resources, their internal knowledge structures, and the causal map representation. Students unfamiliar with causal structures or how to represent knowledge using them will most likely struggle to succeed in completing the *Betty's Brain* learning task (Segedy, Kinnebrew, & Biswas, 2013; Roscoe et al., 2013).

Assessing the quality of constructed solutions describes the processes students employ as they submit their solutions to automated tests within the system and interpret the resulting feedback. In *Betty's Brain*, learners receive feedback in the form of Betty's quiz results: a list of questions that are either addressed appropriately by the model (i.e., Betty can answer these questions correctly), not addressed by the model (i.e., Betty cannot answer these questions), or addressed incorrectly by the model (i.e., Betty generates an incorrect answer to these questions). Learners are expected to use this information to determine which of their causal links are correct, which are incorrect, and what information is missing. This requires understanding how to interpret question grades, identify the causal links used to generate an answer, and evaluate the assessment information obtained via quizzes and question explanations. If students do not understand the relationship between a question, its quiz grade, and the links used to answer it, then they will most likely experience difficulty in obtaining meaningful information from quizzes.

The task model, along with the model of SRL presented in Figure 1, identifies and draws connections among the cognitive and metacognitive processes critical for learning in OELEs. Students need to leverage their metacognitive knowledge and task understanding in order to select intermediate goals for completing their tasks and then create plans for coordinating their use of skills and strategies in service of achieving those goals. Creating these plans requires understanding the purposes of, and relationships among, the tasks identified in the task model. Effective plans will utilize information gained from both information acquisition and solution assessment activities in order to build and refine a causal map that more closely approximates the expected solution. Because students are likely to make mistakes in constructing their solutions, they need to understand how to utilize the results of solution assessments to direct their thinking as they reflect on the sources of their errors.

4.2 Modelling Learner Behaviours with Coherence Analysis

The Coherence Analysis (CA) approach analyzes learners' behaviours by combining information from sequences of student actions to produce measures of *action coherence*. CA interprets students'

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

behaviours in terms of the information they encounter in the system and whether or not this information is utilized during subsequent actions. When students take actions that put them into contact with information that can help them improve their current solution, they have *generated potential* that should *motivate future actions*. The assumption is that if students can recognize relevant information in the resources and quiz results, then they should act on that information. If they do not act on information that they encountered previously, CA assumes that they did not recognize or understand the relevance of that information. This may stem from incomplete or incorrect understanding of the science topic, the learning task, and/or strategies for completing the learning task. Additionally, when students edit their map when they have not encountered any information that could motivate that edit, CA assumes that they are guessing[1]. These two notions come together in the definition of action coherence:

> *Two ordered actions ($x \to y$) taken by a student in an OELE are* ***action coherent*** *if the second action, y, is based on information generated by the first action, x. In this case, x provides* ***support*** *for y, and y is* ***supported*** *by x. Should a learner execute x without subsequently executing y, the learner has created* ***unused potential*** *in relation to y. Note that actions x and y need not be consecutive.*

The task model (Figure 3) implies two critical coherence relations: 1) applying information acquired from the hypertext resources to editing the map; and 2) applying inferred link correctness information (as obtained via quizzes) to editing the map. More specifically, an information seeking action (e.g., reading about a causal relationship) may generate support for a future solution construction action (e.g., adding that causal relationship to the map). Similarly, a solution construction action can be supported by a solution assessment action. An example of the latter situation occurs in *Betty's Brain* when a student deletes a causal link from their map after observing that Betty used that link to generate an incorrect answer to a quiz question.

CA assumes that learners with higher levels of action coherence possess stronger metacognitive knowledge and task understanding. Thus, these learners will perform a larger proportion of supported actions and take advantage of a larger proportion of the potential that their actions generate. In the analyses presented in this paper, we incorporated the following coherence relations:

- Accessing a resource page that discusses two concepts *provides support for* adding, removing, or editing a causal link that connects those concepts.
- Viewing assessment information (usually quiz results) that proves that a specific causal link is correct *provides support for* adding that causal link to the map (if not present) and annotating it

[1] In reality, students may be applying their prior knowledge; however, CA assumes that since students are typically wrestling with their emerging understanding of the domain, they should verify their prior knowledge before attempting a solution.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

as being correct (if not annotated)[2].

- Viewing assessment information (usually quiz results) that proves that a specific causal link is incorrect *provides support for* deleting it from the map (if present).[3]

Action coherence metrics measure whether or not learners' actions take advantage of previously encountered information. To measure whether or not a learner's actions *contradict* the information generated during previous activities, CA also incorporates measures of *action incoherence*:

> *Two ordered actions ($x \rightarrow y$) taken by a student in an OELE are **action incoherent** if the second action, y, is action coherent with the negation of information generated by the first action, x. In this case, x provides **negative support** for y, and y is **contradicted** by x.*

CA assumes that learners with higher levels of incoherence among their actions possess a weaker understanding of the science domain and the relations between different concepts in the domain. For example, when students have a misconception, they may add an incorrect link to their map due to their incorrect prior knowledge (Segedy, Kinnebrew, & Biswas, 2011). During solution assessment, they may obtain evidence that the link is incorrect and then delete it. However, in deleting the link, they may not restructure their own understanding of the problem domain, and, as a consequence, their established misconception may lead them to add the same incorrect link to the map at a later point in time. It is important to note that while incoherence is the natural complement to coherence in our analysis framework, space limitations compel us to focus on the primary (coherence-based) CA metrics in this paper, leaving a detailed analysis of students' action incoherence for future work.

Low levels of action coherence (and high levels of action incoherence) may indicate that learners do not possess sufficient task understanding or all of the metacognitive knowledge necessary for generating coherent plans. However, these CA-derived metrics are general measures of performance, and learners may exhibit low levels of action coherence for a variety of reasons. They may be struggling with 1) the task understanding and metacognitive knowledge underlying the coherence relations, 2) the related cognitive processes, and/or 3) their understanding of the domain content.

Our hypotheses in developing CA were as follows:

1. Students' CA-derived metrics would predict their learning and success in teaching Betty;
2. Students' prior levels of task understanding would predict their CA-derived metrics while using *Betty's Brain*.

[2] A quiz can only prove that a link is correct when it is already on Betty's map; however, a student can view an old quiz after deleting a link proven correct by that old quiz. In this case, viewing the old quiz would provide support for adding that link back to the map.

[3] A student can view an old quiz that proves a link is incorrect even if that link is not currently on their map.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

To explore these hypotheses, we applied CA to a recent classroom study of students using *Betty's Brain*. This study is presented next.

5 EXPERIMENTAL STUDY WITH BETTY'S BRAIN

The goals of this experimental study were to test the two hypotheses listed above. In addition, we sought to investigate whether or not CA-derived metrics would reveal common problem-solving approaches as a set of distinct behaviour profiles from the study data. The data presented in this paper comes from a larger experiment with *Betty's Brain* in which students completed two instructional units: climate change and human thermoregulation. During the first unit on climate change, students received different types of support from Mr. Davis. While analyses on this data have revealed statistically significant learning gains overall, they failed to reveal any significant effects of the type of support from Mr. Davis on students' learning and performance. Therefore, this paper focuses on the behaviour of all students, irrespective of the type of support received in the first unit, as they worked on the second unit, human thermoregulation, where students did not receive any support from the agents.

5.1 Participants

Ninety-nine 6th grade students from four mid-Tennessee science classrooms participated in the study. The participating school was an academic magnet school with competitive admission requirements. To enrol in this school, students need to pass all of their classes and achieve an average grade of B+ during the previous academic year. Demographic data for individual students were not released; however, the participants were typical of the school environment. Of the school's 701 students, 7.8% identified as Asian, 26.2% as Black, 4.0% as Hispanic, and 61.8% as White. None of the students was eligible for English as a Second Language programs, 1.4% of the students participated in special education programs, and 26.8% of the students were served by the Free and Reduced Price Lunch program. None of our study participants was enrolled in special education programs. One student was excused from the study due to an unrelated injury, therefore, the sample included data from 98 students.

5.2 Topic Unit and Text Resources

Students used *Betty's Brain* to learn about human thermoregulation when exposed to cold temperatures. The expert map, shown in Figure 4, contained 13 concepts and 15 links representing cold detection (cold temperatures, heat loss, body temperature, cold detection, hypothalamus response) and three bodily responses to cold: goose bumps (skin contraction, raised skin hairs, warm air near skin, heat loss), vasoconstriction (blood vessel constriction, blood flow to the skin, heat loss), and shivering (skeletal muscle contractions, friction in the body, heat in the body). The resources were organized into two introductory pages discussing the nervous system and homeostasis, one page discussing cold detection, and three pages discussing the three bodily responses to cold temperatures, one response per page. Additionally, a dictionary section discussed some of the concepts, one per page. The text was 15 pages (1,974 words) with a Flesch-Kincaid reading grade level of 9.0.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

5.3 *Betty's Brain* Interface and Features

The version of *Betty's Brain* used in this study was similar to the version presented above and illustrated in Figure 2. Students had access to hypertext resources, causal map editing tools, and the quiz feature. They could also ask Betty to answer questions and explain her answers, and they could ask Mr. Davis to grade Betty's answer to a specific question. However, Mr. Davis avoided grading answers where Betty used a single link to generate the answer. This was done to prevent students from gaming the system (Baker et al., 2006) by repeatedly adding a link to Betty's map and asking Mr. Davis if Betty's answer to a question using only that link was correct. If students were unsure of what to do, they could ask Mr. Davis to explain concepts important for success in *Betty's Brain* (e.g., *what are cause and effect relationships, and how do I find them while reading?*).

In addition, all students had access to a *Teacher's Guide* and a second set of hypertext resources that explained skills and strategies for seeking information, constructing the causal map, and assessing the causal map. For information seeking, the guide discussed how to identify causal links in text passages that use different presentation formats. For example, some passages present a causal link by describing what happens when the source concept *decreases* (e.g., "When cold detection decreases, the hypothalamus response also decreases"). For constructing the causal map, the guide explained how to use the causal map interface to add, edit, and remove concepts and links. It also explained the mechanics of causal reasoning (e.g., how to use a causal map to answer questions). For assessing the causal map, the guide discussed strategies for using quizzes, explanations, and Mr. Davis's answer evaluations to identify correct and incorrect links on Betty's map. In total, the guide contained 31 pages (6,247 words) with a Flesch-Kincaid reading grade level of 6.6.

5.4 Learning Assessments

Learning was assessed using a pre-test–post-test design with two parts: a set of computer-based exercises and a set of paper-and-pencil questions. The computer-based test consisted of 20 causal reasoning items, 10 causal link extraction items, and 14 quiz evaluation items designed to test students' understanding of the skills discussed in Section 4.1. The written test consisted of six multiple-choice science content items and four short answer science content questions. Further details on these assessments are available in Segedy (2014), Appendix C.

5.4.1 Causal Reasoning Items

Causal reasoning items (n=20) presented students with an abstract causal map (i.e., concepts were named *A*, *B*, etc…) and asked students to reason with the map to answer questions (e.g., "*If concept A increases, what will happen to concept B?*"). Each problem presented students with four possible choices: *B* will increase; *B* will decrease; *B* will not be affected; and it depends on which causal relations are stronger. Students were awarded one point for each question they answered correctly. An abstract causal map from this assessment is shown in Figure 5. Two causal reasoning items associated with this

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

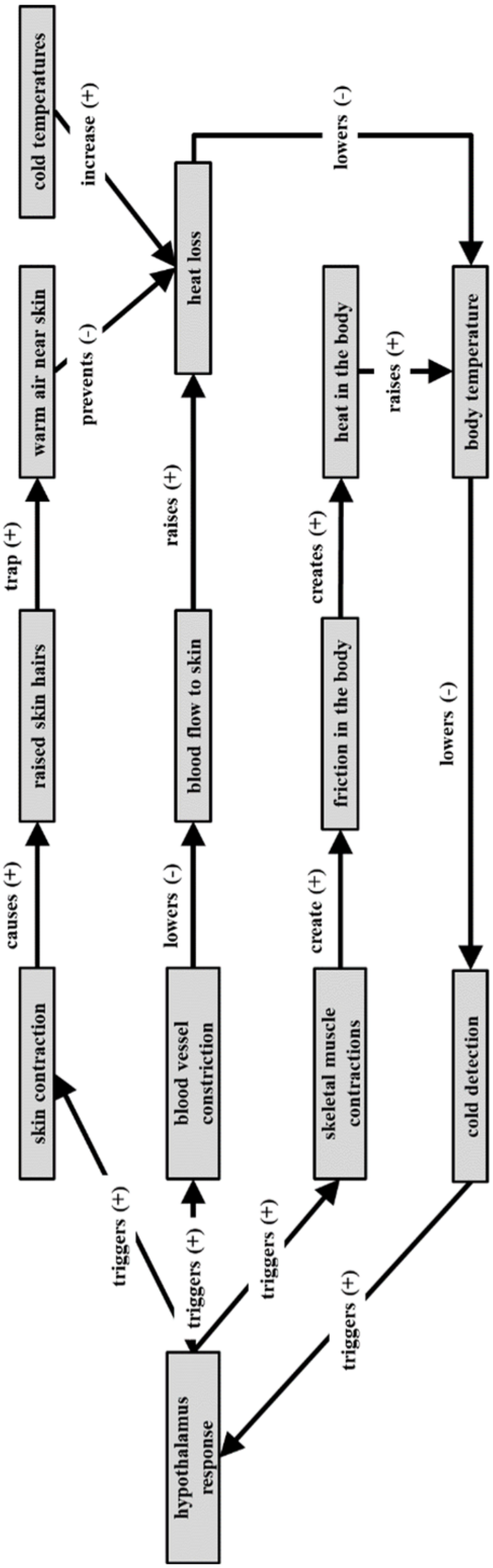


Figure 4: Thermoregulation expert map

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

map were as follows: 1) *If N increases, then what will happen to P?*; and 2) *If N decreases, then what will happen to P?* The causal reasoning items were found to be reliable on both the pre-test (α = 0.74) and the post-test (α = 0.81).

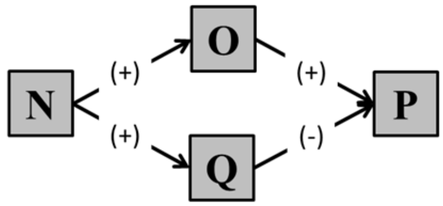


Figure 5. An example causal reasoning item.

5.4.2 Causal Link Extraction Items

Causal link extraction items (n=10) presented students with a text passage discussing the relationship between two abstract entities "Ticks" and "Tacks" (e.g., *"Tacks increase when Ticks decrease"*), and they were asked to choose the corresponding causal link described by that passage from the following choices: *Tacks increase Ticks*, *Tacks decrease Ticks*, *Ticks increase Tacks*, and *Ticks decrease Tacks*. Students were awarded one point for each correct answer. The ten causal link extraction items and their correct answers are included in Table 1. These items were found to be reliable on the pre-test (α = 0.71) and the post-test (α = 0.76).

Table 1. Causal link extraction items and their correct answers

	Text Passage	Correct Causal Link
1.	Tacks increase Ticks.	Tacks increase Ticks.
2.	A decrease in Ticks decreases Tacks.	Ticks increase Tacks.
3.	Tacks are decreased by Ticks.	Ticks decrease Tacks.
4.	Ticks are decreased by a decrease in Tacks.	Tacks increase Ticks.
5.	When Ticks increase, Tacks increase too.	Ticks increase Tacks.
6.	When Tacks decrease, Ticks increase.	Tacks decrease Ticks.
7.	When Tacks increase, Ticks decrease.	Tacks decrease Ticks.
8.	Ticks decrease when Tacks increase.	Tacks decrease Ticks.
9.	Tacks decrease when Ticks decrease.	Ticks increase Tacks.
10.	Ticks are increased when Tacks increase.	Tacks increase Ticks.

5.4.3 Quiz Evaluation Items

Quiz evaluation items (n=14) presented students with a quiz whose questions, answers, and grades were linked to an abstract causal map (see Figure 6). Students received one point for every problem in which they correctly annotated links as correct or incorrect according to the information in the quiz[4]. These

[4] See the "Betty's Brain" section for more information about how Betty's answers can be used to infer correct and incorrect causal links.

items were found to be somewhat reliable on both the pre-test (α = 0.62) and the post-test (α = 0.68).

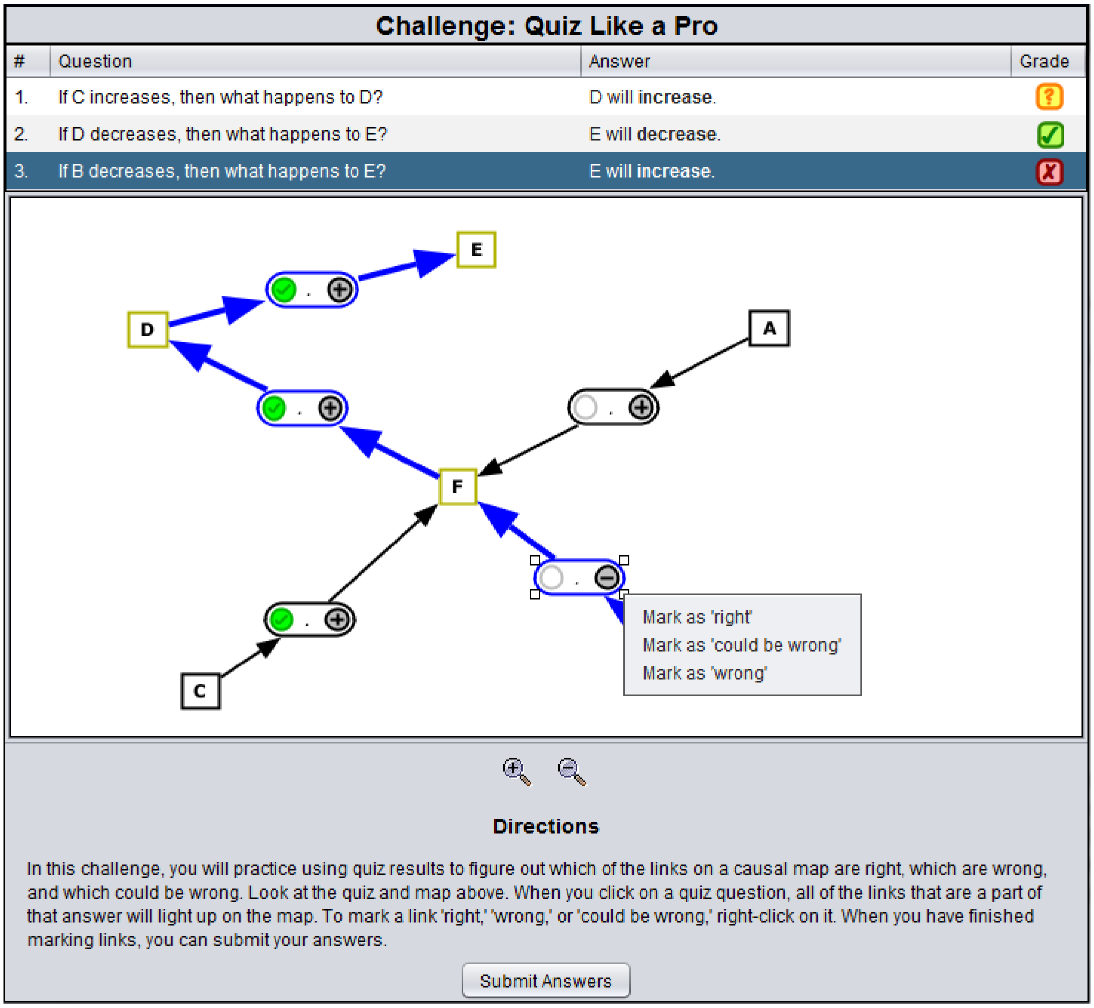


Figure 6. The quiz evaluation problem interface

5.4.4 Science Content Multiple-Choice Items

Science content multiple-choice items (n=6), each with four choices, tested students' knowledge of concepts, processes, and causal relations among concepts in the thermoregulation domain. These items are shown in Table 2.

5.4.5 Science Content Short Answer Items

Short answer items asked students to combine the causal relations among concepts to explain how the human body detects and responds to cold temperatures. The items are listed in Table 3. These items were coded by identifying the chain of causal relationships in learners' answers, and these chains were then scored by comparing them to the chain of causal relationships used to derive the answer from the expert map. One point was awarded for each causal relationship in the student's answer that was the same as or closely related to a relation specified in the expert map. For example, to answer question 1 correctly, students needed to note that skin contraction raises hairs near the skin (1 point), that these raised hairs trap warm air and keep it near the skin (1 point), and that this warm air near the skin reduces the rate at which heat is lost from the body (1 point). The maximum combined score for these questions was 11. Two coders independently scored five of the pre- and post-tests with over 85% agreement, at which point one of the coders individually coded the remaining answers and computed the scores.

Table 2. Science content multiple-choice items.

	Item
1.	What is thermoregulation?
2.	How does the hypothalamus regulate body temperature when the body gets too cold?
3.	How does shivering help regulate body temperature in cold temperatures?
4.	How do blood vessels change when the body is exposed to cold temperatures?
5.	How do raised skin hairs affect body heat?
6.	When a person drinks alcohol, their blood vessels become wider. How would drinking alcohol affect a person outside on a cold day?

Table 3. Science content short answer items.

	Items
1.	Explain, step-by-step, how skin contraction reduces heat loss from the body.
2.	Explain, step-by-step, how skeletal muscle contractions increase body temperature.
3.	Explain, step-by-step, how blood vessel constriction decreases heat loss from the body.
4.	Explain, step-by-step, how cold temperatures cause a hypothalamus response in the brain.

5.5 Log File Analysis

This version of *Betty's Brain* generated *event logs* that captured every *action* taken by the student, Betty, and Mr. Davis. A logged action corresponds to an atomic *expression of intent*, such as deleting a causal link or asking Betty to take a quiz. In addition, the logs contain information on every *view* that was displayed when the system was running. A logged view captures the information visible to a user during a specific time interval. For example, a view is created each time a page of the hypertext resources is visible. Unlike actions, which are distinct and orderable, views can overlap each other and span across multiple actions.

The log files provided the information required to calculate a measure of task performance for each student. By tracking the evolution of a student's causal map, we could compute how the student's causal map score changed over time. The *map score* at any point in time is calculated as the number of correct links (i.e., links that appear in the expert map) minus the number of incorrect links in the student's map. A student's *best map score* was computed as the highest map score they attained during

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

the intervention[5]. The log files also served as input to CA, which automatically calculated the following metrics for each student:

1. *Edit Frequency*: the number of causal link edits and annotations made by the student divided by number of minutes that the student was logged onto the system.
2. *Unsupported edit percentage*: the percentage of causal link edits and annotations not supported by any of the previous views that occurred within a five-minute window of the edit.
3. *Information viewing time*: the amount of time spent viewing either the science resource pages or Betty's graded answers. *Information viewing percentage* is the percentage of the student's time on the system classified as *information viewing time*.
4. *Potential generation time*: the amount of *information viewing time* spent viewing information that could support causal map edits that would improve the map score. To calculate this, we annotated each hypertext resource page with information about the concepts and links discussed on that page. *Potential generation percentage* is the percentage of *information viewing time* classified as *potential generation time*.
5. *Used potential time*: the amount of *potential generation time* associated with views that both occur within a prior five-minute window and also support an ensuing causal map edit. *Used potential percentage* is the percentage of *potential generation time* classified as *used potential time*.

Metrics one and two capture the quantity and quality of a student's causal link edits and annotations, where supported edits and annotations are considered to be of higher quality. Metrics three, four, and five capture the quantity and quality of the student's time viewing either the resources or Betty's graded answers. These metrics speak to the student's ability to *seek* and *identify* information that may help her build or refine her map (potential generation percentage) and then *utilize* information from those pages in future map-editing activities (used potential percentage). In these analyses, a page view generated potential and supported edits only if it lasted at least 10 seconds. Similarly, students had to view quiz results for at least 2 seconds. These cut-offs helped filter out irrelevant actions (e.g., rapidly flipping through the resource pages without reading them).

We also calculated a measure of *disengaged time*, which is defined as the sum of all periods of time, at least 5 minutes long, during which the student neither 1) viewed a source of information (i.e., science resources and quiz results) for at least 30 seconds; nor 2) added, changed, deleted, or annotated concepts or links. This metric represents periods of time during which the learner is *not measurably engaged* with the system. *Disengaged percentage* is the percentage of the student's time on the system classified as disengaged time.

[5] Not every student's final map was the best map she had created. For example, a student might decide to delete her entire map and start over near the end of the intervention.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

As a complement to the CA-derived metrics, we employed an information-theoretic differential sequence mining approach (Kinnebrew, Mack, Biswas, & Chang, 2014) to analyze students' action sequences further. This allowed us to identify sequential action patterns that best differentiated groups of students defined by the CA-derived metrics. The analysis considered the following student actions: 1) reading resource pages; 2) adding, removing, or editing causal links in the map (further distinguished by whether or not the edit improved the map score); 3) asking Betty to answer causal questions; 4) having Betty take quizzes; 5) asking Betty to explain her answer to a question; 6) creating, editing, and viewing notes; and 7) annotating links to keep track of their correctness. The derivation of action definitions from raw activity logs is discussed further in (Kinnebrew et al., 2013).

5.6 Procedure

The study was conducted over a period of approximately 6 weeks. At the beginning of the study, the first author spent 20 minutes introducing students to the causal reasoning methods used in the system. In particular, this lesson focused on understanding how to interpret and reason with both individual links and chains of links (i.e., a sequence of one or more links). Students then spent two weeks completing the climate change unit, which are not reported in this paper. At the beginning of the climate change unit, students were introduced to the software by Mr. Davis, who explained the task goal (i.e., teach Betty the correct causal map) and each of the *Betty's Brain* system features. As Mr. Davis explained a feature, he required students to use the feature in a specific way. For example, Mr. Davis asked students to add the concept "wolves" to their maps and he did not let them proceed until they had followed his instructions. Students practiced adding and deleting concepts and links, annotating links, asking Betty to take a quiz, and viewing Betty's quiz results. Mr. Davis also explained the importance of these features in successfully completing the *Betty's Brain* task. For example, he noted that students needed to identify relevant causal relations as they read the science resources and then teach these relations to Betty.

A two-week break separated the climate change and thermoregulation units. At the start of the thermoregulation unit, students spent two days completing the thermoregulation pre-tests. Students then spent four class periods using *Betty's Brain* to learn about thermoregulation. They completed the thermoregulation post-test approximately 1.5 weeks after the pre-test.

6 RESULTS

6.1 Learning and Performance Results

Table 4 summarizes the means (and standard deviations) of the students' pre-test and post-test scores, significant tests for gains, and a measure of effect size (Cohen's d). Overall, students exhibited strong gains on science multiple choice (d = 1.04) and short answer items (d = 1.55), suggesting that *Betty's Brain* facilitated students' ability to recognize and reason with relationships and definitions important for understanding thermoregulation. Conversely, students did not show statistically significant gains on the remaining three measures. However, during the first unit, students did exhibit statistically significant gains on causal reasoning ($p < 0.01$, $d = 0.27$), causal link extraction ($p < 0.01$, $d = 0.72$), and quiz evaluation items ($p < 0.01$, $d = 0.75$), suggesting that *Betty's Brain* facilitated students' ability to reason with causal maps, identify links in text passages, and interpret Betty's quiz results during the first unit.

Table 4. Means (and standard deviations) of assessment test scores

Measure	Maximum	Pre-test	Post-test	*t*	*p*	Cohen's *d*
Science Multiple-Choice	6	2.46 (1.07)	3.90 (1.63)	7.87	0.001	1.04
Science Short Answer	11	1.09 (1.14)	4.63 (2.55)	13.83	0.001	1.55
Causal Reasoning	20	11.44 (3.78)	11.61 (4.05)	0.72	0.474	0.07
Causal Link Extraction	10	6.06 (1.98)	6.09 (2.17)	0.22	0.824	0.02
Quiz Evaluation	14	5.27 (2.35)	5.63 (2.51)	1.79	0.076	0.18

Figure 7 displays the distribution of best map scores achieved by students ($\mu = 6.87$, $\sigma = 5.24$). As in previous studies with *Betty's Brain*, student performance on the task varied widely (Kinnebrew et al., 2013), with 37 students scoring below 5, 28 students scoring between 5 and 10, and 33 students scoring higher than 10. The maximum score students could obtain was 15, and 13 of the 98 students (13.3%) attained the maximum score.

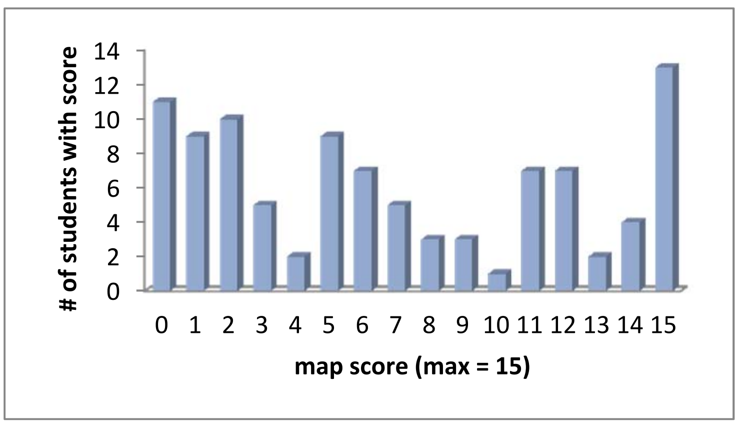


Figure 7. Map score distribution

6.2 Relationships between CA-Derived Metrics, Learning, and Performance

To investigate our first hypothesis, that students' CA-derived metrics would predict their learning and success in teaching Betty, we first analyzed correlations between the CA metrics, learning gains, and students' best map scores. The bottom row of Table 5 shows the overall descriptive statistics for the CA metrics. Students edited their maps fairly often (0.60 times per minute), and on average, 55.7% of these edits were supported. Students spent roughly one-third of their total time viewing information, but only 65.3% of this viewing time was spent on information that could support causal map edits. Students used, on average, a majority of the potential that they generated (62.3%), and they were mostly engaged in their learning (88.8%).

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

Table 5. CA metrics, learning gain, and performance correlations

Measure	Best Map	Sci. Content	Short Answer	Edit Freq.	Unsup. Edit %	Info. Viewing %	Potential Gen. %	Used Potential %	Disengaged %
Sci. Content	0.30**	1							
Short Answer	0.58**	0.29**	1						
Edit Freq.	0.56**	0.16	0.40**	1					
Unsup. Edit %	-0.49**	-0.12	-0.15	-0.14	1				
Info. Viewing %	0.11	-0.02	-0.07	-0.35**	-0.27**	1			
Potential Gen. %	0.41**	0.10	0.26**	0.14	-0.48**	-0.01	1		
Used Potential %	0.60**	0.16	0.26*	0.57**	-0.43**	-0.18	0.34**	1	
Disengaged %	-0.46**	0.05	-0.17	-0.49**	0.28**	-0.39**	-0.15	-0.38**	1
Mean (SD)	6.87 (5.21)	1.42 (1.89)	3.54 (2.52)	0.60 (0.36)	44.3% (21.9%)	37.0% (11.2%)	65.3% (17.1%)	62.3% (21.0%)	11.2% (10.5%)

Note. $*p \leq 0.05$. $**p \leq 0.01$.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

To test for the relationships between students' CA-derived metrics and their learning and performance, we calculated the pairwise correlations between students' CA metrics, their short answer learning gains, and their best map scores (Table 5). The results show that several of the CA metrics were significantly and moderately-to-strongly correlated with students' best map scores. Best map scores were positively correlated with edit frequency (r = 0.56), potential generation percentage (r = 0.41), and used potential percentage (r = 0.60). Best map scores were negatively correlated with unsupported edit percentage (r = –0.49) and disengaged percentage (r = –0.46). Students' CA metrics were also correlated with their short answer learning gains. More specifically, short answer learning gains were positively correlated with edit frequency (r = 0.40), potential generation percentage (r = 0.26), and used potential percentage (r = 0.26).

Interestingly, several CA-derived metrics were significantly correlated with each other. Students with higher levels of disengagement performed fewer edits per minute (r = –0.49), a higher proportion of which were unsupported (r = 0.28). They also spent a smaller percentage of their time viewing sources of information (r = –0.39) and took advantage of proportionally less of the information they encountered (r = –0.38).

To investigate these relationships further, we conducted multiple regression analyses to predict each of the learning gain measures and the best map score with the six CA metrics. The CA metrics predicted best map scores (F = 22.87, p < 0.001, R^2 = 0.601) and gains on short answer items (F = 4.544, p < 0.001, R^2 = 0.231) with statistical significance. With respect to map scores, the CA metrics of edit frequency (*Beta* = 0.56, t = 5.10, p < 0.01), information viewing percentage (*Beta* = 0.36, t = 3.44, p < 0.01), potential generation percentage (*Beta* = 0.20, t = 2.52, p = 0.01), and used potential percentage (*Beta* = 0.26, t = 2.80, p < 0.01) each added significantly to the prediction. With respect to short answer items, edit frequency (*Beta* = 0.56, t = 3.73, p < 0.01) and potential generation percentage (*Beta* = 0.25, t = 2.29, p = 0.02) each added significantly to the prediction. Conversely, the CA metrics did not predict gains on multiple choice (F= 1.60, p = 0.156, R^2 = 0.095), causal reasoning (F = 0.53, p = 0.782, R^2 = 0.034), causal link extraction (F = 1.02, p = 0.417, R^2 = 0.063), or quiz evaluation items (F = 1.20, p = 0.316, R^2 = 0.073) with statistical significance.

Together, these analyses provide potential insight into why particular students experienced more or less success. Negative correlations between unsupported edit percentage and information viewing percentage, potential generation percentage, and used potential percentage along with the positive correlation between unsupported edit percentage and disengaged percentage may suggest a behaviour profile characterized by disengagement, effort avoidance, and/or a difficulty in identifying causal links in the resources.

In summary, these results provide support for our first hypothesis. Students' CA-derived metrics were collectively predictive of their short answer learning gains and collectively and individually predictive of

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

their success in teaching Betty. Students who edited their maps more often, spent more time viewing information, viewed proportionally more relevant sources of information, and attempted to apply that information (via supported edits) achieved higher map scores. To gain further insight into this and other possible behaviour profiles, we performed a more comprehensive behaviour analysis in the next section.

To test our second hypothesis — that students' prior levels of task understanding would predict their CA-derived metrics while using *Betty's Brain* — we calculated correlations between CA metrics and students' pre-test skill levels (6). Most correlations are weak. However, some specific pre-test skill levels were weakly and significantly correlated with the CA metrics. Students with higher causal reasoning scores edited their maps more often (r = 0.23) and used proportionally more of the potential they generated (r = 0.27). Those with higher causal link extraction scores edited their maps more often (r = 0.34), had higher potential generation percentage (r = 0.26), and higher levels of used potential percentage (r = 0.21). Finally, students with higher quiz evaluation pre-test scores had higher levels of used potential percentages (r = 0.24).

To investigate this further, we conducted regression analyses to predict each of the six behaviour metrics from students' pre-test skill levels. The pre-test skill levels predicted edit frequency (F = 4.91, p = 0.003, R^2 = 0.136), potential generation percentage (F = 2.76, p = 0.047, R^2 = 0.284), and used potential percentage (F = 4.09, p = 0.009, R^2 = 0.115) with statistical significance. In these tests, only students' causal link extraction score added to the prediction of edit frequency with statistical significance (*Beta* = 0.29, t = 2.50, p = 0.01). Conversely, the pre-test skill levels did not predict unsupported edit percentage (F = 2.40, p = 0.072, R^2 = 0.071), information viewing percentage (F = 0.28, p = 0.843, R^2 = 0.009), or disengaged percentage (F = 1.93, p = 0.129, R^2 = 0.058) with statistical significance.

Table 6. Correlations between skill level at pre-test and behaviour metrics

	Causal Reasoning	**Causal Link Extraction**	**Quiz Evaluation**
Causal Link Extraction	0.53**	1	
Quiz Evaluation	0.17	0.20*	1
Edit Frequency	0.23*	0.34**	0.19
Unsupported Edit %	–0.14	–0.24*	–0.16
Info Viewing %	0.07	–0.01	–0.02
Potential Generation %	0.17	0.26**	0.16
Used Potential %	0.27**	0.21*	0.24*
Disengaged %	–0.14	–0.23*	–0.13

Note. $*p \leq 0.05$. $**p \leq 0.01$.

To summarize, we found only limited support for our second hypothesis. Students who were better at interpreting causal relations in text passages during the pre-test edited their maps somewhat more frequently and exhibited slightly higher levels of coherence by generating and using proportionally more potential.

6.3 Exploratory Clustering Analysis

In addition to testing our hypotheses, we performed an exploratory analysis to identify and characterize common behaviour profiles exhibited by students. These profiles may provide insight into students' SRL strategies as they worked toward completing the *Betty's Brain* task. For this analysis, we clustered students with a complete-link hierarchical clustering algorithm (Jain & Dubes, 1988; Murtagh, 1983), where each student was described by their set of CA metrics (listed in the Log File Analysis section above). The Euclidean distance between students' normalized CA metrics was used as the measure of dissimilarity among pairs of students. Clustering was performed using version 2.7 of the Orange data mining toolbox (Demšar, Curk, & Erjavec, 2013).

Figure 8 illustrates the resulting dendrogram. The analysis revealed five relatively distinct clusters containing 24, 39, 5, 6, and 24 students. Table 7 displays the means (and standard deviations) of the CA metrics for each cluster. The clustering results show distinct behaviour profiles among the 98 students in the study. Cluster 1 students (n=24) may be characterized as *frequent researchers and careful editors*; these students spent large proportions of their time (42.4%) viewing sources of information and did not edit their maps very often. When they did edit their maps, the edit was usually supported by recent activities ((unsupported edit percentage = 29.4%). Most of the information they viewed was useful for improving their causal maps (potential generation percentage = 71.4%), but they often did not take advantage of this information (used potential percentage = 58.9%). Cluster 2 students (n=39) may be characterized as *strategic experimenters*. These students spent a fair proportion of their time (33.5%) viewing sources of information, and, like Cluster 1 students, often did not take advantage of this information (used potential percentage = 62.6%). Unlike Cluster 1 students, they performed more map edits, a higher proportion of which were unsupported, as they tried to construct the correct causal model.

Cluster 3 students (n=5) may be characterized as *confused guessers*. These students edited their maps fairly infrequently and usually without support. They spent an average of 58.9% of their time viewing sources of information, but most of their time viewing information did not generate potential (potential generation percentage = 45.8%). One possibility is that these students struggled to differentiate between more and less helpful sources of information. Unfortunately, when they did view useful information, they often did not take advantage of it (used potential percentage = 23.1%), indicating that they may have struggled to understand the relevance of the information they encountered. Students in Cluster 4 (n=6) may be characterized as *disengaged from the task*. On average, these students spent more than 30% of their time on the system (more than 45 minutes of class time) in a state of disengagement. Like confused guessers, disengaged students had a very high proportion of unsupported edits, low potential generation percentage, and low used potential percentage. In addition, their information viewing percentage was much lower, though their edits per minute were slightly higher

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

than the confused guessers.

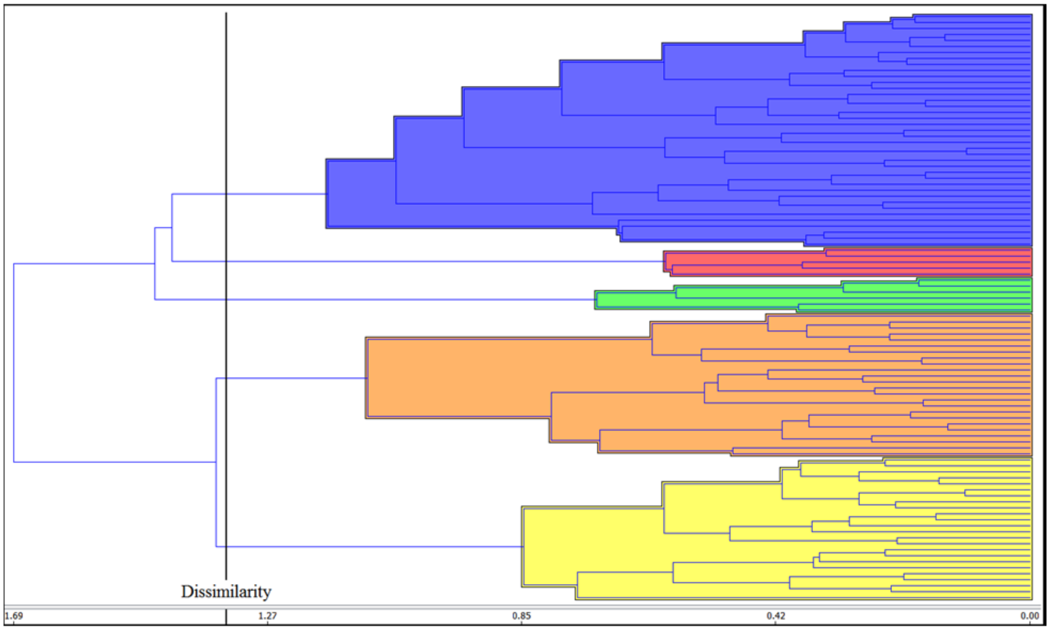


Figure 8. Dendrogram of students' thermoregulation behaviour profiles

Table 7. Means (and standard deviations) of CA-derived metrics by cluster

	Cluster	Edit Freq.	Unsup. Edit %	Info. View %	Potential Gen. %	Used Potential %	Disengaged %
1.	Res./Careful Editors (n=24)	0.30 (0.11)	29.4% (16.1%)	42.4% (11.0%)	71.4% (10.6%)	58.9% (15.4%)	15.7% (9.9%)
2.	Strat. Experimenters (n=39)	0.60 (0.23)	54.4% (14.8%)	33.5% (8.3%)	58.7% (18.9%)	62.6% (16.2%)	10.9% (7.4%)
3.	Confused Guessers (n=5)	0.21 (0.06)	73.5% (13.5%)	58.9% (7.7%)	45.8% (19.4%)	23.1% (12.6%)	4.8% (5.4%)
4.	Disengaged (n=6)	0.33 (0.11)	74.7% (17.4%)	27.0% (9.6%)	54.9% (9.3%)	28.0% (8.7%)	33.6% (8.4%)
5.	Engaged/Efficient (n=24)	1.04 (0.32)	29.1% (15.2%)	35.4% (8.6%)	76.8% (9.5%)	82.0% (9.0%)	3.1% (5.0%)

Cluster 5 students (n=24) are characterized by a high edit frequency (just over 1 edit per minute), and most of these students' edits (70.9%) were supported. Additionally, they spent just over one-third of their time viewing information, and over three-fourths of this time viewing information that generated potential. These students are distinct from students in the other four clusters in that they used a large majority of the potential they generated (82.0%) and were rarely in a state of disengagement (3.1%). In other words, these students appeared to be *engaged and efficient*. Their behaviour is indicative of students who knew how to succeed in *Betty's Brain* and were willing to exert the necessary effort.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

Table 8 shows the pre-test and post-test scores broken down by cluster. Cohen's *d* calculations were computed using equation 8 from Morris & DeShon (2002), which corrects for dependence among the means in within-subjects t-tests. Note that the small cluster sizes of Clusters 3 and 4 (n=5 and 6, respectively) necessitates caution when interpreting statistical tests performed on the data from these two clusters. Nevertheless, a 5x2 repeated-measures ANOVA run on the data revealed a main effect of cluster on short answer questions ($F = 5.09$, $p = 0.001$), causal extraction problems ($F = 2.82$, $p = 0.029$), and quiz evaluation problems ($F = 3.96$, $p = 0.005$). Tukey HSD-adjusted pairwise comparisons between the clusters showed that 1) Cluster 5's short answer scores were significantly higher than the scores of Cluster 1 ($p^{adjusted} = 0.001$) and higher, but not significantly higher, than the scores of Cluster 4 ($p^{adjusted} = 0.064$); 2) Cluster 5's causal extraction scores were significantly higher than the scores of Cluster 4 ($p^{adjusted} = 0.027$); and 3) Cluster 3's quiz evaluation scores were significantly lower than the scores of Clusters 1 ($p^{adjusted} = 0.036$), 2 ($p^{adjusted} = 0.030$), and 5 ($p^{adjusted} = 0.003$).

The analysis also revealed an interaction between time and cluster for short answer questions ($F = 4.86$, $p = 0.001$). Follow-up ANOVAs on the pre-test and post-test short answer scores did not reveal a significant effect of cluster on short answer pre-test scores ($F = 1.92$, $p = 0.102$), but they did find a significant effect of cluster on short answer post-test scores ($F = 5.70$, $p < 0.001$). Tukey HSD-adjusted pairwise comparisons between the clusters showed that Cluster 5's short answer post-test scores were significantly higher than the scores of Clusters 1 ($p^{adjusted} < 0.001$, $d = 1.20$), 2 ($p^{adjusted} = 0.018$, $d = 0.83$), and 4 ($p^{adjusted} = 0.043$, $d = 1.34$). These results show that Cluster 5 students, characterized as engaged and efficient, exhibited significantly higher short answer item gains when compared to most of the other student clusters.

Table 9 displays the means (and standard deviations) of the best map scores achieved by students in each cluster. Because the map scores exhibited a non-normal distribution, we tested for differences among clusters using a Kruskal-Wallis H test. The test identified a statistically significant difference in map scores between the clusters ($\chi^2(4) = 35.70$, $p < 0.001$), with a mean rank score of 41.58 for Cluster 1, 47.91 for Cluster 2, 24.20 for Cluster 3, 14.00 for Cluster 4, and 74.25 for Cluster 5. Follow-up Mann-Whitney tests between the groups showed that Cluster 5 students achieved higher map scores than students in Clusters 1 ($p < 0.001$, $d = 1.50$), 2 ($p = 0.001$, $d = 1.49$), 3 ($p = 0.001$, $d = 3.70$), and 4 ($p < 0.001$, $d = 4.05$). As with the learning results, engaged and efficient students performed significantly better than most other clusters.

To explore more detailed behaviour differences between the identified clusters, we employed the information-theoretic differential sequence mining approach described in Kinnebrew, Mack, Biswas, & Chang (2104) and the Log File Analysis section above. This approach identified the action patterns that best differentiate the five clusters, the top seven of which are presented in Table 10[6]. In previous work,

[6] The top differential activity patterns presented in Table 10 are those that included multiple distinct actions, leaving out trivial patterns of the same action simply repeated multiple times.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

we have argued that these patterns, when interpreted in the context of the task model, may be indicative of strategies students employ in building and refining their maps (Kinnebrew, Segedy, & Biswas, 2014). The pattern that most effectively differentiated clusters involved adding an incorrect link and then annotating an incorrect link as being correct (usually this was the same link just added). This pattern, most frequently used by strategic experimenters and disengaged students, suggests a potential misunderstanding of the use of link annotation functionality. According to the task model, students should only mark links as being correct once those links have been used in Betty's correct quiz answers. On average, the pattern was performed 5 to 6 times by strategic experimenters and disengaged students and less than once by students in the other clusters.

Table 8. Means (and standard deviations) of assessment test scores by cluster

Measure	Max	Cluster	Pre-test	Post-test	*t*	*p*	Cohen's *d*
Science Content	6	1 (Res./Careful Editors)	2.17 (1.40)	3.88 (1.51)	3.48	0.02	0.71
		2 (Strat. Experiment.)	2.59 (1.21)	3.72 (1.73)	4.52	0.01	0.76
		3 (Confused Guessers)	2.20 (1.79)	3.60 (1.95)	1.25	0.28	0.56
		4 (Disengaged)	2.83 (1.47)	3.33 (1.21)	0.75	0.49	0.31
		5 (Engaged/Efficient)	2.58 (1.32)	4.42 (1.59)	5.10	0.01	1.05
Short Answer	11	1 (Res./Careful Editors)	0.73 (1.13)	3.46 (2.50)	5.17	0.01	1.15
		2 (Strat. Experiment.)	1.44 (1.27)	4.53 (2.14)	8.75	0.01	1.47
		3 (Confused Guessers)	0.70 (0.45)	3.80 (2.68)	2.44	0.07	1.24
		4 (Disengaged)	0.67 (0.61)	3.42 (2.06)	3.51	0.02	1.84
		5 (Engaged/Efficient)	1.08 (1.00)	6.44 (2.46)	11.29	0.01	2.69
Causal Reasoning	20	1 (Res./Careful Editors)	10.67 (3.71)	11.17 (4.40)	0.89	0.38	0.19
		2 (Strat. Experiment.)	11.72 (3.75)	11.62 (3.77)	0.25	0.81	0.04
		3 (Confused Guessers)	9.60 (3.36)	9.80 (3.70)	0.22	0.84	0.10
		4 (Disengaged)	8.83 (1.33)	9.83 (1.72)	2.24	0.08	0.97
		5 (Engaged/Efficient)	12.79 (3.97)	12.88 (4.46)	0.20	0.84	0.05
Causal Extraction	10	1 (Res./Careful Editors)	5.79 (1.77)	5.88 (1.87)	0.31	0.76	0.07
		2 (Strat. Experiment.)	5.85 (1.71)	5.97 (2.13)	0.51	0.61	0.08
		3 (Confused Guessers)	5.80 (3.42)	6.20 (3.27)	1.00	0.37	0.45
		4 (Disengaged)	4.83 (1.33)	4.00 (1.55)	1.27	0.26	0.52
		5 (Engaged/Efficient)	7.04 (2.16)	7.00 (2.09)	0.20	0.84	0.04
Quiz Evaluation	14	1 (Res./Careful Editors)	5.21 (2.11)	5.75 (3.19)	1.25	0.23	0.29
		2 (Strat. Experiment.)	5.26 (1.94)	5.64 (1.86)	1.42	0.17	0.22
		3 (Confused Guessers)	2.00 (2.00)	3.00 (2.83)	1.41	0.23	0.73
		4 (Disengaged)	5.00 (2.76)	3.67 (2.58)	1.75	0.14	0.72
		5 (Engaged/Efficient)	6.08 (2.69)	6.54 (2.11)	0.92	0.37	0.19

Table 9. Means (and standard deviations) of map score metrics by cluster

Cluster	Best Map – Correct Links	Best Map – Incorrect Links	Best Map Score (max = 15)
1 (Res./Careful Editors)	6.42 (5.69)	0.96 (1.65)	5.46 (5.27)
2 (Strat. Experiment.)	7.54 (4.76)	1.41 (1.58)	6.13 (4.40)
3 (Confused Guessers)	2.80 (2.39)	0.80 (1.30)	2.00 (2.00)
4 (Disengaged)	1.17 (1.94)	0.00 (0.00)	1.17 (1.94)
5 (Engaged/Efficient)	12.67 (2.85)	0.75 (1.15)	11.92 (3.37)

Table 10. **Means (and standard deviations) of activity pattern frequency by cluster**

Pattern	Res./Careful Editors	Strat. Exper.	Confused Guessers	Disengaged	Engaged/ Efficient
1. [Add Incorrect Link] → [Mark Incorrect Link as Being Correct]	0.04 (0.20)	5.28 (8.81)	0.60 (0.89)	5.83 (6.52)	0.83 (3.29)
2. [Quiz] → [Remove Incorrect Link]	2.00 (4.24)	3.67 (6.18)	0.20 (0.45)	0.33 (0.52)	12.29 (12.33)
3. [Remove Incorrect Link] → [Quiz]	1.25 (1.33)	2.97 (5.73)	0.00 (0.00)	0.17 (0.41)	9.33 (8.65)
4. [Add Incorrect Link] → [Quiz] → [Remove Incorrect Link]	1.08 (4.27)	2.56 (5.53)	0.00 (0.00)	0.17 (0.41)	8.79 (10.47)
5. [Add Incorrect Link] → [Quiz]	4.54 (4.62)	7.72 (7.63)	2.20 (1.30)	2.50 (1.38)	17.54 (13.06)
6. [Check Quiz Answer Explanation] → [Remove Incorrect Link]	1.79 (1.61)	2.54 (2.87)	0.20 (0.45)	0.33 (0.82)	9.08 (6.40)
7. [Remove Incorrect Link] → [Add Incorrect Link] → [Quiz]	1.08 (1.06)	1.59 (2.58)	0.00 (0.00)	0.00 (0.00)	6.46 (6.11)

All of the other top differential activity patterns described in Table 10 involve a combination of map editing (specifically with respect to incorrect links) and quizzing, patterns that have been associated with successful performance in *Betty's Brain* (Kinnebrew et al., 2013). For example, patterns 2 and 6 are characteristic of *supported map edits based on quiz results*. In particular, pattern 6 is characteristic of exploring the quiz results more deeply by viewing Betty's answer explanation. Patterns 3, 5, and 7 are characteristic of *using quizzes to monitor progress*. After the student edits the map, they have Betty take a quiz in order to evaluate the effect of that edit on her quiz performance. Pattern 4 combines these two pattern types into a pattern characteristic of an *edit-and-check* strategy, in which students add a link to their map, use a quiz to monitor the effect of that link on Betty's performance, and then, upon discovering the link is incorrect, remove it from their maps.

All six of these activity patterns display similar relative use across the clusters, having the highest average frequency in engaged and efficient students, moderate frequencies in researchers/careful editors and strategic experimenters, and low frequencies in confused guessers and disengaged students.

Interestingly, these relative frequencies follow the same pattern as overall map scores and number of correct links in comparing performance across clusters, as illustrated in Table 9. In other words, clusters that used these behaviour patterns more often were ones that also achieved greater success in teaching Betty. Moreover, they also had lower average unsupported edit percentages and higher used potential percentages. This suggests, but does not prove, that when these patterns were employed by students, the quizzes provided support for subsequent edits. Altogether, these results indicate that students in more successful clusters were more likely to employ behaviours illustrating productive uses of the quiz for solution assessment.

7 DISCUSSION AND CONCLUSIONS

This paper presented *Coherence Analysis* (CA), a novel approach to measuring aspects of students' self-regulated learning (SRL) behaviours in open-ended learning environments (OELEs). CA focuses on the learner's ability to seek, interpret, and apply information encountered while working in an OELE. By characterizing behaviours in this manner, CA provides insight into students' open-ended problem-solving strategies, as well as the extent to which they understand the nuances of the learning task they are currently completing. We applied CA to data from a recent classroom study with *Betty's Brain* to test two hypotheses: 1) students' CA-derived metrics would predict their learning and success in teaching Betty; and 2) students' prior levels of task understanding would predict their CA-derived metrics while using *Betty's Brain*. Results showed some support for both hypotheses: CA-derived metrics were predictive of students' task performance and learning gains, and students' prior skill levels were (weakly) predictive of some of the CA metrics, suggesting a link between task understanding and effective open-ended problem-solving behaviours. In addition to testing these hypotheses, we applied a clustering analysis to characterize students based on their CA metrics, and this provided insights into common problem-solving approaches used by students in this study.

One important limitation of this work is the fact that we directly assessed students' task understanding (via their skills) during the pre- and post-tests without similarly assessing aspects of their metacognitive knowledge and regulation. Students with high task understanding may still exhibit difficulty in employing metacognitive processes, such as goal setting, planning, monitoring, and reflection. Future work should investigate the relationships between metacognitive knowledge, task understanding, and CA-derived metrics in OELEs. Another limitation is that the CA metrics were based on action coherence metrics without considering action incoherence. In future work, we will examine the relationships between students' learning, performance, action coherence, and action incoherence.

7.1 Coherence Analysis and SRL in OELEs

CA, as distinct from analyses of students' learning and task performance, provides insight into aspects of students' SRL behaviours, particularly in OELEs. Several of the behaviour profiles, identified using cluster analysis with the CA metrics as features, exhibited similar levels of prior knowledge, prior skill levels,

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

success in teaching Betty, and learning while using the system. CA helps us understand how different behaviours can result in the same level of performance and learning. In fact, one of the more interesting findings that emerged from this study is that CA metrics were able to distinguish groups of students, based on their behaviours, beyond what was possible when only focusing on learning gains and map scores. Certainly, students in the engaged and efficient cluster had higher prior skill levels, better map scores, and higher learning gains than students in most other clusters. However, it is harder to distinguish the remaining four clusters in terms of learning gains and performance. Confused guessers scored lower on quiz evaluation items when compared to researchers/careful editors and strategic experimenters. However, there were no other measurable differences in performance and learning gains. Despite this, these groups of students adopted distinct approaches to completing the Betty's Brain learning task, as measured by CA. Future work should investigate these profiles further with more detailed analyses of strategic behaviours. For example, further analysis of observed behaviour patterns and interviewing students with different behaviour profiles may help us understand the intentions that drove students' problem-solving strategy selections. Additional work should also look for interactions between CA-derived metrics and other aspects of SRL, such as affect and self-efficacy.

7.2 Implications for the Design of OELEs

One interesting set of findings from this study involves the predictive relationships between students' task understanding (as measured by their skill levels) and their behaviours (as measured by CA). The results seem to validate, at least to some extent, the *metacognitive knowledge dilemma* presented by Land (2000). This dilemma states that success in OELEs depends not only on students' metacognitive skills, but also on their understanding of the overall task and its components. In this study, students' prior skill levels were predictive of some CA metrics, which, in turn, predicted their success in teaching Betty. Specifically, students with higher task understanding had higher levels of coherence, and students with higher levels of coherence were more successful in their map-building tasks and demonstrated larger learning gains on short answer questions. Therefore, building coherence detectors into OELEs can provide a mechanism for first identifying low levels of coherence and then performing more targeted diagnosis of students' task understanding (perhaps via a method similar to rapid dynamic assessments; Kalyuga & Sweller, 2005). This mechanism, then, could identify and scaffold causes of poor SRL and problem-solving behaviours. For example, the system could provide students with opportunities to practice and develop their skills (Segedy et al., 2013) while explaining relevant problem-solving strategies.

Further, mining of behaviour sequences across identified clusters showed that students with more successful behaviour profiles were more likely to use the quiz productively (patterns 2 to 7 in Table 10). This is especially interesting given the fact that students' quiz evaluation skills were far less predictive of their behaviour than were their causal extraction skills. It makes sense to hypothesize that students cannot take advantage of the quiz functionality unless they can identify causal relations in text passages. Given this, scaffolding agents may support students by first helping them develop their information-

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

seeking skills, and then helping them to develop their solution assessment skills. Future work should investigate these relationships in more detail and how to provide the right scaffolding at the right time.

Another potentially powerful application of CA in OELEs, and one that we are particularly excited about, is presenting CA metrics to classroom teachers for evaluation and formative assessment. Ideally, teachers could use these reports to quickly and easily 1) understand learners' problem-solving approaches, 2) infer potential reasons for the levels of success achieved by students, and 3) make predictions about students' learning and task understanding. Moreover, teachers could use these reports to assign performance and effort grades and implement classroom and homework activities that target the aspects of SRL and problem solving with which students are struggling. However, additional research is required to understand how best to present and use this data with classroom teachers.

In future work, we plan to investigate the predictive power of additional CA-derived metrics via feature engineering and selection (Peng, Long, & Ding, 2005). In this study, we chose six metrics that successfully differentiated students and predicted aspects of their learning and performance. However, other CA-derived metrics may be better predictors. For example, it may be valuable to represent actions based on the *amount of support* they have from previous actions, rather than a binary measure of whether they do/do not have *any* support. As another example, it may be valuable to investigate CA metrics that incorporate more fine-grained aspects of how student behaviour *changes over time*. Ideally, this will allow us to study the development of SRL as students' task understanding and problem-solving skills improve in the *Betty's Brain* environment. Further, by studying aspects of coherence across multiple OELEs, we could also gain insight into how students generalize aspects of SRL and open-ended problem-solving strategies and skills over a more extended period.

ACKNOWLEDGEMENTS

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A120186 to Vanderbilt University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

REFERENCES

Azevedo, R., & Witherspoon, A. (2009). Self-regulated learning with hypermedia. In D. Hacker, J. Donlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 319–339). New York, NY: Taylor & Francis.

Baker, R. S. J. d., Corbett, A., Koedinger, K., Evenson, S., Roll, I., Wagner, A., Naim, M., Raspat, J., Baker, D., & Beck, J. (2006). Adapting to when students game an intelligent tutoring system. In M. Ikeda, K. D. Ashley, & T. Chan (Eds.), *Lecture Notes in Computer Science: Intelligent Tutoring Systems, 4053* (pp. 392–401). Berlin: Springer-Verlag.

Baker, R. S. J. d., Ocumpaugh, J., Gowda, S. M., Kamarainen, A. M., & Metcalf, S. J. (2014). Extending log-based affect detection to a multi-user virtual environment for science. In V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog, & G. Houben (Eds.), *Lecture Notes in Computer Science: User Modeling, Adaptation, and Personalization*, 8538 (pp. 290–300). Cham, Switzerland: Springer International Publishing AG.

Barab, S., Hay, K., Barnett, M., & Keating, T. (2000). Virtual solar system project: Building understanding through model building. *Journal of Research in Science Teaching, 37*(7), 719–756.

Bouchet, F., Harley, J., Trevors, G., & Azevedo, R. (2013). Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning. *Journal of Educational Data Mining, 5*(1), 104–146.

Bransford, J., Brown, A., & Cocking, R. (Eds.). (2000). *How people learn*. Washington, DC: National Academy Press.

Bransford, J., & Schwartz, D. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*(1), 61–101.

Brown, A. (1975). The development of memory: Knowing about knowing, and knowing how to know. In H. Reese (Ed.), *Advances in child development and behavior* (pp. 103–152). New York, NY: Academic Press.

Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281.

Chipman, S., Schraagen, J., & Shalin, V. (2000). Introduction to cognitive task analysis. In J. Schraagen, S. Chipman, & V. Shalin (Eds.), *Cognitive task analysis* (pp. 3–23). Abingdon, UK: Psychology Press.

Clarebout, G., & Elen, J. (2008). Advice on tool use in open learning environments. *Journal of Educational Multimedia and Hypermedia, 17*(1), 81–97.

Clark, D., Nelson, B., Chang, H., Martinez-Garza, M., Slack, K., & D'Angelo, C. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison and learning and affective outcomes for students in Taiwan and the United States. *Computers & Education, 57*(3), 2178–2195.

Coifman, R. R., & Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. Information Theory, IEEE Transactions on, 38(2), 713-718.

Demšar, J., Curk, T., & Erjavec, A. (2013). Orange: Data mining toolbox in Python. *Journal of Machine Learning Research, 14*, 2349−2353.

Flavell, J. (1976). Metacognitive aspects of problem solving. In L. Resnick (Ed.), *The nature of intelligence* (pp. 231–236). Hillsdale, NJ: Erlbaum.

Jain, A., & Dubes, R. (1988). *Algorithms for clustering data*. Upper Saddle River, NJ: Prentice Hall.

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development, 53*(3), 83–93.

Hacker, D., Dunlosky, J., & Graesser, A. (Eds.). (2009). *Handbook of metacognition in education*. New York, NY: Taylor & Francis.

Kinnebrew, J. S., Loretz, K., & Biswas, G. (2013). A contextualized, differential sequence mining method

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

to derive students' learning behavior patterns. *Journal of Educational Data Mining, 5*(1), 190–219.

Kinnebrew, J. S., Mack, D. L. C., Biswas, G., & Chang, C. K. (2014). A differential approach for identifying important student learning behavior patterns with evolving usage over time. *Proceedings of the Mobile Sensing, Mining and Visualization for Human Behavior Inference Workshop at the 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 281–292). Tainan, Taiwan. Retrieved from http://link.springer.com/chapter/10.1007%2F978-3-319-13186-3_27#page-1

Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and Learning, 9*(2), 187–215.

Land, S. (2000). Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development, 48*(3), 61–78.

Land, S., Hannafin, M., & Oliver, K. (2012). Student-centered learning environments: Foundations, assumptions and design. In D. Jonassen & S. Land (Eds.), *Theoretical foundations of learning environments* (pp. 3–25). New York, NY: Routledge.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education, 18*(3), 181–208.

Luckin, R., & du Boulay, B. (1999). Ecolab: The development and evaluation of a Vygotskian design framework. *International Journal of Artificial Intelligence in Education, 10*(2), 198–220.

Luckin, R., & Hammeron, L. (2002). Getting to know me: Helping learners understand their own learning needs through metacognitive scaffolding. In S. Cerri, G. Gouardres, & F. Paraguau (Eds.), *Lecture Notes in Computer Science: Intelligent Tutoring Systems, 2363* (pp. 759–771). Berlin: Springer-Verlag.

Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19.

McNamara, D. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*(1), 1–30.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105–125.

Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal, 26*(4), 354–359.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(8), 1226–1238.

Quintana, C., Reiser, B., Davis, E., Krajcik, J., Fretz, E., Duncan, R., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences, 13*(3), 337–386.

Roscoe, R., Segedy, J., Sulcer, B., Jeong, H., & Biswas, G. (2013). Shallow strategy development in a teachable agent environment designed to support self-regulated learning. *Computers & Education, 62*, 286–297.

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

Sabourin, J., Mott, B., & Lester, J. (2013). Discovering behavior patterns of self-regulated learners in an inquiry-based learning environment. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Lecture Notes in Computer Science: Artificial Intelligence in Education, 7926* (pp. 209–218). Berlin: Springer-Verlag.

Sabourin, J., Shores, L., Mott, B., & Lester, J. (2013). Understanding and predicting student self-regulated learning strategies in game-based environments. *International Journal of Artificial Intelligence in Education, 23*, 94–114.

Schraw, G., Crippen, K., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*(1), 111–139.

Segedy, J. R. (2014). *Adaptive scaffolds in open-ended computer-based learning environments* (Doctoral dissertation). Department of EECS, Vanderbilt University, Nashville, TN.

Segedy, J. R., Biswas, G., & Sulcer, B. (2014). A model-based behavior analysis approach for open-ended environments. *The Journal of Educational Technology & Society, 17*(1), 272–282.

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2011). Modeling learners' cognitive and metacognitive strategies in an open-ended learning environment. *Advances in Cognitive Systems: Papers from the AAAI Fall Symposium* (pp. 297–304). Menlo Park, CA: AAAI Press.

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2013). The effect of contextualized conversational feedback in a complex open-ended learning environment. *Educational Technology Research and Development, 61*(1), 71–89.

Sengupta, P., Kinnebrew, J., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K–12 science education using agent-based computation: A theoretical framework. *Educational and Information Technologies, 18*(2), 351–380.

Snow, E. L., Jackson, G. T., & McNamara, D. S. (2014). Emergent behaviors in computer-based learning environments: Computational signals of catching up. *Computers in Human Behavior, 41*, 62–70.

Spires, H., Rowe, J., Mott, B., & Lester, J. (2011). Problem solving and game-based learning: Effects of middle grade students' hypothesis testing strategies on learning outcomes. *Journal of Educational Computing Research, 44*(4), 453–472.

Suh, J., & Moyer, P. (2007). Developing students' representational fluency using virtual and physical algebra balances. *The Journal of Computers in Mathematics and Science Teaching, 26*(2), 155–173.

van Joolingen, W., de Jong, T., Lazonder, A., Savelsbergh, E., & Manlove, S. (2005). Co-lab: Research and development of an online learning environment for collaborative scientific discovery learning. *Computers in Human Behavior, 21*(4), 671–688.

Veenman, M. (2011). Learning to self-monitor and self-regulate. In R. Mayer & P. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (pp. 197–218). New York, NY: Routledge.

Veenman, M. (2013). Assessing metacognitive skills in computerized learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 157–168), Springer International Handbooks of Education, Vol. 26. New York, NY: Springer

(2015). Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*, *2*(1), 13–48.

Science+Business Media.

Whitebread, D., & Cárdenas, V. (2012). Self-regulated learning and conceptual development in young children: The development of biological understanding. In A. Zohar & Y. J. Dori (Eds.), *Metacognition in science education: Trends in current research* (pp. 101–132), Contemporary Trends and Issues in Science Education, Vol. 40. Netherlands: Springer Science+Business Media.

Winters, F., Greene, J., & Costich, C. (2008). Self-regulation of learning within computer-based learning environments: A critical synthesis. *Educational Psychology Review, 20*(4), 429–444.

Young, A., & Fry, J. (2008). Metacognitive awareness and academic achievement in college students. *Journal of the Scholarship of Teaching and Learning, 8*(2), 1–10.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3–17.

Zimmerman, B., & Schunk, D. (Eds.). (2011). *Handbook of self-regulation of learning and performance*. New York, NY: Routledge.

(2015). *Posterlet*: A Game-Based Assessment of Children's Choices to Seek Feedback and to Revise. *Journal of Learning Analytics*, *2*(1), 49–71.

Posterlet: A Game-Based Assessment of Children's Choices to Seek Feedback and to Revise

Maria Cutumisu, Kristen P. Blair, Doris B. Chin, Daniel L. Schwartz
Graduate School of Education, Stanford University, USA
cutumisu@ualberta.ca

ABSTRACT: We introduce one instance of a game-based assessment designed to measure students' self-regulated learning choices. We describe our overarching measurement strategy and we present *Posterlet*, an assessment game in which students design posters and learn graphic design principles from feedback. We designed *Posterlet* to assess children's choices to seek informative negative feedback and to revise their work. Middle-school students from New York and Illinois played *Posterlet* and then took a post-test, for an overall average of 17 minutes of interaction time. Results showed that the frequency of choosing negative feedback and revision correlated with learning graphic design principles from the game. Seeking negative feedback, but less so revision, further correlated with standardized achievement tests of reading and mathematics. Our research presents a first-of-kind behavioural measure of students' feedback and revision choices and their relations to learning. Within the design context of creating posters, we found correlational evidence that seeking negative feedback and revising are good behaviours for self-regulated learning, and we devised a way to measure these behaviours. This sets the stage for developing and evaluating models of self-regulated learning instruction that help students choose to seek feedback effectively and revise accordingly.

KEYWORDS: self-regulated learning, preparation for future learning, choice, assessment, game, learning, feedback, revision

1 INTRODUCTION

This article describes our initial efforts to develop and empirically validate game-based assessments of self-regulated learning (SRL). Here, we capitalize on the interactive, data-logging possibilities of learning technologies to capture process data about students' selection of positive or negative informative feedback. We demonstrate that the choice to seek constructive critical feedback correlates with greater learning within our specific assessment environment. It also correlates with multiple standardized measures of academic achievement. This work begins to realize Butler and Winne's proposal that "research on feedback and research on SRL should be tightly coupled" (1995, p. 248).

We begin by discussing general issues of assessing SRL and our general solution of choice-based assessments. We then briefly review the literature on feedback to identify the empirical gap that our assessment addresses. Next, we describe in some detail the focal assessment, called *Posterlet*, and we present results of a correlational study using a sample of convenience from New York and Illinois. Finally, we summarize the findings, their limitations, and possibilities for future work.

(2015). *Posterlet*: A Game-Based Assessment of Children's Choices to Seek Feedback and to Revise. *Journal of Learning Analytics*, *2*(1), 49–71.

1.1 Assessing Self-Regulated Learning

A major goal of formal and informal education is to prepare students to be autonomous learners who have the "will and skill" to learn without the strict guidance of a parent, teacher, or computer (Pintrich & De Groot, 1990, p. 38). The goal of fostering independent learners falls within the research traditions of SRL and self-directed learning. For example, Garrison states, "self-directed learning is also a necessity if students are to learn how to learn and become continuous learners" (1997, p. 29).

A frequently acknowledged impediment to developing both theories and practices of SRL is the lack of precise behavioural measures (e.g., Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007; Howard-Rose & Winne, 1993). The authors of the concluding chapter of the 2000 *Handbook of Self-Regulation* state, "current research relies heavily on self-report measures. Thus, more observational and performance measures relevant to self-regulation processes and outcomes are urgently needed. Because there is a fundamental problem with using self-reports and survey methods to demonstrate dynamic processes, we sorely need better ways to operationalize the self-regulation construct" (Zeidner, Boekaerts, & Pintrich, 2000, p. 757).

When we consider existing behavioural assessments, the situation is not ideal. Most tests are poorly suited to evaluating SRL effects. They measure whether students have mastered the content, but they do not capture the SRL behaviours that may or may not have produced the measured learning. They also miss the central concern of SRL instruction, which we view as preparing students for future learning (Bransford & Schwartz, 1999). An ideal assessment would present an environment where students have something to learn, and then evaluate the processes students undertake to complete that learning.

Existing preparation for future learning (PFL) assessments, however, have predominantly focused on knowledge rather than process outcomes. In the typical PFL study, researchers compare which of two forms of instruction better prepares students to learn from an expository resource, such as a lecture, a reading, or a worked example (e.g., Schwartz & Martin, 2004). This determines which instruction prepared students to learn from new explanations, but not the processes students employed to learn. The maturation of computer-based learning environments, as well as rapidly evolving data-analysis techniques, have opened new possibilities for the collection and analysis of student process data (e.g., Stevens & Thadani, 2007). To expand PFL assessments to measure SRL behaviours directly, we are developing simple games in which students are presented with challenges to solve, and we measure students' SRL choices.

1.2 Choice-Based Assessments

Independent learners need to make choices with the imperfect information they have at their disposal. For example, students need to choose what and how to learn. In many educational discussions, student choice is seen as a way to increase motivation and learning during instruction (Iyengar & Lepper, 1999). Here, we take choice as an outcome of instruction. Measuring choice as an outcome may be an important proxy for students' learning in the future, when they must learn on their own.

Our measurement of student choice follows an intent-to-treat logic. Intent-to-treat studies test both whether a particular course of action yields a desirable result and whether people will take that course of action. For example, intent-to-treat studies of a new medication not only ask whether the medication improves patient outcomes, but also whether patients will choose to take that medicine. Similarly, given the practical goals of SRL instruction, the desired outcome is the prescription of useful strategies that students also choose to use.

We design our assessment game environments to follow three measurement principles:

1) *Preparation for Future Learning*. As mentioned above, assessments need to include opportunities to learn, so we can measure whether students employ SRL to learn the available content.
2) *Choice-Based*. Choices about learning need to be free choices, not necessarily right or wrong choices. In other words, the assessment environment must be pedagogically agnostic (i.e., it cannot tilt children towards one choice or another) and allow children to advance in the game regardless of their learning choices.
3) *Typical Learning Performance*. Students often exert maximal performance for high-stakes tests, which may not reflect the choices they would make for everyday learning (Sackett, Zedeck, & Fogli, 1988; Klehe & Anderson, 2007). Therefore, we create assessment environments that do not feel like tests, under the assumption that typical performance is a better proxy for independent learning behaviours, where there is no test to drive performance.

Our game-based assessments share many similarities to other learning technologies that collect and analyze process data. However, there are also some notable practical differences. One primary difference is that we are designing stand-alone assessments. The assessments are not embedded within a specific curriculum, and therefore, they can be used to compare the effects of different courses of instruction. Many computer-based learning environments, such as intelligent tutoring systems (e.g., Koedinger, Anderson, Hadley, & Mark, 1997), measure learning as students work through lessons within the system. These systems can evaluate student knowledge growth over time, aiming to determine the best sequence of problems or tutor moves that will maximize learning from the tutor (Chi, VanLehn, & Litman, 2010). Recently, this work has further begun to consider affect (e.g., boredom and engaged concentration), gaming the system, and off-task behaviours (Fancsali, 2014; Baker et al., 2008; Baker, Corbett, & Koedinger, 2004; Baker, Corbett, Roll, & Koedinger, 2008; Roll, Baker, Aleven, & Koedinger, 2014). The primary purpose of the measurements is to determine ways to help students learn more from the system. This is of high importance, but it constitutes a different goal than examining whether various forms of SRL instruction prepare students to learn in new contexts.

For instance, in the context of an intelligent tutoring system (ITS), Roll and colleagues (Roll, Aleven, McLaren, & Koedinger, 2011a, 2011b) found that automated tutoring on help-seeking and self-assessment behaviours helped students transfer these behaviours to another lesson within an ITS environment. However, there were no available external assessments to determine whether these SRL behaviours would transfer beyond an ITS. Therefore, we hope to help researchers determine whether students transfer SRL behaviours from initial instruction to a new setting of learning by creating a suite of targeted choice-based assessments.

A second difference is that choice-based assessments have *a priori* measurement goals close to the surface rather than inferred through extensive data manipulation. Data-mining techniques have been employed to examine students' learning choices in games, often focusing on predicting future student moves or likelihood of success within the system (Peacock et al., 2013; Lee, Liu, & Popović, 2014; Snow, Allen, Russell, & McNamara, 2014). However, a communicative challenge of data-mining techniques is that the claim connecting behaviours to constructs often requires a complex chain of data transformations, difficult for stakeholders to interpret (e.g., a clustering algorithm that creates an abstract centroid in a multi-featured space). Hence, we have tried to close the distance between the raw behaviours and their assessment interpretations (e.g., by simply counting a student's frequency of choosing negative versus positive feedback). Ideally, making the construct more transparent will aid in educator decision making.

A third difference is that we try to keep the assessments relatively short. In one assessment study, Shute, Ventura, Bauer, and Zapata-Rivera (2009) had an *a priori* assessment goal of measuring student persistence using the physics learning game, Newton's Playground (Ventura & Shute, 2013). They demonstrated convergent validity by correlating their assessment results against existing measures of persistence. However, the assessment depended on several hours of student data, in part because the goal was to adapt the game dynamically to address individual student persistence needs. Given the many demands on classroom time, a multi-hour assessment designed to measure one aspect of SRL is unlikely to scale well. Thus, we have opted for assessments of 10–15 minutes. This increases the flexibility of deployment, although we do pay the price of having less information about any given student.

Ultimately, we aim to build a suite of choice-based assessments to examine the transfer of SRL behaviours, including the choices to engage in critical thinking (Chi, Schwartz, Chin, & Blair, 2014), to read to learn (Chin, Blair, & Schwartz, 2015), and to persist after failure (Chase, Chin, Oppezzo, & Schwartz, 2009). Ideally, these and more choice-based assessments can indicate whether specific educational experiences foster independent learners, the main desiderate of SRL research.

Here, we describe research on a game-based assessment called *Posterlet*. We designed *Posterlet* to measure two behaviours purported to be important for learning: students' *choices to seek negative feedback* and *to revise* (Cutumisu, Blair, Schwartz, & Chin, 2015). We posed three primary research questions to explore the relation between learning choices and learning outcomes:

- Do choices to seek negative versus positive feedback, and choices to revise, correlate with learning from the game?

- Do feedback and revision choices within the game correlate with broader learning outcomes outside the game (i.e., standardized achievement outcomes)?
- Do patterns of choice differ by school, suggesting that SRL choices can be influenced by experience?

2 LEARNING FROM FEEDBACK

The literature on feedback is voluminous and permits many different frameworks (e.g., based on types of feedback [Mory, 2004] or based on functions of feedback [Butler & Winne, 1995]). Different research traditions use different criteria for what counts as feedback, because feedback may refer to any signal relevant to one's thoughts, behaviours, or products. In behaviourist theory, feedback strictly refers to positive or negative reinforcement (e.g., reward and punishment). There was even a period when researchers debated whether "knowledge of results" comprised feedback (i.e., information indicating how to adjust behaviour [Schmidt, Young, Swinnen, & Shapiro, 1989]). Conversely, in control theory, feedback only refers to information that indicates the degree of discrepancy between a current and a desired state (Powers, 1978). In the SRL tradition, further distinctions arise when feedback can refer to internally generated signals, as might be gained from self-monitoring during self-explanation activities (Butler & Winne, 1995), or to external feedback generated by the environment or by another person (Zimmerman, 1990; Okita & Schwartz, 2013), as well as whether the feedback is task- or person-directed (Hattie & Timperley, 2007).

Blurring across the many possible distinctions, Kluger and DeNisi (1998) performed a useful meta-analysis comprising 131 studies. Feedback improved performance by 0.4 standard deviations on average compared to no feedback controls. Of special importance to our investigation, one-third of the studies actually found feedback to be worse than no feedback at all. The prevailing explanation for the benefits and drawbacks of feedback emphasize negative feedback. The authors found that negative feedback was more effective for continued performance than positive feedback, presumably because positive feedback indicates that one has done enough, whereas negative feedback indicates the need for change. Yet, despite the signal for a need to change and learn, negative feedback runs the risk of triggering ego threat issues that lead people to shut down rather than heed the feedback (Hattie & Timperley, 2007). This suggests that students' attitudes towards negative feedback could have large implications for learning. Zimmerman highlighted this possibility among his list of three critical features of students' self-regulated learning: "their use of self-regulated learning strategies, their *responsiveness to self-oriented feedback about learning effectiveness*, and their interdependent motivational processes" (1990, p. 6; italics added for emphasis). Hence, we thought it important to develop an assessment that measured student behaviours and learning with respect to positive and negative feedback. This way, we can closely examine the relation of negative and positive feedback to learning, and we can determine whether there are important individual SRL differences.

It is noteworthy that the feedback literature was launched from the behaviourist tradition (Thorndike, 1927). Nearly all learning studies of external feedback use supervised feedback, where the teacher, experimenter, or computer decides when and how to deliver feedback. Learners do not have independent control — the feedback arrives without choice. However, in many situations, people need to choose to seek feedback. For example, when writing a manuscript, authors may (or may not) ask their peers to read it before submission. They may choose peers who they know provide generally supportive comments or they may choose peers who are more likely to provide strong critique.

There is scant literature on people's feedback-seeking behaviours. For example, to our knowledge, research on the effects of immediate versus delayed feedback and their impact on learning, self-monitoring, and motivation has not included an opportunity for students to choose between immediate or delayed feedback (e.g., Schooler & Anderson, 1990; Lepper, Woolverton, Mumme, & Gurtner, 1993). Work on self-monitoring (e.g., self-explanation [Chi, 2000]) emphasizes individuals' abilities and willingness to detect inconsistences in their own thoughts, not their choice to receive feedback. Closer to the question is the work by Roll and colleagues (Roll, Baker, Aleven, & Koedinger, 2014) on help-seeking behaviours in ITSs, examining students' tendencies to ask for the right answer rather than to fix their understanding based on partial hints. Their research examines how learner choices vary as a function of the informative value of the hints, rather than of the feedback valence. Corbett and Anderson (2001), using a cognitive tutor context for Lisp programming, examined the timing of tutorial advice. They investigated whether various combinations of control over error correction and feedback/help (immediate/on-demand, as well as right/wrong, informative, or none) influenced learning outcomes. However, they only found that the various tutor conditions, when combined, improved performance compared to a system that just provided right/wrong information after the submission of a complete program. An empirical question remains whether help seeking and feedback seeking belonged within the same theoretical SRL construct. Creating an assessment of the choice to seek positive or negative feedback could help refine and differentiate how these behaviours fit within theories of SRL.

The most directly relevant prior work on choices to seek positive and negative feedback comes from consumer research. Finkelstein and Fishbach (2012) compared novices and experts (true domain experts and self-perceived experts). They found that novices sought more positive feedback, whereas experts sought more negative feedback. The demonstration that greater (perceived) expertise correlates with greater selection of negative feedback points to possible causal mechanisms (e.g., self-efficacy). Unfortunately, they did not measure learning outcomes, so the relation to SRL remains undetermined.

In sum, there are reasons to believe that behaviours and attitudes regarding negative feedback influence learning, but there is no evidence to date whether independent choices about seeking feedback are important for learning. Thus, we decided to investigate this question by creating *Posterlet*.

3 THE ASSESSMENT ENVIRONMENT: *POSTERLET*

Posterlet is a game-based assessment that enables students to design posters for their school's Fun Fair. The environment offers students the choice between negative and positive feedback to help them learn about graphic design principles. It also measures whether students choose to revise their work after feedback. Behaviours after feedback, such as revising and help seeking, can be important aspects of learning because they enable students to practice the correct skill. On the other hand, revision may be primarily useful for improving the final product, but not for learning.

Figure 1 summarizes a cycle of *Posterlet*. On each of the three cycles of the game, students start the round by choosing an activity booth: basketball toss, science of bubbles, estimation station, pie-baking, or electricity (Step 1).

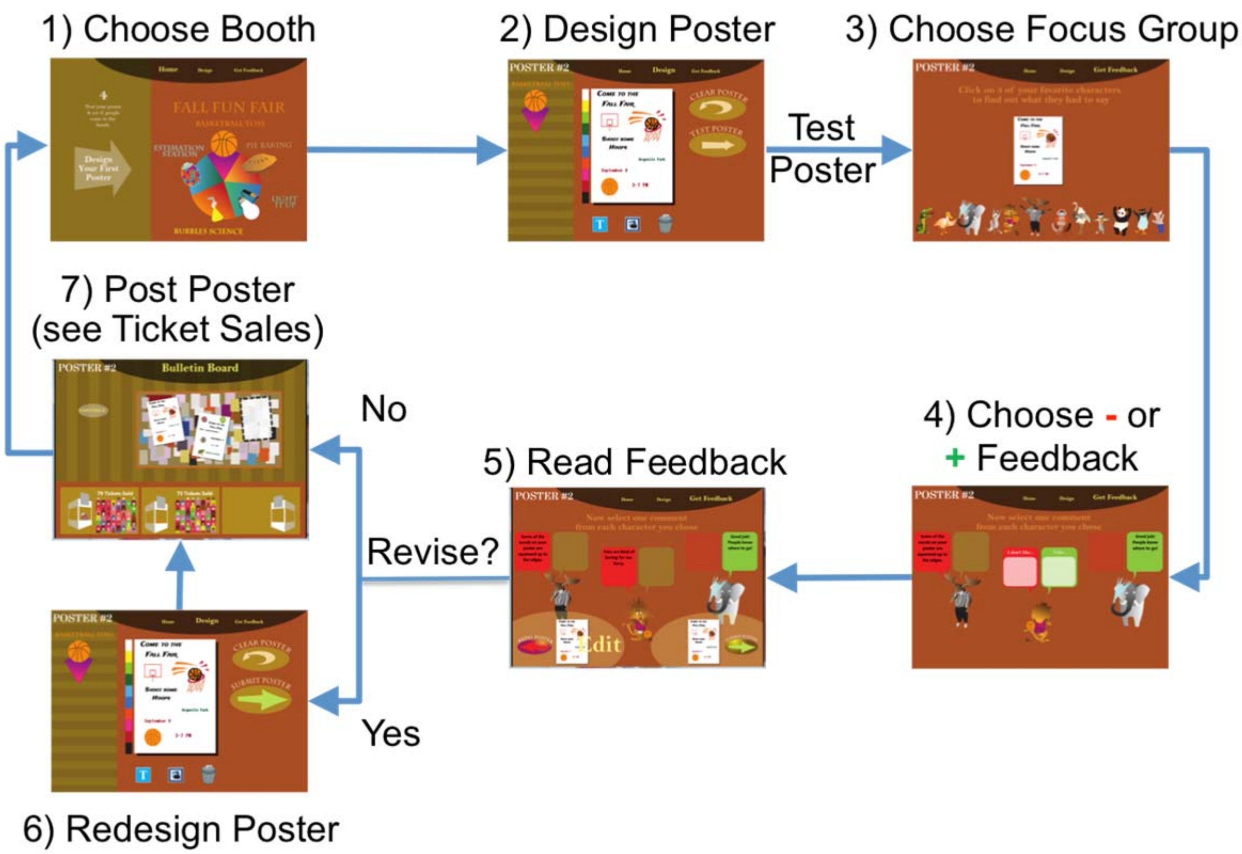


Figure 1. *Posterlet*: a game to assess students' learning choices in a design context. This diagram shows the state of the poster design cycle for the second poster.

In Step 2, students design a poster for their chosen booth. Students can select predefined phrases from a text palette and images from an image palette. In addition to placing text and images wherever they choose on the poster's canvas, students may customize the appearance of text (size, font, alignment, and colour), resize the images, and change the poster's background colour.

After designing their poster, students select three focus-group members from a cast of twelve animal characters (Step 3). Figure 2 shows how students choose either positive or negative feedback from each member of their focus group (Step 4).

(2015). *Posterlet*: A Game-Based Assessment of Children's Choices to Seek Feedback and to Revise. *Journal of Learning Analytics*, *2*(1), 49–71.

Figure 2. Students may choose either positive ("I like…") or negative feedback ("I don't like…") from each of three characters they selected from the focus group. Here, the student has chosen one positive and one negative feedback, and has not yet chosen feedback from the third character.

Here, feedback is delivered by computer characters so, in a mild sense, it constitutes social feedback. Both the negative and the positive feedback refer to the execution of graphic design principles, not to artistic flair (see Table 1). Positive and negative feedback are equally informative and non-directive. In *Posterlet*, negative feedback is non-punishing. Moreover, we are measuring the effects of the feedback's valence (positive versus negative), not of the feedback's informational content (informative versus uninformative), on learning.

Table 1. Examples of positive and negative feedback phrases for each type of feedback (informative or uninformative) and each category (information, readability, or space use)

Type	Category	Positive	Negative
Informative	Information	It's nice that the poster says how much the booth costs.	You didn't say how much the booth was.
	Readability	Your poster has big letters. Really easy to read.	People need to be able to read it. Some of your words are too small.
	Space Use	Great job on getting your pictures away from the edge.	Your picture is squished up next to the edge.
Uninformative	Generic	Yay! Fairs are fun!	Hmm, I don't really like fairs.

After reading the feedback (Step 5), students must choose whether to revise their poster (Step 6) or submit it (Step 7). Once students submit their poster, they can see how many people buy tickets for their booth. The game tracks and displays a poster quality score as the number of tickets sold at each booth, potentially providing extra engagement and motivation for students to improve their scores. Students complete this poster design cycle for three booths. The students do not have an opportunity to seek feedback on a revised poster. Therefore, in total, they have nine decisions for positive or negative feedback (three per poster), and three decisions for revision (one per poster).

While students may learn simply through the process of designing and reflecting on their posters, the feedback provides specific information for learning 21 graphic design principles gathered from a consulting professional graphic artist (Figure 3).

INFORMATION	READABILITY	SPACE USE
location absent	text cut off by bounding box	space used by graphics out of range
date absent	text size small	top half empty
time absent	text style unreadable	bottom half empty
ticket price absent	text & graphics overlap	graphics touching edge
booth description absent	text and other text overlap	text flush with edge
graphics irrelevant	graphics size small	text flush with other text
text absent	text contrast low	-
images absent	-	-

Figure 3. The three categories of the 21 graphic design principles used by the feedback system to generate feedback: information, readability, and space use

These principles correspond to the three categories of informative feedback: pertinent information, readability, and space use. A graphical analysis system embedded in the game evaluates in real-time whether each principle is satisfied for any given poster. For instance, to detect a principle about colour contrast, we implemented an algorithm that computed the contrast and relative luminance[1] between the text colour and the background colour. The system generates poster-specific feedback, employing a prioritization scheme to select which principle to emphasize in the feedback. The scheme ensures a balanced coverage of the three feedback categories and does not repeat prior feedback. In situations when all eligible design principles have been exhausted, the system provides an uninformative generic phrase (see Table 1).

[1] http://www.w3.org/TR/2008/REC-WCAG20-20081211/#relativeluminancedef

(2015). *Posterlet*: A Game-Based Assessment of Children's Choices to Seek Feedback and to Revise. *Journal of Learning Analytics*, *2*(1), 49–71.

4 THE EMPIRICAL INVESTIGATION

A unique feature of choice-based assessments, compared to knowledge-based assessments, is the burden to validate choices (i.e., to establish whether some SRL choices are in fact better than others for learning). Unlike assessments of knowledge, where the correct answer for 2+2 is 4 and not 5, choices face a special challenge. The reason becomes apparent upon a close read of Pintrich and De Groot's seminal SRL article in which students self-reported their uses of cognitive strategies. One of the survey items asked, "When I study for a test I practice saying the important facts over and over to myself" (1990, p. 40). This is a poor cognitive strategy for learning, yet the researchers considered it an indicator of effective cognitive strategy use. To avoid this type of mistake, it is important to have evidence that a proposed SRL strategy is indeed good for learning. In the case of choosing negative feedback, there is no relevant prior literature, so we need to show that it is better for learning, at least in this instance. This is the major goal of the current research.

Several hundred middle-school students completed *Posterlet*. We logged their choices regarding feedback and revision. We correlated these choices with their uses of graphic design principles in their posters and their abilities to judge a poster on a subsequent post-test. We also correlated *Posterlet* behaviours with standardized achievement scores to determine if the choices *within Posterlet* have a relation to academic outcomes that occur *outside Posterlet* in very different contexts.

4.1 Method

4.1.1 Participants and Procedures

Data were collected in the spring of 2013 from students at two public middle schools in urban settings; one school was in New York City (NYC), the second in Chicago. These two schools implement a game-based pedagogy designed to instil a variety of 21st-century skills and attitudes. *Posterlet* was one of several assessments administered by external school evaluators. We do not have access to the results of the other assessments. In New York City, we collected game logs from 278 children in Grades 6 to 9 (mean age=12.1 years, SD=1.0). Of these children, 231 further completed the post-test of graphic design principles. We received standardized test scores for 119 students. From those students who indicated their ethnicity (n=137), 17% were African-American, 47% Caucasian, 29% Hispanic, and 7% identified as "other." In Chicago, we collected the logs of 203 children in Grades 6 to 8 (mean age=12.2 years, SD=0.7). Of these children, 194 completed the post-test. We received standardized test scores for 65 students. For those who self-identified, the ethnicity breakdown was 63% African-American, 8% Caucasian, 27% Hispanic, and 1% "other." Finally, we also collected log files for an additional 36 children from these two schools, without indication of the particular school to which they belonged. Standardized scores were unavailable for these 36 children and only 29 of them completed the post-test.

To maximize the available data for each analysis, we used the subset of students who had complete data for each specific research question (rather than only analyzing students who had complete data on all measures). For example, to analyze the relations among feedback, revision, and poster performance, we

(2015). *Posterlet*: A Game-Based Assessment of Children's Choices to Seek Feedback and to Revise. *Journal of Learning Analytics*, *2*(1), 49–71.

began with the full sample. However, to analyze the relations among feedback, revision, and the post-test, we used the subset of students who completed *Posterlet* and the post-test.

In their regular classrooms, participants individually played the *Posterlet* game by designing posters of their choosing for three of the five Fun Fair booths. They then took the subsequent post-test. We collected data online automatically from both the game and the post-test. NYC students spent an average of M=12.2 minutes (SD=5.9) on *Posterlet* and an average of M=3.2 minutes (SD=2.7) on the post-test. Chicago students spent an average of M=14.9 minutes (SD=6.2) on *Posterlet* and an average of M=4.0 minutes (SD=2.3) on the post-test. Not all students completed the posters or the post-test due to time considerations. We were not present during *Posterlet* administration, so we relied on game duration to remove students who were ±2 SD from the mean. Our logic was that we should be conservative and eliminate any student who may not have followed the proctors' rules. This pruned the data pool by 9.1%: *Posterlet* corpus (n=272 for NYC, n=172 for Chicago) and post-test corpus (n=226 for NYC, n=163 for Chicago). Our results persisted even when we included the outliers in the analyses.

4.1.2 Measures

We employed measures of students' learning-relevant choices, learning of the graphic design principles, and academic achievement.

4.1.2.1 Measures of Learning-Relevant Choices

We focus on two learning choices: *negative feedback* and *revision*. Negative feedback refers to the number of times a student chose the negative feedback option ("I don't like…"). Students had nine total feedback choices, so the number of positive feedback choices ("I like…") is necessarily the inverse. Revision tallies the number of times students chose to revise their posters out of a maximum of three (one per poster).

4.1.2.2 Measures of Learning the Graphic Design Principles

We collected two indicators of how well students learned the graphic design principles: *Poster Quality* and *Post-test*. Poster Quality is a performance measure computed from the quality of the student posters and was determined by the same graphical analysis system that generated feedback. For each poster, the system scored each of the 21 graphic design principles by providing one point for each good feature (i.e., a feature always used correctly on that poster), subtracting one point for each poor feature (i.e., a feature used incorrectly on that poster), and providing zero points for features that were inapplicable or not present. Each poster score is the sum of the scores for each of the 21 graphic design principles and ranges from -21 to 21. The scores for the final poster on each round (revised poster, or initial poster if no revision) were combined to create an overall poster quality score.

(2015). *Posterlet*: A Game-Based Assessment of Children's Choices to Seek Feedback and to Revise. *Journal of Learning Analytics*, *2*(1), 49–71.

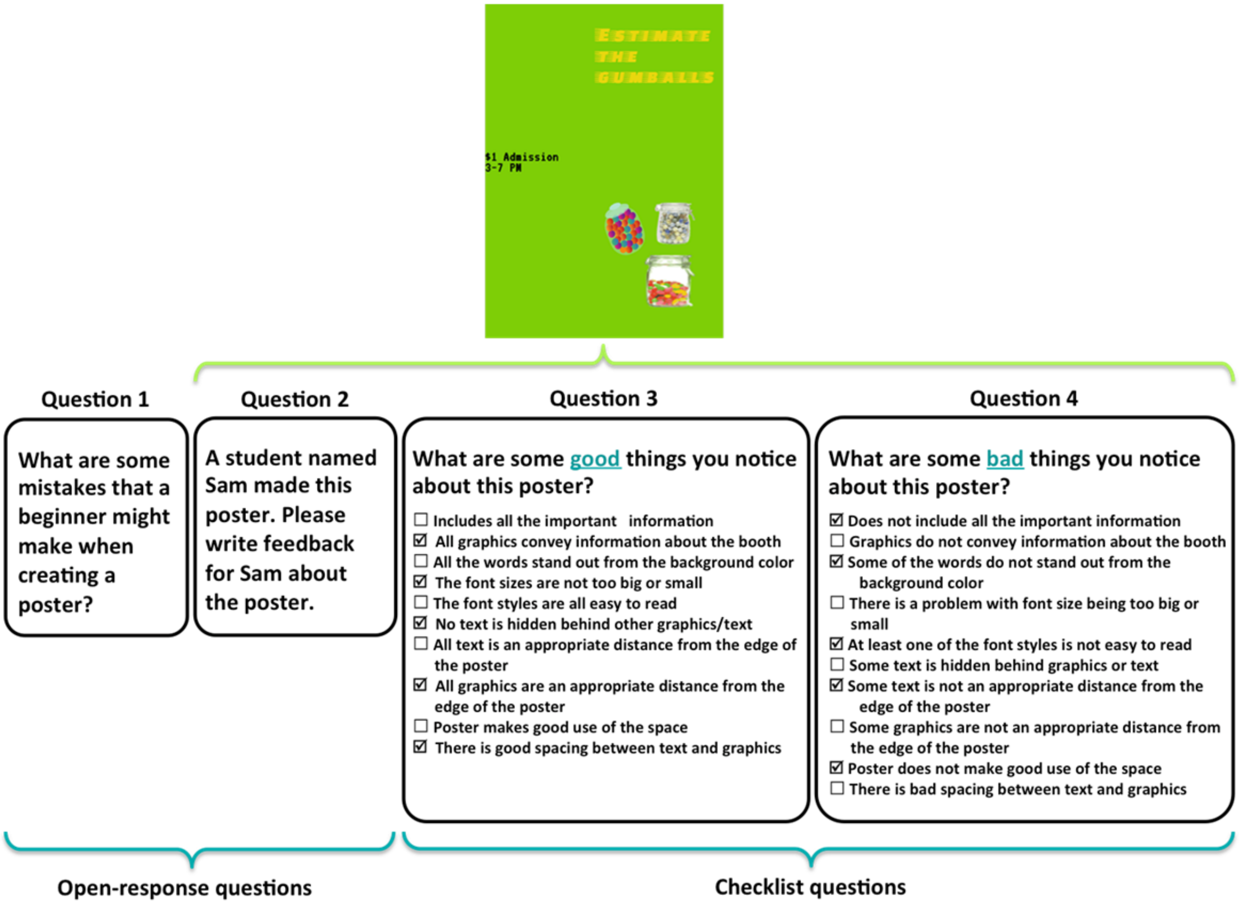


Figure 4. The four post-test questions: the items checked are the correct answers for Question 3 and Question 4, respectively

The post-test comprised four questions, as shown in Figure 4. The first open-response question asked students to list mistakes a beginner might make when creating a poster. The second open-response question provided a target poster and asked students to write feedback. We scored students' responses by counting the number of graphic design principles implied in each response (max=21). Open responses were coded individually by two evaluators with an inter-rater reliability score of r>.8. Two checklist items referred to the same target poster. In one item, students checked all the good things about the poster. In the other item, students checked all the bad things about the poster. Students received one point for correct answers that they did not otherwise contradict across the "what is good" and "what is bad" questions. The answers to these two questions are complementary, five being correct and five being incorrect for each question. We computed a total post-test score by adding the normalized scores (Z-scores) for each of the four questions.

4.1.2.3 Measures of Academic Achievement

For a subset of the students, we received Mathematics and English Language Arts achievement scores based on their respective state standardized tests: New York State Testing Program (NYSTP) and Illinois

(2015). *Posterlet*: A Game-Based Assessment of Children's Choices to Seek Feedback and to Revise. *Journal of Learning Analytics*, *2*(1), 49–71.

Standards Achievement Test (ISAT). The fact that the students from the two schools had different standardized achievement tests makes direct comparison of achievement data across schools impossible. Nevertheless, if we find that the correlations between the choice variables and the standardized achievement scores are similar across schools and tests, this would indicate a potentially stable relation between choice behaviours within *Posterlet* and multiple broad measures of school learning.

4.2 Results

Across the game, students improved their poster quality, which we take as evidence that they learned; round 1=8.4 (SD=5.6), round 2=9.4 (SD=5.2), round 3=9.6 (SD=5.3); $F(2,471)=14.9$, $p<.001$. The question then is whether their choices regarding negative feedback and revision correlated with learning from the game and, more far reaching, with learning outcomes in school.

We describe the results in four sections, which we preview here. The first section considers whether in-game choices correlate with learning graphic design principles. It shows that, in the context of *Posterlet*, the choice to seek more negative feedback and the choice to revise independently correlate with student learning from the game. This is a first-of-kind demonstration, which warrants future testing to determine if this relation holds for other contexts and tasks, as well as for the causal variables driving the correlations. The second section considers the effect of choosing negative feedback on time spent reading feedback. It demonstrates that the more negative feedback students choose for a given poster, the longer they dwell on the feedback. This may further help explain how negative feedback, at least when self-selected, contributes to greater learning. The third section evaluates whether the *Posterlet* results correlate with academic achievement more broadly. We could not collect *in situ* measures of feedback-relevant behaviours in school and correlate those measures with *Posterlet*. Nevertheless, the stable correlations between negative feedback in *Posterlet* and standardized achievement provide promissory evidence that seeking negative feedback in *Posterlet* may have a meaningful association with how well children are doing in school. Finally, the fourth section demonstrates that *Posterlet* can detect differences between the two school samples in their choice of negative feedback. While we do not know the causes of these differences, it does indicate that *Posterlet* may be deployable as a way to evaluate the effects of different courses of instruction (e.g., instruction that teaches SRL behaviours).

4.2.1 In-Game Choices Correlate with Learning Graphic Design Principles

Table 2 shows the zero-order correlations between frequency of negative feedback selection, revision, overall poster quality, and post-test score. The more negative feedback students chose, the better they performed on the post-test and on the overall poster quality. Given that the score for positive feedback is the inverse of negative feedback, this result indicates that seeking positive feedback is negatively correlated with learning to the same degree. It is important to note that these results do not mean that students who prefer positive feedback did not learn from that feedback. They just learned less than their counterparts who sought more negative feedback.

The correlation between revision and overall poster quality is to be expected, given that the overall poster quality score comprised revised (and improved) posters for those students who chose to revise. More importantly, the choice to revise positively correlated with post-test scores, which did not depend on the quality of the posters that students produced.

Table 2. Correlations between negative feedback, revision, and internal learning outcomes

Measures	Negative Feedback	Revision	Poster Quality	Post-test
		n=473	n=473	n=414
Negative Feedback	--	$.47^{***}$	$.28^{***}$	$.23^{***}$
Revision		--	$.34^{***}$	$.24^{***}$
Poster Quality			--	$.39^{***}$

Note: *** *p*<.001

Is student prior knowledge responsible for these effects? Perhaps students who knew more about designing posters sought more negative feedback, revised more, and performed better on the post-test. To investigate this possibility, we treated the students' first poster (prior to any feedback or revision) as a pre-test. When controlling for differences in pre-test scores, all the correlations remain significant, *p'*s<.01. The partial correlations, when removing the effect of pre-test, are as follows: post-test x negative feedback *r*=.20, post-test x revision *r*=.21, overall poster quality x negative feedback *r*=.22, overall poster quality x revision *r*=.36. The significant pattern of correlations held up for both schools.

Negative feedback and revision were highly correlated. To determine if negative feedback and revision were independent predictors of learning from the game, we entered them simultaneously in two multiple regressions, one using the post-test as the dependent measure and one using overall poster quality. For the post-test, both negative feedback [$t(411)=2.7$, $p=.006$, $\beta=.15$, B=.17, SE=.06] and revision [$t(411)=2.8$, $p=.005$, $\beta=.16$, B=.36, SE=.13] were significant predictors [$F(2,411)=16.11$, $p < .001$, R Square=.07, Adjusted R Square=.07]. Similarly, for the overall poster quality score, negative feedback [$t(470)=3.2$, $p=.002$, $\beta=.15$, B=.89, SE=.28] and revision [$t(470)=5.4$, $p<.001$, $\beta=.26$, B=3.23, SE=.59] were significant independent predictors [$F(2,470)=36.07$, $p < .001$, R Square=.13, Adjusted R Square=.13].

In sum, two measures of learning — the performance-based poster quality score and the declarative post-test — correlated independently with both the choice to seek more negative feedback and with the choice to revise. This pattern persisted when controlling for prior knowledge. Thus, within the context of the design-oriented *Posterlet*, seeking negative feedback had a positive association with learning, even though both positive and negative feedback were equally informative.

4.2.2 The Amount of Chosen Negative Feedback Positively Correlates with Dwell Time

The standard explanation for the learning benefits of negative feedback is that it provides information for improvement, whereas positive feedback reinforces existing behaviours, not new learning (Hattie & Timperley, 2007). However, a second possibility is that people may also pay more attention to negative

(2015). *Posterlet*: A Game-Based Assessment of Children's Choices to Seek Feedback and to Revise. *Journal of Learning Analytics*, *2*(1), 49–71.

feedback, which could further contribute to learning. For example, Hancock, Stock, and Kulhavy (1992) found that if students aged 15 to 18 were confident of a correct answer on SAT-like questions and they received negative feedback, they spent more time reading that feedback. We assume the students were preparing to take the SAT, so they had high motivation to learn. However, Chase et al. (2009) found that children tended to pay less attention to negative feedback they received in a knowledge construction task, such as building a concept map. *Choosing* to receive negative feedback may mitigate the tendency to under-process (unwanted) negative feedback. To test the relation of negative feedback and processing, we further examined the amount of time the students spent dwelling on the feedback. The expectation was that students would spend more time processing negative feedback than positive feedback.

As a first step, we correlated the feedback-screen dwell time for each poster with the number of negative feedback messages (0–3) the student chose for that poster. The correlations were as follows: $r_{\text{Poster 1}}$=–.02, p=.67; $r_{\text{Poster 2}}$=.20, p<.001; $r_{\text{Poster 3}}$=.21, p<.001. For posters 2 and 3, there was a positive association between the amount of chosen negative feedback and the dwell time, which we interpret as a closer read of the feedback. The lack of correlation for poster 1 may be attributed to the novelty of the environment, so that for poster 1 there was more exploration and more variability: M_{round1}=19.1 seconds, SD=17.2; M_{round2}=14.8 seconds, SD=8.7, and M_{round3}=12.9 seconds, SD=6.7.

The correlation between the dwell time and amount of negative feedback does not stem from differences in message length: there is no statistically significant difference in the length of the negative and positive feedback messages. However, a reasonable explanation is that students who choose more negative feedback were generally more diligent and they read their feedback more carefully and slowly. A different explanation is that the correlation reflects students selectively paying closer attention to their negative feedback than to their positive feedback. We looked at dwell times within subjects to test this hypothesis. We compared the dwell times for each student for rounds 2 and 3. The logic is that, if the correlation is specific to negative feedback, then a student who chose more negative feedback on round 2 than on round 3 should also exhibit longer dwell times for round 2 than for round 3. Similarly, if a student chose more negative feedback on round 3, then dwell times should be longer for round 3. This is exactly what we found. We constructed a variable that computed the amount of negative feedback chosen for round 2 minus the amount of negative feedback chosen for round 3. We also constructed a variable that computed the feedback dwell time for round 2 minus the feedback dwell time for round 3. The correlation of these two variables was significant; r=.14, p=.003. This indicates that individuals spent more time processing self-selected negative feedback compared to self-selected positive feedback.

The log data permit further refined analyses. For instance, one might examine whether students who received negative feedback about a specific principle (e.g., font too small to read) performed better on that specific principle when they made their next poster. A difficulty with this analysis is that students would often try out completely different designs on their next poster, so that they did not immediately employ what they had learned. More generally, we searched the log files for many of these types of micro-changes. We were unable to find patterns more telling than the coarse analyses above, which we applied *a priori*.

4.2.3 Seeking Negative Feedback in the Game Correlates with Academic Achievement

Are we only measuring behaviours and correlations that hold in the context of the game? Ideally, we could observe students choosing feedback across many topics and settings and determine whether their choices correlate with learning in those environments. However, prior to incurring the expense of broad generalization research, a prudent first step is to determine whether our assessment shows any relation to other markers of learning. We received standardized English Language Arts and Mathematics achievement tests for a subset of the students. We correlated learning choices collected from the game with these measures. Table 3 shows correlations by school. The association of negative feedback with achievement outcomes is significant, moderate, and similar across the four tests, whereas revision shows more modest and variable correlations. Notably, the correlations of negative feedback are greater here for the achievement outcomes than for the measures of graphic design principles taught in the game (Table 2). This is likely a function of our learning measures, which had not had the benefit of multiple revisions to improve their precision and reliability.

Table 3. Correlations between negative feedback, revision, and outside assessments

Measures	**New York City** (n=119)		**Chicago** (n=65)	
	Reading (NYSTP ELA)	Math (NYSTP Math)	Reading (ISAT)	Math (ISAT)
Negative Feedback	.33***	.39***	.41***	.33**
Revision	.08	.28**	.31*	.21

Note: *** *p*<.001, ** *p*<.01, * *p*<.05

The fact that the choice to seek negative feedback holds relatively high correlations across different populations of students and across different standardized achievement measures is a promising indicator that *Posterlet* measures something of importance. This finding is somewhat surprising because *Posterlet* presents a context of design, whereas the achievement tests emphasize demonstrably correct answers. The fact that negative feedback shows stable correlations with achievement measures, whereas revision does not, provides useful divergent validity. It implies that *Posterlet* is measuring two different things, one of which seems more relevant to academic achievement. One value of divergent validity is that it contributes to theory building. For example, if one believes that the correlation of negative feedback, revision, and learning are due to a latent variable, such as students' IQ or growth mindset (Dweck, 2006), then one needs to explain further why negative feedback has a consistent relation to multiple achievement measures, whereas revision does not.

4.2.4 Choices to Seek Negative Feedback Are Contingent on Experience

Assessments of human abilities always raise the question of whether performance differences result from a relatively immutable trait, such as IQ, or whether the abilities are mutable. In the case of *Posterlet*, the question is whether there is a reason to believe that a course of instruction could change

(2015). *Posterlet*: A Game-Based Assessment of Children's Choices to Seek Feedback and to Revise. *Journal of Learning Analytics*, *2*(1), 49–71.

student willingness to seek negative feedback. While we want to believe the answer is yes, it would be reassuring to find evidence to support the idea that feedback seeking is malleable, before engaging in costly instructional research. As a simple exploration, we tested whether the NYC and Chicago schools exhibited different patterns of choices. The NYC students chose negative feedback (M=3.96, SD=2.48) significantly more often than Chicago students (M=3.24, SD=2.14), $t(402)=-3.2$, $p<.01$, *df* adjusted for unequal variances. There were no appreciable differences in the rates of revision; $M_{NYC}=1.1$ (SD=1.14), $M_{Chicago}=1.2$ (SD=1.14), $t(442)=.68$, $p=.50$. Thus, students from the two schools showed strong differences in the willingness to seek negative feedback, but not in the willingness to revise. The schools further differed in post-test performance, with NYC outperforming Chicago on the two checklist questions; $t(391)=2.4$, $SE=.25$, $p=.02$. We do not know what caused these differences, which could range from differences between the states, the school climates, the curricular implementations, or even the home environments that led parents to send their children to a game-based school. In the meantime, the results suggest that the choices to seek negative and positive feedback are influenced by experience.

5 DISCUSSION, LIMITATIONS, AND CONCLUSIONS

To support research on self-regulated learning, we are developing relatively short, game-based assessments. It is not our initial goal to test broad SRL theories but rather to provide specificity and behavioural evidence to the theorists. Therefore, the assessments probe for specific behaviours in specific contexts, which may be important SRL behaviours. With *Posterlet*, we tested the frequency of students' choices of negative over positive feedback in the context of designing posters. We also tested whether they chose to revise their work given an opportunity.

An important element of our choice-based assessments is that they include new material that students can learn. This way, the assessments can test whether students have developed SRL behaviours that prepare them for future learning. To be clear, *Posterlet* does not teach students to choose negative feedback and revise. Instead, *Posterlet* measures whether students choose negative feedback and revise, and whether they learn graphic design principles as a consequence. This assessment design has the advantage that it both tests for SRL choices and it also determines whether those SRL choices do, in fact, correlate with learning.

The current work presents our initial empirical validation of *Posterlet*. We collected the log files from hundreds of students who completed *Posterlet* as part of a larger evaluation of two game-themed schools. We found that students who chose more negative feedback also learned more graphic design principles as measured by the quality of their posters and their performance on a post-test. Additionally, these correlations were largely unchanged when controlling for the students' incoming knowledge (as revealed by their first poster before feedback). Stated inversely, students who sought positive feedback learned less. We also found that the choice to revise correlated with improved learning according to the same outcome measures. The choices of negative feedback and revision were correlated, so it is difficult to know if the negative feedback drove greater revision. Nevertheless, negative feedback and revision were separable and independent contributors to student learning, so they should not be subsumed under the same SRL construct.

To our knowledge, this is a first-of-kind demonstration that choosing negative over positive feedback correlates with learning. Moreover, we found that students who chose negative feedback spent more time attending to that feedback. This was not due to a person-level variable, such as diligence or IQ, because the same child would spend more time reading feedback when choosing more negative feedback. This is an additional first-of-kind demonstration. Prior explanations of why negative feedback enhances learning have focused on the error-correction signal. Here, we also found that when people have the opportunity to choose the valence of their feedback, they spend more time processing negative feedback, which, by nearly any cognitive theory, should lead to greater learning (see also Long & Aleven, 2013).

Somewhat surprisingly, choosing negative feedback in the game also exhibited stable correlations with four separate measures of standardized achievement ($.32<r<.41$). The correlations are of moderate size, but the fact that there are any correlations is somewhat surprising, because *Posterlet* presents a context of design, whereas standardized achievement tests reflect long-term experiences in more academic contexts. It is a matter for future research to determine the bridge of transfer between behaviours in *Posterlet* and academic outcomes. One can imagine several possibilities for how the choice to avoid negative feedback would play out in children's academic achievement. For instance, children may disregard negative feedback to avoid ego threat (Chase et al., 2009) and children may not ask other people to look at their work for fear of negative feedback. Moreover, there may be a broader disposition of which the choice to seek negative feedback is simply one manifestation. There are a variety of possible candidates that range from self-efficacy to growth mindset. However, to apply these theories, it will be important to account for why the choice to revise did not exhibit equally stable correlations across the standardized achievement tests ($.08<r<.31$). Growth mindset, for instance, should predict both the motivation to gain information for improvement (negative feedback) and the motivation to take action (revise) based on that information. However, only the choice to seek negative feedback exhibited a consistent relation with academic achievement.

The current results reflect initial work and, as such, the findings must be scoped within the typical limitations of early correlational research that has not had a chance to test for replicability, generality, and causality. First, we were not present during the administration of the assessment. The proctors, whom we never met, were blind to our hypotheses and had not seen *Posterlet* previously. Still, it is possible that some unknown events drove the results.

Second, our main findings are that the choice of more negative feedback is associated with greater learning and longer feedback dwell times. It is important to recognize this is a single demonstration with a unique assessment using a single age group. Whether this finding holds in more everyday settings and for other learner profiles has not been demonstrated, and therefore, cannot be claimed (but see Cutumisu & Schwartz, 2014; Cutumisu, Chin, & Schwartz, 2014). The correlations with academic achievement measures suggest there is some relation to other contexts of learning, but the nature of that relation is unknown. This is typical of behavioural research, which depends on replication and testing under a variety of circumstances before claiming generalizability. In the meantime, the finding may serve as a useful data point for SRL theories.

Third, our research was primarily correlational and designed to evaluate our assessment. We cannot make causal claims that choosing more negative feedback is the cause of better learning. As always, there may be latent variables (e.g., persistence) that drove the correlations of negative feedback selection, revision, and learning. Similarly, we cannot make causal claims about how the choice to seek negative feedback in *Posterlet* is related to performance on the standardized achievement measures. However, given the initial demonstration, it is now possible to consider research designs that could pursue questions of causality.

Fourth, we have no way to determine the sources of individual differences in the selection of negative feedback, which could be a function of parental income, school curricula, weather, and many other factors (Aikens & Barbarin, 2008). For example, we found that the NYC students chose more negative feedback than the Chicago students. For all we know, this may be a result of being raised in NYC and may have nothing to do with their respective school experiences. From our perspective, the source of individual differences is of less interest than the question of whether we can instruct students to seek negative feedback, and whether, in turn, this will improve their learning. An experiment on this question is currently underway and we are using *Posterlet* as a post-test to determine whether students transfer their SRL lessons from a classroom setting to the fanciful design world of *Posterlet*.

Thus far, the work on *Posterlet* has been theoretically agnostic. *Posterlet* was not designed to test any broad SRL theory or theory about types of learners (e.g., fixed versus growth mindset). The goal was to create a theory-neutral, process-based behavioural assessment of a potential SRL behaviour and, simultaneously, to determine if this SRL behaviour is useful for learning. Now that this initial work is done, it is possible to begin theoretical work.

One important place to focus that theoretical work is the difference between spontaneously choosing to engage in an SRL behaviour versus being directed to engage in that behaviour. For example, letting patients choose their level of pain medication led to lower doses than when the doses were prescribed by medical staff (Haydon et al., 2011). Similarly, letting students choose negative feedback rather than assigning it to them may lead them to learn more from that feedback. Therefore, we are currently using *Posterlet* to investigate the effect of chosen versus assigned negative feedback. If choosing leads to better learning, it implies that our focus should not simply be on teaching an SRL strategy, but also on convincing students to choose that strategy, so they will reap the most benefit. While this point may seem obvious, one non-intuitive hypothesis is that the benefits of SRL will be more pronounced in transfer settings, where students choose to use those strategies, than in the original instructional context that required students to use those strategies.

In conclusion, students playing *Posterlet* exercise free choice in deciding to seek negative feedback or to revise, these choices being equally acceptable, rather than right or wrong. Assessing student choices provides a new approach for evaluating process skills that have been elusive to more traditional testing, but that are of great interest to many educators (Schwartz & Arena, 2013). Instead of considering choice as self-selection and source of motivation, we view choice as one of the most important outcomes of learning. By capturing students' choices, we approach our goal of measuring students' propensity for SRL. We aim to build a suite of choice-based assessments that can differentiate which educational experiences foster independent learners who can make good choices for learning in the future.

ACKNOWLEDGEMENTS

Funding from the Gordon and Betty Moore foundation and the NSF (Grant #1228831) supported this work. The results and interpretations do not represent those of the granting agencies. We thank Richard Arum for kindly including our assessment within his larger project, Neil Levine for his artwork and list of graphic design principles, as well as Jacob Haigh for programming *Posterlet* and Howard Palmer for his work on the Choicelet architecture. We also thank all the students from New York City and Chicago who participated in our study, as well as their teachers and proctors.

REFERENCES

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100*(2), 235–251.

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). Detecting student misuse of intelligent tutoring systems. *Proceedings of the 7th International Conference on Intelligent Tutoring Systems* (pp. 531–540). Berlin/Heidelberg: Springer.

Baker, R. S., Corbett, A. T., Roll, I., & Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction, 18*(3), 287–314.

Baker, R. S., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research*, *19*(2), 185–224.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education*, Vol. 24 (pp. 61–100). Washington, DC: American Educational Research Association (AERA).

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, *65*(3), 245–281.

Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*, *18*(4), 334–352.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology*, Vol. 5 (pp. 161–237). Mahwah, NJ: Erlbaum.

Chi, M., Schwartz, D. L., Chin, D. B., & Blair, K. P. (2014). Choice-based assessment: Can choices made in digital games predict 6th-grade students' math test scores? In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 36–43). Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/long%20papers/36_EDM-2014-Full.pdf

Chi, M., VanLehn, K., & Litman, D. (2010). Do micro-level tutorial decisions matter: Applying reinforcement learning to induce pedagogical tutorial tactics. *Proceedings of the 7th International Conference on Intelligent Tutoring Systems* (pp. 224–234). Berlin/Heidelberg: Springer.

Chin, D. B., Blair, K. P., & Schwartz, D. L. (2015). Got game? A choice-based learning assessment of data visualization skills. Paper presented at *American Educational Research Association Annual Conference*, Chicago, IL, April 16–20.

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 245–252). New York, NY: ACM Press.

Cutumisu, M., Blair, K. P., Schwartz, D. L., & Chin, D. B. (2015). A game-based assessment of children's choices to seek feedback and to revise. Paper presented at *American Educational Research Association Annual Conference*, Chicago, IL, April 16–20.

Cutumisu, M., Chin, D. B., & Schwartz, D. L. (2014). A game-based assessment of students' choices to seek feedback and to revise. *Proceedings of the 11th International Conference on Cognition and Exploratory Learning in Digital Age* (pp. 17–24), Best Paper, Porto, Portugal, October 25–27.

Cutumisu, M., & Schwartz, D. L. (2014). Choosing negative feedback improves learning for students of all ages: A game-based assessment of seeking negative feedback and revising. *Proceedings of the London International Conference in Education* (pp. 171–176), London, UK, November 10–12.

Dweck, C. (2006). *Mindset: The new psychology of success*. New York, NY: Random House LLC.

Fancsali, S. E. (2014). Causal discovery with models: Behavior, affect, and learning in cognitive tutor algebra. *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 28–35). Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/long%20papers/28_EDM-2014-Full.pdf

Finkelstein, S. R., & Fishbach, A. (2012). Tell me what I did wrong: Experts seek and respond to negative feedback. *Journal of Consumer Research*, *39*(1), 22–38.

Garrison, D. R. (1997). Self-directed learning: Toward a comprehensive model. *Adult Education Quarterly*, *48*(1), 18–33.

Hancock, T. E., Stock, W. A., & Kulhavy, R. W. (1992). Predicting feedback effects from response-certitude estimates. *Bulletin of the Psychonomic Society*, *30*(2), 173–176.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Haydon, M. L., Larson, D., Reed, E., Shrivastava, V. K., Preslicka, C. W., & Nageotte, M. P. (2011). Obstetric outcomes and maternal satisfaction in nulliparous women using patient-controlled epidural analgesia. *American Journal of Obstetrics and Gynecology*, *205*(3), 271, e1–6.

Howard-Rose, D., & Winne, P. H. (1993). Measuring component and sets of cognitive processes in self-regulated learning. *Journal of Educational Psychology*, *85*(4), 591–604.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology, 76*(3), 349–366.

Klehe, U. C., & Anderson, N. (2007). Working hard and working smart: Motivation and ability during typical and maximum performance. *Journal of Applied Psychology*, *92*(4), 978–992.

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science, 7*(3), 67–72.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*, 30–43.

Lee, S. J., Liu, Y. E., & Popović, Z. (2014). Learning individual behavior in an educational game: A data-driven approach. *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 114–121). Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/long%20papers/114_EDM-2014-Full.pdf

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 75–105). Hillsdale, NJ: Erlbaum.

Long, Y., & Aleven, V. (2013). Skill diaries: Improve student learning in an intelligent tutoring system with periodic self-assessment. *Artificial Intelligence in Education: Lecture Notes in Computer Science, 7926*, 249-258. Berlin/Heidelberg: Springer.

Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology*, 2 (pp. 745–783). New York, NY: Taylor & Francis.

Okita, S. Y., & Schwartz, D. L. (2013). Learning by teaching human pupils and teachable agents: The importance of recursive feedback. *Journal of the Learning Sciences*, *22*(3), 375–412.

Peacock, S. B., Smith, C. P., Martin, T., Aghababyan, A., Popović, Z., Andersen, E., & Liu, Y. E. (2013). Learning fractions through splitting in an online game. *Proceedings of the American Educational Research Association Annual Conference (AERA)*, San Francisco, CA, April 29. Retrieved from http://www.aera.net/Publications/OnlinePaperRepository/AERAOnlinePaperRepository/tabid/12720/Owner/604309/Default.aspx

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40.

Powers, W. T. (1978). Quantitative analysis of purposive systems: Some spadework at the foundations of scientific psychology. *Psychological Review, 85*, 417–435.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011a). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, *21*(2), 267–280.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011b). Metacognitive practice makes perfect: Improving students' self-assessment skills with an intelligent tutoring system. In *Artificial Intelligence in Education: Lecture Notes in Computer Science, 6738,* 288–295. Berlin/Heidelberg: Springer.

Roll, I., Baker, R. S., Aleven, V., & Koedinger, K. R. (2014). On the benefits of seeking (and avoiding) help in online problem-solving environments. *Journal of the Learning Scienc*es, 23, 537–560. doi: 10.1080/10508406.2014.883977

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, *73*(3), 482.

Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 352–359.

Schooler, L. J., & Anderson, J. R. (1990). The disruptive potential of immediate feedback. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 702–708), Cambridge, MA.

Schwartz, D. L., & Arena, D. (2013). *Measuring what matters most: Choice-based assessments for the digital age*. Cambridge, MA: MIT Press.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for learning: The hidden efficiency of original student production in statistics instruction. *Cognition and Instruction, 22*, 129–184.

Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Abingdon, UK: Routledge.

Snow, E. L., Allen, L. K., Russell, D. G., & McNamara, D. S. (2014). Who's in control? Categorizing nuanced patterns of behaviors within a game-based intelligent tutoring system. *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 185–191). Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/long%20papers/185_EDM-2014-Full.pdf

Stevens, R. H., & Thadani, V. (2007). Quantifying student's scientific problem solving efficiency and effectiveness. *Technology, Instruction, Cognition and Learning*, *5*(4), 325–337.

Thorndike, E. L. (1927). The law of effect. *American Journal of Psychology, 39*, 212–222.

Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior*, *29*(6), 2568–2572.

Zeidner, M., Boekaerts, M., & Pintrich, P. R. (2000). Self-regulation: Directions and challenges for future research. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 749–768). San Diego, CA: Academic Press.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational psychologist*, *25*(1), 3–17.

JOURNAL OF LEARNING ANALYTICS

(2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques

Christoph Sonnenberg and Maria Bannert
University of Wuerzburg, Instructional Media, Germany
christoph.sonnenberg@uni-wuerzburg.de

ABSTRACT: According to research examining self-regulated learning (SRL), we regard individual regulation as a specific sequence of regulatory activities. Ideally, students perform various learning activities, such as analyzing, monitoring, and evaluating cognitive and motivational aspects during learning. Metacognitive prompts can foster SRL by inducing regulatory activities, which, in turn, improve the learning outcome. However, the specific effects of metacognitive support on the dynamic characteristics of SRL are not understood. Therefore, the aim of our study was to analyze the effects of metacognitive prompts on learning processes and outcomes during a computer-based learning task. Participants of the experimental group (EG, n=35) were supported by metacognitive prompts, whereas participants of the control group (CG, n=35) received no support. Data regarding learning processes were obtained by concurrent think-aloud protocols. The EG exhibited significantly more metacognitive learning events than did the CG. Furthermore, these regulatory activities correspond positively with learning outcomes. Process mining techniques were used to analyze sequential patterns. Our findings indicate differences in the process models of the EG and CG and demonstrate the added value of taking the order of learning activities into account by discovering regulatory patterns.

KEYWORDS: self-regulated learning, metacognitive prompting, process analysis, process mining, think-aloud data, HeuristicsMiner algorithm

1 INTRODUCTION

Recent research in the field of self-regulated learning (SRL) has moved to a process-orientated or event-based view to investigate how learning processes unfold over time and how scaffolds influence the dynamic nature of regulatory activities. Two recent special issues indicate the importance of investigating sequential and temporal patterns in learning processes and present new methodological contributions for the analysis of time and order in learning activities (Martin & Sherin, 2013; Molenaar & Järvelä, 2014). Technical advances allow the recording of learning-related behaviour on a very detailed level and largely unobtrusively for learners (e.g., Azevedo et al., 2013; Winne & Nesbit, 2009). As such, researchers have focused more on behavioural process data and less on measures of aptitude (Azevedo, 2009; Bannert, 2009; Veenman, van Hout-Wolters, & Afflerbach, 2006). When focusing on process data, differences among learners are explained on the event level with respect to regularities and patterns (Winne & Perry, 2000), allowing researchers to gain new insights into the process of learning.

(2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

Process analysis methods beyond the variable-centred *coding and counting* approach (Kapur, 2011) can provide valuable information on the specific effects of scaffolds (e.g., metacognitive prompts) and are able to inform researchers about how to optimize an applied supporting strategy further (e.g., Jeong et al., 2008; Johnson, Azevedo, & D'Mello, 2011; Molenaar & Chiu, 2014). Moreover, findings on the sequential and temporal structure of SRL processes can provide knowledge for the development of SRL theories on the micro-level (Molenaar & Järvelä, 2014).

Our approach applies the techniques of process mining (Trčka, Pechenizkiy, & van der Aalst, 2010) on process data obtained by concurrent think-aloud protocols (Ericsson & Simon, 1993). For example, we have compared process patterns of students with high versus low learning performance in a recent study (Bannert, Reimann, & Sonnenberg, 2014) and demonstrated that process mining techniques can reveal differences in the sequential patterns of regulatory processes. Now, we are investigating the effects of *metacognitive prompts* (Bannert, 2009) by means of an in-depth analysis using process mining techniques. An analysis of differences in the process models between students supported by metacognitive prompts and students without prompts can provide information on how to promote beneficial regulatory patterns and thereby improve learning.

The paper is structured as follows: First, we introduce research focusing on the support of SRL through metacognitive prompts. Second, we describe SRL models that emphasize the importance of different learning events and event patterns. Third, some of the foundations of analyzing learning processes with process mining are introduced. Fourth, we analyze process data from coded think-aloud protocols of an experimental study. In addition to the traditional frequency-based approach, the relative arrangement of learning activities is taken into account using process mining techniques. Finally, the results of these analyses are compared, and the effects of metacognitive support on the sequential structure of SRL processes are discussed.

2 THEORETICAL BACKGROUND

2.1 Metacognitive Support through Prompts

Current research in metacognition and SRL shows that learners often do not spontaneously use metacognitive skills during learning, which in turn leads to poorer learning outcomes (e.g., Azevedo, 2009; Bannert & Mengelkamp, 2013; Greene, Dellinger, Tüysüzoglu, & Costa, 2013; Winne & Hadwin, 2008; Zimmerman, 2008). The students' awareness and control of their own manner of learning is important, especially in technology-enhanced and open-ended learning settings (Azevedo, 2005; Lin, 2001; Lin, Hmelo, Kinzer, & Secules, 1999). In most open-ended learning environments, it is constantly necessary to make decisions on what to do and where to go next and to evaluate the retrieved information with respect to current learning goals (Schnotz, 1998). Therefore, the general purpose of our research is to provide metacognitive support for hypermedia learning through metacognitive prompts.

Instructional prompts are scaffolds that induce and stimulate students' cognitive, metacognitive, and motivational activities during learning (Bannert, 2009). The underlying assumption is that students have already acquired these processes, but they do not recall or execute them spontaneously in a specific learning situation (*production deficit*; Veenman et al., 2006; Veenman, 2007). Metacognitive prompts aim at inducing regulatory activities such as orientation, goal specification, planning, monitoring and control, and evaluation strategies (Bannert, 2007; Veenman, 1993) by asking students to reflect upon, monitor, and control their own learning process.

Previous research has demonstrated beneficial effects from metacognitive prompting (e.g., Azevedo, Cromley, Moos, Greene, & Winters, 2011; Ge, 2013; Johnson et al., 2011; Lin & Lehman, 1999; Veenman, 1993; Winne & Hadwin, 2013). For example, Lin and Lehman (1999) prompted students to give reasons for their actions to increase the awareness of their own strategies by utilizing a pop-up window at certain times in a computer-based simulation environment (e.g., "What is your plan for solving the problem?"). Their findings showed significantly higher performance on contextually dissimilar problems (i.e., far transfer performance) for the students supported by prompts. Based on an analysis of think-aloud data, Johnson et al. (2011) showed that prompts given by a human tutor during learning in a hypermedia learning environment influenced the deployment of regulatory processes and temporal dependencies. Compared to a control group, the externally assisted condition also achieved a better learning outcome.

In previous experiments, we investigated the effects of different types of metacognitive prompts during hypermedia learning (Bannert & Mengelkamp, 2013; Bannert & Reimann, 2012). The prompts stimulated or even suggested appropriate metacognitive learning activities for university students during a hypermedia learning session lasting approximately 40 minutes. For example, in one of our experiments, students were prompted after each navigational step in a learning environment to verbalize the reasons why they had chosen the next step (so-called *reflection prompts*; Bannert, 2006). Overall, the findings confirm the positive effects of all investigated types of metacognitive prompts on transfer performance and the use of learning strategies during learning.

Our most recent work (Bannert, Sonnenberg, Mengelkamp, & Pieger, 2015) investigates the effects of a new type of metacognitive prompt (so-called *self-directed metacognitive prompts)* on navigation behaviour and learning outcomes. In summary, the findings show that such prompts enhance strategic navigation behaviour (i.e., students visited relevant webpages significantly more often and spent more time on them) and transfer performance (i.e., students performed better at applying knowledge of basic concepts to solve prototypical problems compared with a control group). In addition, learner characteristics (e.g., prior domain knowledge or verbal abilities) were obtained by questionnaires, but they had no effects as covariates in our analyses. The present study extends this contribution by focusing on the sequential analysis of coded think-aloud data obtained during learning.
Despite the findings about the general effectiveness of metacognitive prompts, the specific effects of prompts on learning processes remain unexplained. More precisely, a closer look at the effects of

prompts on the sequential and temporal structure of SRL processes is necessary (e.g., Jeong et al., 2008; Johnson et al., 2011). Understanding this process at the micro-level would allow researchers to better design metacognitive support. For example, regulatory patterns associated with successful learning but that could not be fostered by metacognitive prompts could be identified. Subsequently, the metacognitive support could be adapted by taking information about these patterns into consideration. Therefore, we focus on analyzing the sequential order of learning activities obtained by concurrent think-aloud protocols during learning.

2.2 Regulatory Patterns in SRL

Boekaerts (1997) describes SRL as a complex interaction of cognitive, metacognitive, and motivational regulatory components. With respect to assumptions in SRL models (e.g., Winne & Hadwin, 2008; Zimmerman, 2008), successful studying corresponds with an active performance of different regulatory activities during learning. These regulatory activities include employing orientation to obtain an overview of the learning task and resources, planning the course of learning, monitoring and controlling all learning steps, and evaluating the learning product. Research in SRL has confirmed that successful learning is associated with the active deployment of these regulatory activities (e.g., Azevedo, Guthrie, & Seibert, 2004; Bannert, 2009; Johnson et al., 2011; Moos & Azevedo, 2009).

Most SRL models share the common assumption of a time-ordered sequence of regulatory activities, although they do not imply a strict order (Azevedo, 2009). Usually, three cyclic phases of forethought, performance, and reflection (Zimmerman, 2000) are distinguished. The forethought phase comprises task analysis, goal setting, and strategic planning. During the performance phase, self-observations for adaptations (monitoring) and control strategies (self-instruction or time management) are deployed. Finally, the reflection phase includes self-judgments and self-reactions, which, in turn, can inform the next forethought phase. The COPES model (Winne & Hadwin, 2008) represents a more elaborate description of regulatory processes in terms of an information-processing model. Here, learning occurs in three phases, namely, task definition, goal setting and planning, and studying tactics, and a fourth optional phase, adaptations to metacognition. In addition, monitoring and control are crucial elements in the COPES model. Monitoring is used to detect differences between current conditions (e.g., learning progress) and standards (e.g., predefined learning goals), which, in turn, activates control processes to reduce discrepancies (e.g., engaging more intensively in a certain topic).

2.3 Microanalysis Using Process Mining Techniques

In a recent study (Bannert et al., 2014), we suggested process mining (PM) as a promising method in SRL research. PM allows researchers to describe and test process models of learning that incorporate an event-based view and that are at the high end of process granularity. These process models are able to represent the workflow of activities (van der Aalst, Weijters, & Maruster, 2004). Therefore, we argue that PM is adequate for investigating regulatory patterns based on process assumptions conceptualized in SRL research, as described in the previous section. For example, PM or data-mining techniques can extract patterns by analyzing process data (e.g., think-aloud protocols or log files), and the resulting patterns can be compared to assumptions of SRL models (e.g., the assumption of a time-ordered sequence of regulatory activities in successful learning or the concept of dynamic and cyclic patterns). Therefore, the observed behaviour in process data could be aligned with SRL models.

PM is an approach that can be used in the context of *Educational Data Mining* (Romero, Ventura, Pechenizkiy, & Baker, 2010). In this context, PM represents student activities as a process model derived from their log traces while using a computer-based learning environment. In general, PM methods allow researchers to discover process models inductively from activity sequences stored in an event log, test process models through conformance checking with additional event data, or the extension of existing models (Trčka et al., 2010). Especially in the context of computer-supported learning research, PM techniques are increasingly used to study learning from a process-oriented perspective (Reimann & Yacef, 2013; Schoor & Bannert, 2012). For example, PM techniques can be applied to modelling sequences of learning activities that have been recorded in log files or coded think-aloud data.

By using PM techniques to discover process patterns in SRL activities, we assume that the present process data — comprising temporally ordered event sequences — is directed by one or more mental processes, with each set of processes corresponding to a process model. Hence, a process model represents a system of states and transitions that produced the sequence of learning events. Usually, the performance of this system is driven by a plan for action. In the context of SRL, this plan can be a learning strategy or an external resource provided to the learner (e.g., prompts). A process model is able to express a holistic view of a process by modelling a system comprising states and transitions rather than a process-as-sequence perspective (Reimann, 2009).

With respect to related approaches, hidden Markov models (e.g., Jeong et al., 2008) also allow for expressing the holistic nature of a process by taking into account the entire sample of behaviour. However, this approach uses time-consuming iterative procedures; generally, the researcher has to pre-define the appropriate number of states, and the interpretation of the output model is often difficult (van der Aalst, 2011). There are, however, approaches for automatically selecting the appropriate number of states using the Bayesian Information Criterion (e.g., Li & Biswas, 2002). Additionally, hidden Markov models, as well as simple transition graphs and other low-level models, represent a lower abstraction level than the PM notation language (e.g., inability to represent concurrency, which typically results in more complex models). Finally, PM techniques have the advantage of explicitly dealing with noise (i.e., exceptional or infrequent behaviour), which is necessary when analyzing real-life event traces. For these reasons, we recommend PM techniques for analyzing sequences of learning activities (see Bannert et al., 2014 for more information regarding the comparison to other process analysis methods).

2.4 Research Questions and Hypotheses

Metacognitive prompts ask students explicitly to reflect, monitor, and control their own learning process. They focus students' attention on their own thoughts and on understanding the activities in which they are engaged during learning (e.g., Bannert, 2006; Hoffman & Spatariu, 2011). Hence, it is assumed that prompting students to monitor and evaluate their own manner of learning will allow them to activate their repertoire of metacognitive knowledge and learning strategies, which will consequently enhance their learning process and learning outcome. However, according to previous work and research on metacognitive prompting, the use of metacognitive prompts has to be explained and practiced in advance to guarantee an adequate application during learning (e.g., Bannert, 2007; Veenman, 2007). Based on the findings of studies investigating the effects of metacognitive prompts (e.g., Azevedo et al., 2004; Bannert, 2009), we expect that students supported by metacognitive prompts will engage in more regulatory activities, as obtained by coded think-aloud protocols. Moreover, scaffolded SRL processes should result in better learning performance; that is, a positive effect on learning outcomes mediated by improved regulatory behaviour. Whereas these two hypotheses are based on a variable-centred view of learning processes, we assume that an event-centred analysis that takes into account the relative arrangements of multiple learning activities can provide additional information about the sequential structure of the regulatory behaviour induced by the prompts (e.g., a sequence of orientation activities, searching for relevant information, cognitive processing, and evaluation of progress are typically executed). Therefore, the effectiveness of metacognitive prompts can be analyzed on a micro-level, and the results can be used to derive implications for the improvement of metacognitive support. In detail, the following research questions are addressed in the present study:

1. Does metacognitive prompting during learning influence SRL processes by engaging students in more metacognitive learning events?
2. Does the number of metacognitive learning events mediate the effect of metacognitive prompting on learning outcomes?
3. Which sequential patterns of SRL activities are induced by metacognitive prompting compared to a control group without support?

2.5 Process Mining Using the HeuristicsMiner Algorithm

To analyze the relative arrangement of learning activities, we employed the PM approach (Trčka et al., 2010). The basic idea of PM is to use an event log to generate a process model describing this log inductively (process discovery). Furthermore, theoretical models or empirically mined models can be compared to event logs (conformance checking), and existing models can be extended (model extension). Fluxicon Disco Version 1.7.2 (2014) software was used for data preparation. Next, the event log was imported into the ProM framework Version 5.2 (2008), and PM was conducted. The ProM framework comprises a variety of PM algorithms that can be assigned to the functions of discovery, model checking, or model extension. For our analysis, we used the HeuristicsMiner algorithm (Weijters, van der Aalst, & de Medeiros, 2006) for process discovery.

(2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

We selected the HeuristicsMiner algorithm based on a comparison of seven state-of-the-art process discovery algorithms on the dimensions of accuracy and comprehensibility, provided by de Weerdt, de Backer, Vanthienen, and Baesens (2012). Accuracy is defined as the capability of a sound capturing of behaviour in an event log, omitting over- and underfitting (i.e., a process model should balance between generality and precision). Comprehensibility comprises simplicity and structuredness of the resulting process models, and thereby determines the complexity and ease of interpretation of the output. For the first time, real-life event logs containing log data from different information systems were used for benchmarking PM algorithms. Among the seven algorithms, the HeuristicsMiner was the best technique for the real-life logs used and the authors conclude that "HeuristicsMiner seems the most appropriate and robust technique in a real-life context in terms of accuracy, comprehensibility, and scalability" (De Weerdt et al., 2012, p. 671). In the following, we explain the general principle and functionality of this algorithm in more detail.

2.6 General Principle of the HeuristicsMiner

The general principle of the HeuristicsMiner algorithm is to take into account the sequential order of events for mining a process model that represents the control flow of an event log (Weijters et al., 2006). The event log containing case IDs, time stamps, and activities represents the data input. Based on this input, the algorithm searches for causal dependencies between activities by computing a dependency graph that indicates the certainty of a relation between two activities (e.g., event *a* is followed by event *b* with a certainty of 0.90). Finally, a so-called *heuristic net* is generated as an output model that constitutes a visual representation of the dependencies among all activity classes in the event log. The resulting process model can be adjusted by setting thresholds for the inclusion of relations in the heuristic net (for more details on parameter settings, see below).

In addition, the HeuristicsMiner is based on two main assumptions. First, each non-initial activity has at least one other activity that triggers its performance, and each non-final activity is followed by at least one dependent activity. This assumption is used in the so-called *all activities connected heuristic* (Weijters et al., 2006). Second, the event log contains a representative sample of the observed behaviour, which usually contains a certain amount of noise, especially if traces of human behaviour are stored in the event log. For example, in our study, a perfect trace of verbal utterances for all performed learning steps is unlikely. Therefore, the event log contains noise caused, for example, by a missing learning step that was not uttered or by disagreement among the raters during the coding procedure. It must be noted that there is also noise in other types of data (e.g., log file data). Consequently, an analysis method is needed that can abstract from noise and that can concentrate on the main relations among learning activities. It is a specific feature of the HeuristicsMiner to be robust to noise in the data. This is the main reason for the appropriateness of applying this PM algorithm to our event log.

An additional advantage of the HeuristicsMiner algorithm is that the mined model (heuristic net) can be converted into a formal *petri net*. A petri net can be described as a bipartite directed graph with a finite set of places, a finite set of transitions, and two sets of directed arcs, from places to transitions and from transitions to places (Reisig, 1985). Thus, the resulting process model can be used as input for other PM algorithms, and it can be utilized in subsequent analyses (e.g., conformance checking between the model and a new event log). In contrast, the output model of another promising process discovery algorithm within the ProM framework that we used in previous process analyses (Bannert et al., 2014; Schoor & Bannert, 2012), called the *Fuzzy Miner* (Günther & van der Aalst, 2007), cannot be converted into a petri net (De Weerdt et al., 2012). Therefore, the HeuristicsMiner was the first choice for our present analysis.

2.7 Functionality and Application of the HeuristicsMiner

Considering its functionality, the HeuristicsMiner algorithm uses several parameters that guide the creation of a process model and that can be adjusted to set the level of abstraction from noise and low-frequency behaviour. First, a frequency-based metric is used to determine the degree of certainty of a relation between two events, *A* and *B*, based on an event log. The dependency values, ranging between –1 and 1, between all possible combinations of events are computed using the following formula (Weijters et al., 2006, p. 7):

$$A \Rightarrow_W B = \left(\frac{|a >_W b| - |b >_W a|}{|a >_W b| + |b >_W a| + 1}\right)$$

Based on an event log *W*, the certainty of a dependency relation between two events, $A \Rightarrow_W B$, is computed using the number of times event *a* is followed by event *b*, subtracted from the number of times event *b* is followed by event *a*, and divided by the number of occurrences of these two relations, plus 1. The number of correct (*a* follows *b*) and incorrect (*b* follows *a*) event sequences influences the dependency value by the +1 in the denominator. For example, an event log containing only correct sequences (*a* is always followed by *b*, but never vice versa), but with a low frequency of five observations, results in a certainty of 5/6 = 0.83, whereas in the case of a high frequency of 50 observations, the certainty of a dependency relation between *a* and *b* would be 50/51 = 0.98.

Moreover, the computed dependency values are used to construct a heuristic net (i.e., the output model). However, not all dependency relations are kept in the process model. Instead, the HeuristicsMiner algorithm concentrates on the main causal dependencies and abstracts from noise and low-frequency behaviour. At first, the all activities connected heuristic is applied. Therefore, only the best candidates (with the highest $A \Rightarrow_W B$ values) regarding the dependency values are kept in the output model. Second, three threshold parameters are used for the selection of further dependency relations. The *dependency threshold* determines the cut-off value for the inclusion of dependency relations in the output model. Furthermore, the *positive observation threshold* defines the minimum number of necessary observed sequences. Finally, the *relative to best threshold* determines that only additional dependency relations with a lower difference to the best candidate are included in the output model. We refer to Weijters et al. (2006) for more information about these threshold parameters.

In our analysis, the threshold parameters were kept at their default values of dependency threshold = 0.9, positive observation behaviour = 10, and relative-to-best-threshold = 0.05. As explained above, these threshold parameters can be used to adjust the level of abstraction of the output model. For example, reducing the cutoff-values would result in additional dependency relations in the model and thus increase the complexity. However, there were no reasons for changing the default-values in our case.

Furthermore, the HeuristicsMiner algorithm can also address *short loops* of lengths one (e.g., ACCB) and two (e.g., ACDCDB) as well as *long distance dependencies*; that is, a dependency based on choices made in other parts of the process model. Moreover, the algorithm considers *AND-relations* (two events are executed concurrently) and *OR-Relations* (e.g., either event *b* or event *c* can be executed after event *a*) to construct the heuristic net.

In general, searching for an optimal process model based on a present event log can be challenging, especially if there is a certain amount of noise and less-frequent behaviour in the data. Therefore, it is possible to compare the resulting process model with the event log using a *fitness* value (Rozinat & van der Aalst, 2008). The fitness indicates the gap between the observed behaviour, that is, the set of event sequences in the log, and the mined process model.

By applying the HeuristicsMiner algorithm to our event log, we assume that the present set of sequences of learning events is caused by one or multiple underlying processes. However, it might be possible that there is a high variety in SRL activities within the sample. In this case, using very robust algorithms such as the HeuristicsMiner can result in over-generalization (underfitting); that is, the mined model allows for much more behaviour than what is actually observed (De Medeiros et al., 2008). Therefore, the event log could be modelled more precisely by generating different process models for subsets of participants instead of a single model for all cases. This approach is called *trace clustering*, which can improve the discovery of process models (De Weerdt, vanden Broucke, Vanthienen, & Baesens, 2013; Greco, Guzzo, Pontieri, & Saccà, 2006). A plug-in has been implemented in the ProM framework that combines the HeuristicsMiner algorithm with a trace clustering procedure, namely, *DWS mining* (Disjunctive Workflow Schema; De Medeiros et al., 2008). The basic idea of DWS mining is to split the log into clusters iteratively until the mined process model for each cluster reaches high precision. A process model has a high precision if it only allows for behaviour that was observed in the event log. Consequently, a cluster is further partitioned if the mined model allows for more behaviour than is expressed by the cases within this cluster. For more information on the DWS mining plugin, refer to De Medeiros et al. (2008). In our analysis, we kept the default parameter settings for clustering the log traces.

3 METHOD

The present study extends a previous contribution (Bannert et al., 2015) that investigates the effects of metacognitive prompting on navigation behaviour and learning outcome referring to the same participants, but to different research questions and to mostly different data.

3.1 Sample and Research Design

A total of n=70 undergraduate students from a German university participated in the study (mean age = 20.07, *SD* = 1.88, 82.9% female). All participants were either majoring in media communications or in human–computer systems. Participants were recruited via an online recruitment system administered by our institute, and each student received 40 Euros (approximately $47 USD) for participating.

Altogether, the experimental study was based on a between-subject design and comprised two sessions. In the first session, learner characteristics were obtained as potential covariates (e.g., prior domain knowledge), especially in the case of an unbalanced distribution of characteristics among the groups by randomization (which is possible for the relatively small sample size). Approximately one week later, the participants were randomly assigned to either the experimental group (n=35) or the control group (n=35) and individually participated in hypermedia learning. The experimental group learned with metacognitive prompts, whereas the control group learned without prompts. Figure 1 presents an overview of the research design.

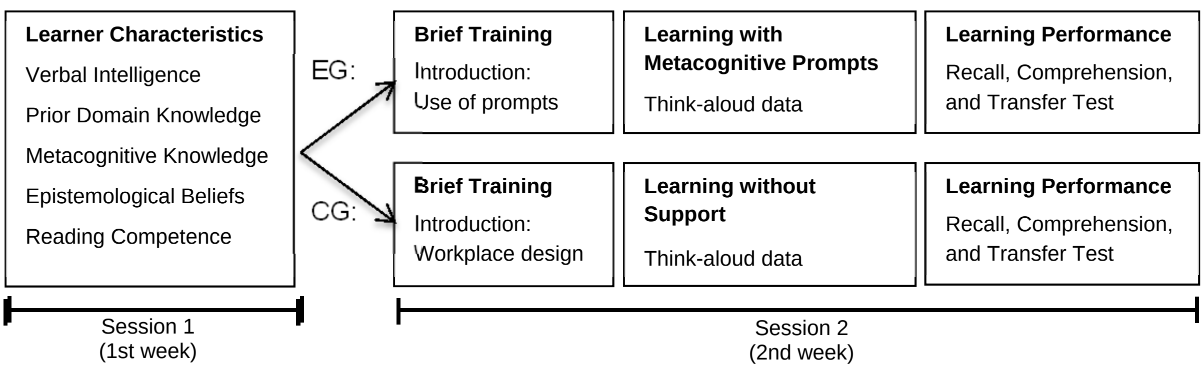


Figure 1. Research design

3.2 Learning Material and Performance Measurement

3.2.1 Learning Environment and Metacognitive Prompts

The learning material comprised a chapter on the topic of learning theories (classical conditioning, operant conditioning, and observational learning) presented in a hypermedia learning environment. For example, the content of one node included a description of the Skinner-box with reference to the concept of operant conditioning, and illustrated with a picture. In total, the material comprised 50 nodes with 13,000 words, 20 pictures and tables, and 300 hyperlinks. Within this chapter, the material relevant for the learning task comprised 10 nodes with 2,300 words, 5 pictures and tables, and 60 hyperlinks. The remaining pages were not relevant for the learning task. These pages included overviews, summaries, and pages with information on concepts not relevant for the learning goals. The Flesch-Kincaid grade-level score of the complete learning material was 19.01.

(2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

Navigation in the learning environment was possible in four different ways: 1) a hierarchical navigation menu, 2) a next-page and previous-page button on top of each page, 3) the backward- and forward-button of the browser, and 4) hyperlinks embedded in the text.

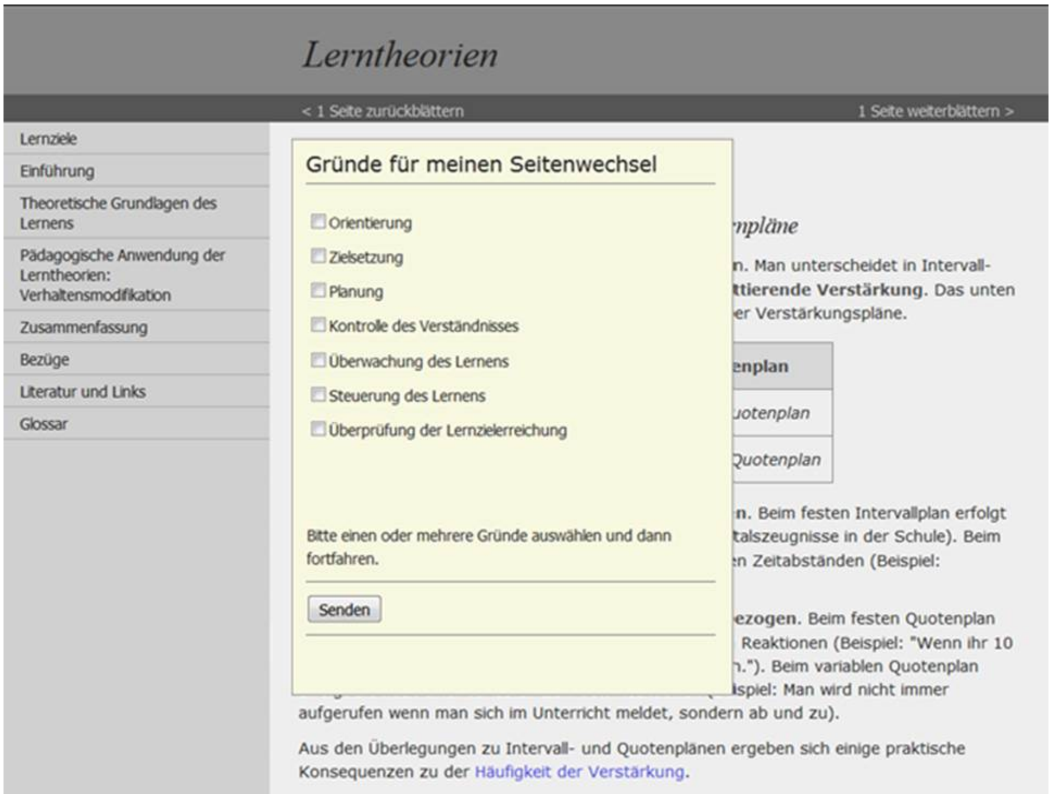


Figure 2: Learning environment with metacognitive prompt. Students are asked to select one or more reasons for node selection in a hypermedia learning environment by choosing among a list of strategic reasons (e.g., orientation, goal specification, planning) eight times during learning

Support via metacognitive prompts was implemented in the learning environment. A prompt appeared in the form of a pop-up window placed in the middle of the screen eight times during learning. Each prompt contained a list of strategic reasons for node selection. At least one reason had to be selected before continuing with learning. Figure 2 shows the hypermedia learning environment with a metacognitive prompt.

3.2.2 Knowledge Tests

Learning performance was measured with three knowledge tests on different levels based on Bloom's taxonomy of cognitive learning (Bloom, 1956). The measurement comprised a free recall test, a comprehension test, and a transfer test. In the free recall test, students were instructed to write down all basic concepts they could remember. The comprehension test, which assessed knowledge of facts, comprised 22 multiple-choice items, each with one correct and three incorrect answers. Transfer was measured by asking students to apply basic concepts and knowledge of facts to solve eight prototypical problems in educational settings that were not explicitly addressed in the learning material (maximum score: 40 points). For example, students were asked to explain how a teacher should behave in response to a described classroom discipline problem based on the principles of operant conditioning. The answers of the participants were rated on a researcher-developed rating scale by two research assistants (Cohen's Kappa = .84). In case of disagreement among the raters, one of the authors determined the final score. More information on the learning material and knowledge tests used in our prompting studies is provided by Bannert and Reimann (2012) and by Bannert and Mengelkamp (2013).

3.3 Procedure

Approximately one week before the learning session, the learner characteristics *verbal intelligence*, *prior domain knowledge*, *metacognitive strategy knowledge*, *epistemological beliefs*, and *reading competency* were measured by questionnaires. More information on the instruments is provided in Bannert et al. (2015).

The learning session started with an introduction phase. First, the navigation in the hypermedia learning environment was explained by the experimenter. Then, the participant was asked to practice all possible ways of navigating the learning environment by using a practice lesson. After that, a series of exercises had to be performed in the practice lesson using concurrent thinking aloud during the task. The experimenter provided feedback and, if necessary, additional exercises until the participant firmly mastered the think-aloud technique.

Subsequently, the students in the experimental group received an introduction (approximately 10 minutes) to the use of metacognitive prompts, which included a description of the importance of reflecting on one's own learning steps, an explanation of the reasons for strategic node selection listed in the prompts, and the correct use of the prompts. It is necessary to explain the use of metacognitive prompts to the students to guarantee adequate application during learning (e.g., Veenman, 2007). After that, they were instructed to configure the prompts by arranging the list with reasons for node selection and by defining eight time stamps when the prompts should be presented during learning. To keep the workload for both groups equivalent, participants in the control group received an introduction to workplace design, which is not relevant for the stimulation of metacognitive learning activities. Instead of prompt configuration, they were asked to arrange their workplace before learning. Both introductions were realized by the experimenter using a sheet of instructions and advice visible to the participant.

Following this, the learning phase started. All participants received a sheet with their learning task, which instructed them to learn the basic concepts of operant conditioning within 40 minutes. Moreover, they were provided with a list of seven example concepts that had to be learned (e.g., Skinner Box, Positive Reinforcement). Students in the experimental group received metacognitive support by prompts, whereas the control group learned without prompts. All participants were completely free to navigate in the learning environment and to use their learning strategies. During learning, notes could be taken on a blank sheet of paper (e.g., for summarizing or structuring information), but the participants were not allowed to use their notes to work on the knowledge tests. The participants were instructed to read and think aloud during the whole learning phase as practiced before, and these activities were videotaped. If a participant stopped thinking aloud for more than five seconds, the experimenter reminded her or him by saying, “Please think aloud.”

Directly after learning, the students worked on the three knowledge tests described above. Overall, the duration of the session was approximately two hours.

3.4 Coding Scheme

A coding scheme based on our theoretical framework of self-regulated hypermedia learning (Bannert, 2007) was used for segmenting and coding the students’ verbal protocols. Our theoretical framework characterizes hypermedia learning into the major categories *Metacognition*, *Cognition*, and *Motivation*. In addition, it distinguishes several sub-categories within the categories Metacognition and Cognition, as further described below.

Table 1 presents the coding categories and provides descriptions and examples. The coding scheme comprises the main categories Metacognition, Cognition, Motivation, and Other. Metacognition includes the sub-categories *Orientation*, *Goal specification*, *Planning*, *Searching for information, Judgment of its relevance*, *Evaluating goal attainment*, and finally *Monitoring and regulation*. Cognition contains *Reading*, *Repeating information*, and deeper processing, that is, *Elaboration* and *Organization of information*. The main category of Motivation includes all positive and negative utterances on the task, the situation, or oneself. Finally, all task-irrelevant utterances, non-classifiable utterances, and the handling of the prompts for the experimental group were assigned to the category Other.

The coding was conducted based on the procedure presented by Chi (1997). Segmentation of the verbal protocols was based on meaning. A segment was assigned for every definable learning activity. Multiple or nested codes were not allowed. Four trained research assistants coded the verbal protocols of all 70 participants. A random sample of three participants from each of the experimental group and the control group was selected to compute the interrater’s reliability. The reliability, based on 1,385 segments, showed substantial agreement: Cohen’s Kappa = .78, which is seen as sufficient for the following analysis.

3.5 Analysis

An example of the coded data used for the process analysis is presented in Table 2. The data comprise three types of information: 1) a Case ID that clearly distinguishes the participants, 2) a time stamp that indicates the beginning of an event, and 3) a learning activity — that is, the assigned category of the coding scheme (CODE). Using this information, it is possible to compute not only the frequency of events but also to determine the relative arrangement of multiple events. For example, in the short section of

(2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

Table 2, MONITOR is the most frequent activity (3 occurrences). Furthermore, MONITOR is directly followed by READ twice and directly followed by ORGANIZATION once.

Table 1. Coding scheme for analyzing students' learning activities

Code	Coding Category	Description and Examples
	Metacognition	
ORIENT	Orientation	Task clarification, overview of material. *I will sketch the menu first.*
SETGOAL	Goal specification	Goal setting and sub-goaling *I have to learn the basic concepts of operant conditioning.*
PLAN	Planning	Planning how to proceed *First I will decide in which sequence I have to learn and which pages to read.*
SEARCH	Search	Searching for information *Where is the page with the information about plans of reinforcement?*
EVALUATE	Judgment	Judgments about the relevance of information *Skinner's Vita is not relevant for my learning task.*
EVAL	Evaluation	Checking and evaluating *Did I process all the topics?*
MONITOR	Monitoring	Monitoring one's own learning *Ah, now I understand the principle.*
	Cognition	
READ	Reading	Reading out loud
REPEAT	Repeating	Repeating
ELABORATE	Elaboration	Deeper processing, *paraphrasing, connecting, inferring*
ORGANIZATION	Organization	Organization *drawing a map, writing down major concepts*
	Motivation	
MOT	Motivation	Positive, negative, neutral motivational utterances regarding a task, person, or situation *The task is very interesting.*
	Other	
REST	Other	Off-topic statements, comments on technique, not interpretable statements, pauses *May I make notes? The mouse doesn't work well.*

Table 2. Section of an event log

Case ID	*Timestamp*	*Learning Activity (CODE)*
		…
Case 1	04:14	ORIENT
Case 1	04:19	MONITOR
Case 1	04:24	READ
Case 1	04:37	REPEAT
Case 1	04:43	ORIENT
Case 1	05:01	MONITOR
Case 1	05:03	READ
Case 1	06:51	MONITOR
Case 1	06:56	ORGANIZATION
Case 1	07:30	REPEAT
		…

In the first step of our analysis, we took the frequencies of coded learning activities into account (frequency analysis) and used the frequency of metacognitive events to examine the expected mediation effect on transfer performance (mediation analysis). Subsequently, the sequential order of the coded learning activities is analyzed using a PM algorithm to discover differences in the process models of the experimental and control groups (process mining). Therefore, our view on temporality corresponds to the relative arrangement of multiple events.

4 RESULTS

A preliminary analysis showed that the randomized assignment of participants resulted in two subsamples with similar learner characteristics. With the exception of one subscale of reading competency — namely, *text comprehension* (measured by ELVES; Richter & van Holt, 2005) — no significant differences were found. In the case of the subscale text comprehension, students in the control group scored significantly better than those in the experimental group ($t(69) = 2.97$, $p = .004$, $d = 0.72$; two-tailed testing). In conclusion, this analysis indicates that the following results are not caused by unbalanced subsamples.

4.1 Frequency Analysis

Table 3 presents the descriptive and test statistics of all coded events for the experimental group and the control group. In addition to the minimum and maximum occurrence of each category, absolute frequencies, means, and standard deviations are listed. A total of 8,743 events were coded for the students in the experimental group, and a total of 8,087 events were coded for the students in the control group. For the experimental group, there were, on average, approximately 250 events coded in

40 minutes of learning time, with 116 metacognitive, 107 cognitive, 2 motivational, and 25 other utterances. Participants in the control group showed a mean of approximately 231 events, with 98 metacognitive, 106 cognitive, 3 motivational, and 24 other utterances.

A one-tailed t-test for independent samples showed that the experimental and control groups significantly differ in the number of metacognitive utterances ($t(69) = 1.80$, $p = .038$, $d = 0.44$). As expected, students in the experimental group who had been supported through metacognitive prompts showed a higher number of metacognitive learning activities ($M = 116.43$, $SD = 45.97$) than students in the control group without prompts ($M = 98.49$, $SD = 36.72$). Moreover, both groups showed a similar number of utterances in the remaining main categories Cognition ($M_{EG} = 107.43$, $SD_{EG} = 36.01$; $M_{CG} = 106.31$, $SD_{CG} = 44.65$), Motivation ($M_{EG} = 2.06$, $SD_{EG} = 4.14$; $M_{CG} = 2.54$, $SD_{CG} = 4.40$), and Other ($M_{EG} = 25.34$, $SD_{EG} = 15.08$; $M_{CG} = 23.71$, $SD_{CG} = 12.19$). For these three categories, the t-tests for independent samples were not significant.

Concerning the descriptive statistics of the subcategories of Metacognition, the experimental group showed more Monitoring ($M_{EG} = 71.17$, $SD_{EG} = 37.62$; $M_{CG} = 58.00$, $SD_{CG} = 27.46$), Orientation ($M_{EG} = 14.31$, $SD_{EG} = 7.23$; $M_{CG} = 11.14$, $SD_{CG} = 5.80$), Evaluation ($M_{EG} = 3.63$, $SD_{EG} = 3.08$; $M_{CG} = 2.49$, $SD_{CG} = 3.03$), and Planning ($M_{EG} = 1.74$, $SD_{EG} = 1.65$; $M_{CG} = 0.74$, $SD_{CG} = 1.09$) compared to the control group. In both groups, the highest frequency occurred for Monitoring, followed by Orientation, Searching, and Judgment, whereas Planning and Goal specification were rarely executed by students. As reported in Table 3, on the right side, differences between the experimental and control groups are significant for Orientation, Planning, and Monitoring.

Within the main category Cognition, participants of the control group showed more reading activities ($M_{EG} = 40.66$, $SD_{EG} = 15.75$; $M_{CG} = 44.49$, $SD_{CG} = 18.19$) but less Elaboration ($M_{EG} = 21.91$, $SD_{EG} = 12.92$; $M_{CG} = 18.40$, $SD_{CG} = 17.54$) and less Organization ($M_{EG} = 26.37$, $SD_{EG} = 13.87$; $M_{CG} = 24.60$, $SD_{CG} = 12.51$) than participants of the experimental group. However, all differences regarding these categories are non-significant. Finally, motivational events seldom occurred in both groups and with non-significant differences.

4.2 Mediation Analysis

A mediation analysis was conducted to investigate whether the observed relationship between the treatment group and learning performance (outcome variable) is mediated by the number of metacognitive events during learning. Regarding measurements of learning outcome, only transfer performance (i.e., a post-test score) differed significantly between the experimental and control groups ($M_{EG} = 20.61$, $SD_{EG} = 3.97$; $M_{KG} = 18.79$, $SD_{KG} = 4.30$; $t(69) = 1.85$, $p = .035$, $d = 0.45$; for more details on learning outcomes, see Bannert et al., 2015). Furthermore, both the number of metacognitive events and its sub-category Monitoring significantly correlate with transfer performance (Metacognitive events: $r = .22$, $p = .033$; Monitoring: $r = .32$, $p = .003$). Therefore, these two variables are regarded as possible mediators, and only transfer performance is included as an outcome variable.

(2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

Table 3: Absolute frequencies, means, and test statistics of all coded learning events for the experimental group and the control group

	Experimental Group (n=35)					Control Group (n=35)							
	Min	Max	Absolute Frequency	*M*	*SD*	Min	Max	Absolute Frequency	*M*	*SD*	*t*	*p*	*d*
Metacognition	34	242	4075	116.43	45.97	34	173	3447	98.49	36.72	1.804	.038	0.44
Orientation	5	30	501	14.31	7.23	2	28	390	11.14	5.80	2.025	.024	0.49
Planning	0	5	61	1.74	1.65	0	4	26	0.74	1.09	2.987	.002	0.73
Goal specification	0	10	72	2.06	2.36	0	8	67	1.91	1.84	0.282	.309	0.07
Search	1	32	414	11.83	7.54	1	57	453	12.94	10.64	-0.506	.308	-0.12
Judgment	2	23	409	11.69	5.70	0	33	394	11.26	7.35	0.273	.393	0.07
Evaluation	0	15	127	3.63	3.08	0	15	87	2.49	3.03	1.565	.061	0.38
Monitoring	11	203	2491	71.17	37.62	9	124	2030	58.00	27.46	1.673	.049	0.41
Cognition	47	201	3760	107.43	36.01	30	193	3721	106.31	44.65	0.115	.909	0.03
Reading	20	84	1423	40.66	15.75	19	89	1557	44.49	18.19	-0.941	.350	-0.03
Repeating	2	45	647	18.49	11.04	2	59	659	18.83	12.25	-0.123	.902	-0.03
Elaboration	3	55	767	21.91	12.92	0	56	644	18.40	17.54	0.955	.343	0.23
Organization	3	61	923	26.37	13.87	0	58	861	24.60	12.51	0.561	.576	0.14
Motivation	0	18	72	2.06	4.14	0	22	89	2.54	4.40	-0.476	.636	0.11
Other	8	69	887	25.34	15.08	6	58	830	23.71	12.19	0.497	.621	0.12
Sum of all coded events	126	473	8743	249.80	76.64	112	390	8087	231.06	76.27	1.023	.310	0.25

Note: Since we expected metacognitive prompting to increase the number of metacognitive utterances, we conducted one-tailed testing for metacognitive categories; elsewhere we conducted two-tailed testing; p < .05.

(2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

We used the PROCESS custom dialog box for SPSS based on the regression-based approach of Hayes (2013) to run the mediation analysis, which calculates bootstrapped confidence intervals (BCa CI) for the indirect effect and the measurement of effect size. There was a significant indirect effect of the treatment on transfer performance through the number of metacognitive events, b = 0.33, BCa CI [–0.01, 1.09]. Kappa-squared (Preacher & Kelley, 2011) was used to measure the effect size. The detected effect is relatively small, K^2 = .039, 95% BCa CI [.004, .127]. Furthermore, there was a significant indirect effect through the number of Monitoring events, b = 0.48, BCa CI [0.01, 1.24]. Again, this represents a small effect, K^2 = .058, 95% BCa CI [.006, .142].

In summary, the number of metacognitive events and of its sub-category Monitoring could be identified as mediator variables. Metacognitive prompting increased the occurrence of metacognitive events, especially of Monitoring, which in turn enhanced the transfer performance. Due to the mediation effect of the sub-category Monitoring being even slightly larger than the effect of all metacognitive events, we conclude that the mediation is mainly driven by Monitoring. Figure 3 presents the mediation model, including Monitoring as mediator variable.

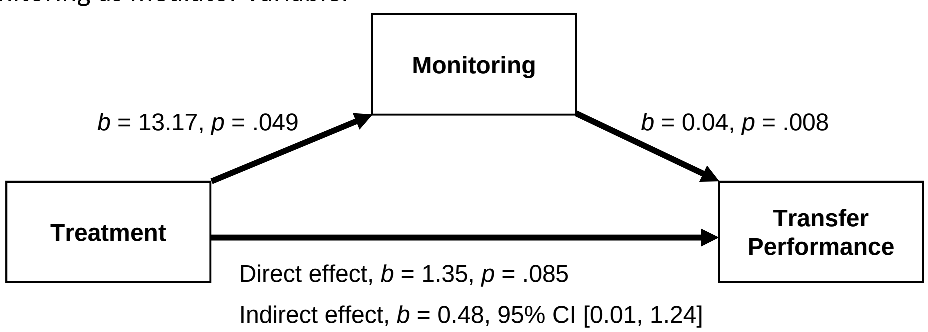


Figure 3: Mediation through the number of monitoring events

4.3 Process Analysis Using the HeuristicsMiner Algorithm

To apply the HeuristicsMiner, we had to simplify the categories of the coding scheme (see Table 1) for three reasons. First, the event classes Planning, Goal specification, and Motivation showed a very low frequency in our event log. Second, the HeuristicsMiner algorithm should preferably be used on data without too many different event classes (Rozinat, 2010). Finally, theoretical SRL models describe regulation processes mainly with the three phases of forethought, performance, and reflection (e.g., Zimmerman, 2000). The simplification was conducted as follows: We aggregated the metacognitive events Orientation, Planning, and Goal specification into a new event class called *Analyze*. Furthermore, the event class Judgment was added to Monitoring. In addition, the cognitive events Elaboration and Organization were combined to form a new event class called *Process*, that is, deeper processing. Finally, the event classes Motivation and Other were excluded from the process analysis. Altogether, seven event classes, listed in Table 4, were used for the analysis with PM techniques. With respect to the mean number of events, there was only a significant difference between both groups for the category Analyze (t(69) = 2.36, p = .011, d = 0.57).

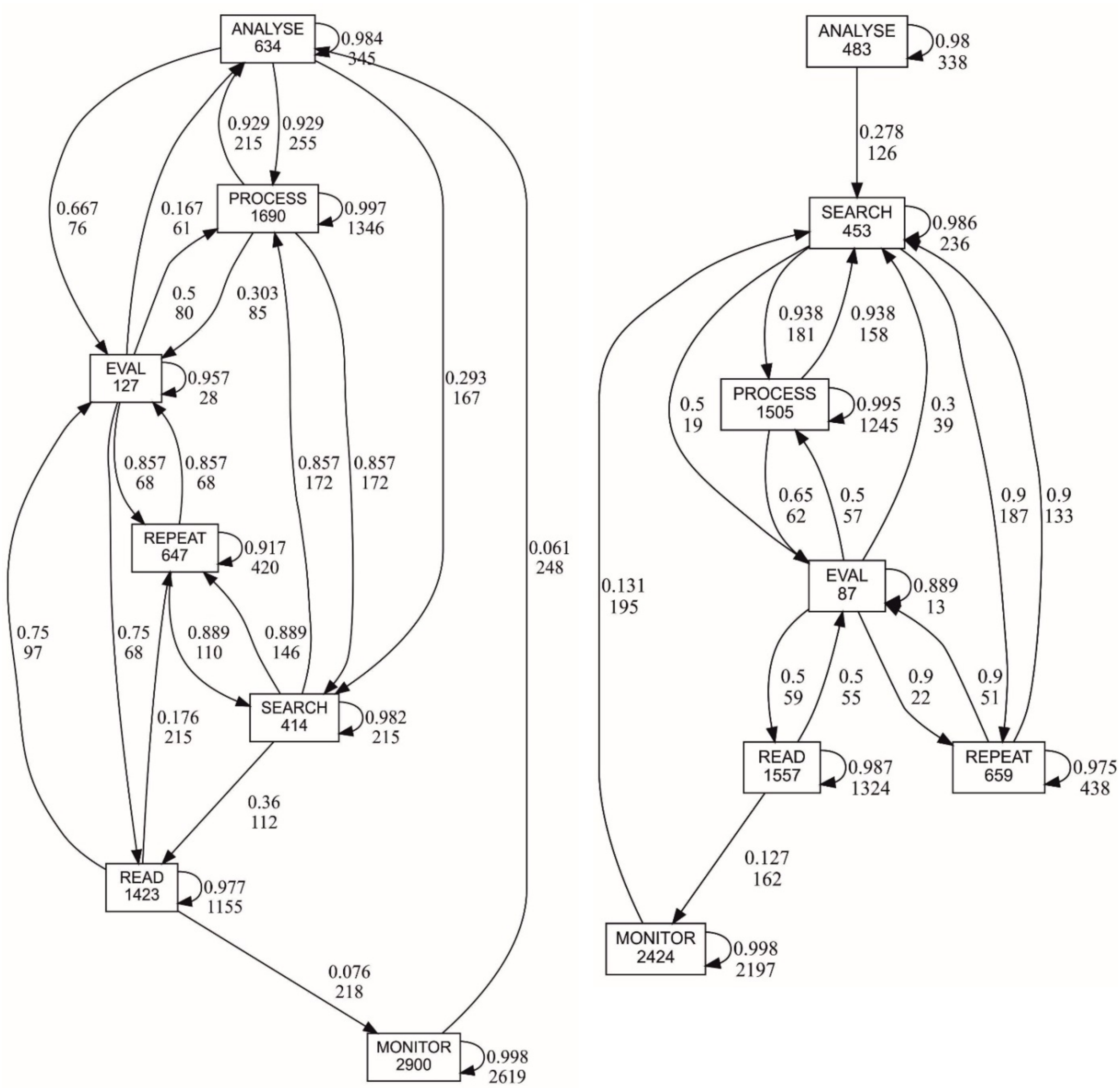
Table 4: Absolute frequencies, means, and test statistics of aggregated categories for the experimental group and the control group

	Experimental Group (n=35)					Control Group (n=35)							
	Min	Max	Absolute Frequency	*M*	*SD*	Min	Max	Absolute Frequency	*M*	*SD*	*t*	*p*	*d*
Analyze	5	37	634	18.11	8.58	3	33	483	13.80	6.61	2.356	.011	0.57
Search	1	32	414	11.83	7.54	1	57	453	12.94	10.64	-0.506	.310	-0.12
Evaluation	0	15	127	3.63	3.08	0	15	87	2.49	3.03	1.565	.061	0.38
Monitoring	20	211	2900	82.86	39.37	12	131	2424	69.26	30.71	1.612	.056	0.39
Reading	20	84	1423	40.66	15.75	19	89	1557	44.49	18.19	-0.941	.350	-0.03
Repeating	2	45	647	18.49	11.04	2	59	659	18.83	12.25	-0.123	.902	-0.03
Process	9	93	1690	48.29	21.54	3	84	1505	43.00	20.73	1.046	.299	0.25

Note: Since we expected metacognitive prompting to increase the number of metacognitive utterances, we conducted one-tailed testing for metacognitive categories; elsewhere we conducted two-tailed testing; $p < .05$.

(2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

The relative arrangement of learning activities was analyzed by applying the HeuristicsMiner algorithm in combination with the DWS mining plugin. The trace clustering did not split the cases into clusters for the participants of the experimental group (n=35) or the control group (n=35). This means a single process model can already express the event log with sufficient precision for both groups.



Control group *Experimental group*

Figure 4. Process models for the experimental group (n=35) and for the control group (n=35) represented as a heuristic net. Metacognitive Activities: ANALYZE = Orientation, Planning, and Goal specification; EVAL = Evaluation; MONITOR = Monitoring and Judgment. Cognitive Activities: READ = Reading; REPEAT = Repeating; PROCESS = Elaborate and Organization

(2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

The resulting process models are displayed in Figure 4, where they are represented as heuristic nets for the experimental and control groups. These visual representations of the process models comprise square boxes that represent the event classes and arcs between these boxes that indicate the dependency between two event classes. The number in the event box represents the occurrence of an event class in the log. The arcs are labelled with two types of information. The upper number displays the dependency measure, which indicates the certainty of a dependency relation between two activities. A value close to 1.0 indicates a high certainty that a dependency relation exists. The lower number shows the number of times this transition is used, that is, how often event *a* is followed by event *b*. An arc pointing back at the same box indicates a self-loop, meaning that an event class often occurred multiple times in a row in loops of length one or length two (e.g., ACCB, ACDCDB).

The fitness between the mined model and the event log used for generating this model was measured using the so-called *Improved Continuous Semantic Fitness* (De Medeiros, 2006; range: –∞ to 1.0). This fitness measure indicates the number of correct parsed event sequences, whereas a punishment for allowed extra behaviour in the model is subtracted from this number. The idea of this measure is to favour a process model that allows for less extra behaviour if several models can correctly parse the same number of event sequences. Both process models show a substantial fitness value: the experimental group model = 0.53, and the control group model = 0.62.

4.3.1 Process model of the experimental group

For the experimental group, a common pattern — that is, a path of transitions with high certainty — is ANALYZE → PROCESS → SEARCH → REPEAT → EVAL → READ → MONITOR → ANALYZE. Moreover, the process model comprises a number of loops with high certainty between two activities. Participants circle between ANALYZE and PROCESS, EVAL and REPEAT, SEARCH and PROCESS, EVAL and READING, and SEARCH and REPEAT. Apparently, these loops always occur between metacognitive and cognitive learning activities but never between two cognitive or two metacognitive events. Furthermore, it is interesting that EVAL is connected with several other learning events, meaning it takes an important position in the structure of the process, although this event class has a relatively low frequency. MONITOR only shows a weak connection in the process model. This event class follows READ and is followed by ANALYZE. Finally, the model shows self-loops for all event classes, indicating that an activity can be performed multiple times in a row.

4.3.2 Process model of the control group

The model of the control group shows the most common path of transitions for SEARCH → PROCESS → EVAL → REPEAT → READ → MONITOR → SEARCH. In contrast to the model of the students in the experimental group, ANALYZE is only weakly connected with SEARCH, and therefore, it is quite isolated. Similar to the experimental group, the low-frequency event class EVAL is also connected with several other learning activities. MONITOR is only weakly connected, whereas this event class follows READ, as in the model of the experimental group, but is followed by SEARCH instead of ANALYZE. In comparison with the experimental group, this process model shows fewer loops with high certainty between two activities (only between SEARCH and PROCESS, SEARCH and REPEAT, and EVAL and REPEAT), but again these loops only occur between metacognitive and cognitive events. Again, all event classes show self-loops.

Overall, the process models of the experimental and control groups especially differ in two points. First, ANALYZE (including the activities Orientation, Planning, and Goal specification) is hardly connected in the model of the control group, but this event class is well embedded in the process of the experimental group. Second, students in the experimental group show more loops between metacognitive and cognitive events, which can be interpreted as "regulation circles." For example, they circle with high certainty between ANALYZE and PROCESS and between EVAL and READ. Despite these differences, both models have in common that EVAL takes an important position in the described process. The frequency analysis could not reveal the importance of this event class, even showing that EVAL is one of the least-frequent categories. Here, the analysis of the sequential order provides additional information. Moreover, in both models, MONITOR, the metacognitive category with the highest frequency, is hardly connected with other learning activities. Based on recent SRL models, we argue that MONITOR does not have a clear position but can follow each learning activity (e.g., A → MONITORING → B → MONITORING → C → MONITORING). The HeuristicMiner algorithm could have failed to position this activity in the process model because its modelling notation does not allow for so-called *duplicate tasks* (i.e., an activity that has more than one label in the process model).

5 DISCUSSION AND IMPLICATIONS FOR FUTURE RESEARCH

In this study, we analyzed think-aloud data from an experimental study to investigate the effects of metacognitive prompts during learning on SRL processes. In addition to an analysis of frequencies of learning events, we focused on exploring the sequential structure of regulation activities using PM techniques.

As expected, the analysis of coded think-aloud data provides deeper insights into the effects of metacognitive prompts on students' regulatory processes during hypermedia learning. The findings of a frequency analysis indicate differences in the number of metacognitive utterances between students in the experimental group, who were prompted by metacognitive prompts, and those in the control group, who learned without prompts. Participants supported by metacognitive prompts articulated significantly more metacognitive activities and achieved better transfer performance. In addition, a mediation analysis revealed that prompting increased the number of metacognitive activities, especially Monitoring, which, in turn, increased the transfer performance. Both results are in line with findings of research on metacognitive prompting (e.g., Azevedo et al., 2004; Bannert, 2009) and the assumed effect mechanism of this type of metacognitive support (e.g., Bannert & Mengelkamp, 2013).

A microanalysis of the relative arrangement of learning activities was conducted by means of PM techniques to discover specific sequential patterns in the learning process of the experimental group versus the control group. This process analysis provided additional information on the effects of metacognitive prompts that could not be revealed by a simple analysis of frequencies of occurring

(2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

learning events. A comparison of the process models of students in the experimental and the control group showed two striking differences. First, activities of orientation, planning, and goal specification (aggregated as ANALYZE) are much better integrated in the process model of the experimental group, whereas this event class was quite isolated in the process model of the control group. Second, more loops between cognitive and metacognitive learning activities were identified in the process model of the experimental group, indicating that more regulation steps occurred. In conclusion, these differences indicate that the use of metacognitive prompts resulted in a better integration of ANALYZE events and a higher number of regulation loops. SRL models (e.g., Winne & Hadwin, 2008; Zimmerman, 2008) emphasize both the importance of orientation phases and an active regulation for successful learning. Following this, the fostered process patterns in the model of the experimental group are in line with current theoretical assumptions. We conclude that these process patterns could be successfully scaffolded through the application of metacognitive prompts. However, the evaluation of learning progress (EVAL) is similarly integrated in both process models, meaning no different effect on this event category could be detected. The findings of the process patterns could be used to optimize our metacognitive prompts further. For example, the design of prompts could be optimized by aiming at scaffolding the sequential deployment of evaluating activities in more detail. SRL models suggest that evaluation activities are followed by an update of the orientation phase. This transition was not represented in the process model of the experimental group. Here, an optimization process could be used.

With respect to the metacognitive support used in this study — that is, an introduction about what metacognitive prompts are, why they are important, and how to use them in combination with metacognitive prompts during learning — it is necessary to discuss which components of support have contributed to the findings. Based on our experience with metacognitive prompts and research on metacognitive prompting, at least a brief training or an introduction to the concept of metacognitive prompts is necessary in advance to guarantee an adequate application of prompts during learning (e.g., Bannert & Mengelkamp, 2013; Bannert & Reimann, 2012). Therefore, it is challenging to determine the individual effect of both the introduction and prompting components. To our knowledge, there is no empirical study that systematically compares the impact of training of prompt use, metacognitive prompting, and their combination. Consequently, this research question should be addressed in future work.

As an inductive approach, the validity of PM depends on the representativity and quality of the data stored in the event log (Reimann, Frerejean, & Thompson, 2009). It is possible that SRL processes, for example, those obtained by think-aloud protocols or log files, comprise a high variety of regulatory behaviour (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; Winne, 2014). Therefore, we applied a technique for trace clustering in combination with a PM algorithm to check whether a single process model for the whole event log is appropriate. The applied trace clustering did not split the cases into subsets of participants. However, new approaches of trace clustering are currently rising in the PM domain (e.g., de Weerdt et al., 2013). These approaches could possibly improve the detection of homogenous subsamples, which in turn could enhance the quality of the mined process models. There are approaches to clustering students according to their interactions and activities in computer-based learning environments based on a set of variables (e.g., Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010; Bouchet, Harley, Trevors, & Azevedo, 2013). However, these approaches do not explicitly include an event-centred perspective and a timing aspect, but are based on frequencies of interactions with the learning environment. Furthermore, a subset of participants can also be selected for process analysis based on learner characteristics (e.g., high vs. low prior knowledge) or learning outcomes (e.g., high vs. low achieving students; see Bannert et al., 2014 for an example).

Regarding further limitations of our analysis, it has to be noted that the resulting process models are dependent on the learning setting (learning environment, learning material, and instructions on the learning task). In addition, they are descriptive models. Moreover, findings depend on the underlying coding scheme and its level of granularity. In general, more research on PM techniques in the field of SRL and metacognition is needed; for example, for deriving guidelines for parameter settings aiming to improve the quality of a mined process model. Therefore, we encourage other researchers to use PM techniques to analyze their process data with respect to the sequential and temporal characteristics of learning events. In addition, a comparison of different methods for sequential and temporal analyses on the same data would be beneficial for discovering the advantages and disadvantages of recent process analysis methods. For example, a comparison could be made of different approaches presented in a special issue on the sequential and temporal characteristics of self- and socially regulated learning (Molenaar & Järvelä, 2014).

The resulting process models of our analysis represent a description of the underlying learning processes in our sample of students. In future studies, the validity of the discovered process patterns should be investigated by checking the conformance of these models to new data sets. For this purpose, the mined process models of the HeuristicsMiner algorithm can be converted into petri nets, and then methods for conformance checking can be applied within the ProM framework (Rozinat & van der Aalst, 2008). In this way, the conformance — that is, the differences between a discovered process model and a new event log — can be determined. Another possible scenario for the application of conformance checking would be the derivation of a system of event sequences on the micro-level based on the theoretical assumptions of SRL models. An illustration of this approach is presented in Bannert et al. (2014). However, more micro-level theories would be needed for this approach. At the moment, only the COPES model (Winne & Hadwin, 2008) provides a detailed level of granularity regarding information processing, but even this model is far from the level of elaboration needed to correspond directly to the granularity of our event data.

Finally, an advantage of PM techniques is the representation of sequential characteristics as visual process models. Primarily, this helps the researcher to grasp easily the course of learning activities and regulatory patterns. However, this type of visual representation can also be used to give process feedback to the students and, thereby, be a resource for learners as well (Reimann et al., 2009). Winne

(2014) recommends supporting students with information on their past learning processes through displays of traces they can interpret. Following this direction in future research on SRL, process models generated by PM techniques could make a substantial contribution in providing feedback to learners.

ACKNOWLEDGMENTS

This research was supported by funds from the German Research Foundation (DFG: BA 2044/7-1). We thank Ferdinand Guggenberger, Valentina Haspert, Michaela Hofbauer and Lisa Müsse, the student research assistants of this project, for their valuable help for coding the think-aloud data.

REFERENCES

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist*, *40*(4), 199–209. doi:10.1207/s15326985ep4004_2

Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning*, *4*(1), 87–95. doi:10.1007/s11409-009-9035-7

Azevedo, R., Cromley, J. G., Moos, D. C., Greene, J. A., & Winters, F. I. (2011). Adaptive content and process scaffolding: A key to facilitating students' self-regulated learning with hypermedia. *Psychological Testing and Assessment Modeling*, *53*(1), 106–140.

Azevedo, R., Guthrie, J., & Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research*, *30*(1–2), 87–111. doi:10.2190/DVWX-GM1T-6THQ-5WC7

Azevedo, R., Trevors, G., Duffy, M., Feyzi-Behnagh, M., Bouchet, F., & Landis, R. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 427–449). New York, NY: Springer. doi:10.1007/978-1-4419-5546-3_28

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research*, *35*(4), 359–375. doi:10.2190/94V6-R58H-3367-G388

Bannert, M. (2007). *Metakognition beim Lernen mit Hypermedia. Erfassung, Beschreibung und Vermittlung wirksamer metakognitiver Lernstrategien und Regulationsaktivitäten [Metacognition and hypermedia learning]*. Münster: Waxmann.

Bannert, M. (2009). Promoting self-regulated learning through prompts: A discussion. *Zeitschrift für Pädagogische Psychologie / German Journal of Educational Psychology*, *23*(2), 139–145. doi:10.1024/1010-0652.23.2.139

Bannert, M., & Mengelkamp, C. (2013). Scaffolding hypermedia learning through metacognitive prompts. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 171–186). New York, NY: Springer. doi:10.1007/978-1-4419-5546-3_12

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, *40*(1), 193–211. doi:10.1007/s11251-011-9167-4

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, *9*(2), 161–185. doi:10.1007/s11409-013-9107-6

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior, 52*, 293–306. doi:10.1016/j.chb.2015.05.038

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology Enhanced Learning*, *5*(2), 123–152. doi:10.1142/S1793206810000839

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York, NY: D. McKay.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*(2), 161–186. doi:10.1016/S0959-4752(96)00015-1

Bouchet, F., Harley, J. M., Trevors, G. J., & Azevedo, R. (2013). Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning. *Journal of Educational Data Mining*, *5*(1), 104–146.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, *6*(3), 271–315. doi:10.1207/s15327809jls0603_1

De Medeiros, A. K. A. (2006). *Genetic Process Mining*. Eindhoven University of Technology, Eindhoven.

De Medeiros, A. K. A., Guzzo, A., Greco, G., van der Aalst, W. M. P., Weijters, A. J. M. M., van Dongen, B. F., & Saccà, D. (2008). Process mining based on clustering: A quest for precision. *Lecture Notes in Computer Science*, *4928*, 17–29. doi:10.1007/978-3-540-78238-4_4

De Weerdt, J., de Backer, M., Vanthienen, J., & Baesens, B. (2012). A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Information Systems*, *37*(7), 654–676. doi:10.1016/j.is.2012.02.004

De Weerdt, J., vanden Broucke, S., Vanthienen, J., & Baesens, B. (2013). Active trace clustering for improved process discovery. *IEEE Transactions on Knowledge and Data Engineering*, *25*(10), 2708–2720. doi:10.1109/TKDE.2013.64

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Fluxicon Disco (Version 1.7.2) [Computer Software]. (2014). Retrieved from http://fluxicon.com/disco/

Ge, X. (2013). Designing learning technologies to support self-regulation during III-structured problem-solving processes. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 213–228). New York, NY: Springer. doi:10.1007/978-1-4419-5546-3_15

Greco, G., Guzzo, A., Pontieri, L., & Saccà, D. (2006). Discovering expressive process models by clustering log traces. *IEEE Transactions on Knowledge and Data Engineering*, *18*(8), 1010–1027. doi:10.1109/TKDE.2006.123

Greene, J. A., Dellinger, K. R., Tüysüzoglu, B. B., & Costa, L.-J. (2013). A two-tiered approach to analyzing self-regulated learning data to inform the design of hypermedia learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 117–128). New York, NY: Springer. doi:10.1007/978-1-4419-5546-3_8

Günther, C. W., & van der Aalst, W. M. P. (2007). Fuzzy mining: Adaptive process simplification based on multi-perspective metrics. *Business Process Management: Lecture Notes in Computer Science*, *4714*, 328–343. doi:10.1007/978-3-540-75183-0

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*(2), 107–124. doi:10.1007/s11409-007-9016-7

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: The Guilford Press.

Hoffman, B., & Spatariu, A. (2011). Metacognitive prompts and mental multiplication: Analyzing strategies with a qualitative lens. *Journal of Interactive Learning Research*, *22*(4), 607–635.

Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. (2008). Using hidden Markov models to characterize student behaviors in learning-by-teaching environments. *Lecture Notes in Computer Science*, *5091*, 614–625. doi:10.1007/978-3-540-69132-7_64

Johnson, A. M., Azevedo, R., & D'Mello, S. K. (2011). The temporal and dynamic nature of self-regulatory processes during independent and externally assisted hypermedia learning. *Cognition and Instruction*, *29*(4), 471–504. doi:10.1080/07370008.2011.610244

Kapur, M. (2011). Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *International Journal of Computer-Supported Collaborative Learning*, *6*(1), 39–56. doi:10.1007/s11412-011-9109-9

Li, C., & Biswas, G. (2002). A Bayesian approach for structural learning with hidden Markov models. *Scientific Programming*, *10*(3), 201–219.

Lin, X. (2001). Designing metacognitive activities. *Educational Technology Research and Development*, *49*(2), 23–40. doi:10.1007/BF02504926

Lin, X., Hmelo, C., Kinzer, C. K., & Secules, T. J. (1999). Designing technology to support reflection. *Educational Technology Research and Development*, *47*(3), 43–62. doi:10.1007/BF02299633

Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, *36*(7), 837–858. doi:10.1002/(SICI)1098-2736(199909)36:7<837::AID-TEA6>3.3.CO;2-L

Martin, T., & Sherin, B. (2013). Learning analytics and computational techniques for detecting and evaluating patterns in learning: An introduction to the special issue. *Journal of the Learning Sciences*, *22*(4), 511–520. doi:10.1080/10508406.2013.840466

Molenaar, I., & Chiu, M. M. (2014). Dissecting sequences of regulation and cognition: Statistical discourse analysis of primary school children's collaborative learning. *Metacognition and Learning*, *9*(2), 137–160. doi:10.1007/s11409-013-9105-8

Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, *9*(2), 75–85. doi:10.1007/s11409-014-9114-2

Moos, D. C., & Azevedo, R. (2009). Self-efficacy and prior domain knowledge: To what extent does monitoring mediate their relationship with hypermedia learning? *Metacognition and Learning*, *4*(3), 197–216. doi:10.1007/s11409-009-9045-5

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*(2), 93–115. doi:10.1037/a0022658

ProM (Version 5.2) [Computer Software]. (2008). Retrieved from http://www.processmining.org

Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 239–257. doi:10.1007/s11412-009-9070-z

Reimann, P., Frerejean, J., & Thompson, K. (2009). Using process mining to identify models of group decision making processes in chat data. In C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), *Computer-Supported Collaborative Learning Practices: CSCL2009 Conference Proceedings* (pp. 98–107). Rhodes, Greece: International Society for the Learning Sciences.

Reimann, P., & Yacef, K. (2013). Using process mining for understanding learning. In R. Luckin, P. Puntambekar, P. Goodyear, B. Grabowski, J. Underwood, & N. Winters (Eds.), *Handbook of design in educational technology* (pp. 472–481). New York, NY: Routledge.

Reisig, W. (1985). *Petri Nets: An introduction*. Berlin: Springer.

Richter, T., & van Holt, N. (2005). ELVES: Ein computergestütztes Diagnostikum zur Erfassung der Effizienz von Teilprozessen des Leseverstehens [ELVES: Computer-based assessment of the efficiency of processes involved in reading comprehension]. *Diagnostica*, *51*(4), 169–182. doi:10.1026/0012-1924.51.4.169

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. (2010). *Handbook of educational data mining*. Boca Raton, FL: Chapman & Hall/CRC.

Rozinat, A. (2010, October 18). ProM tips: Which mining algorithm should you use? [Web log post]. Retrieved from http://fluxicon.com/blog/2010/10/prom-tips-mining-algorithm/

Rozinat, A., & van der Aalst, W. M. P. (2008). Conformance checking of processes based on monitoring real behavior. *Information Systems*, *33*(1), 64–95. doi:10.1016/j.is.2007.07.001

Schnotz, W. (1998). Strategy-specific information access in knowledge acquisition from hypertext. In L. B. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.), *Discourse, tools, and reasoning: Essays on situated cognition* (pp. 336–358). Berlin: Springer.

Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior*, *28*(4), 1321–1331. doi:10.1016/j.chb.2012.02.016

Trčka, N., Pechenizkiy, M., & van der Aalst, W. M. P. (2010). Process mining from educational data. In C. Romero, S. Ventura, M. Pechenizkiy, & R. Baker (Eds.), *Handbook of educational data mining* (pp. 123–142). Boca Raton, FL: Chapman & Hall/CRC.

Van der Aalst, W. M. P. (2011). *Process mining: Discovery, conformance and enhancement of business processes*. Berlin/Heidelberg: Springer. doi:10.1007/978-3-642-19345-3

Van der Aalst, W., Weijters, T., & Maruster, L. (2004). Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, *16*(9), 1128–1142. doi:10.1109/TKDE.2004.47

Veenman, M. (1993). *Metacognitive ability and metacognitive skill: Determinants of discovery learning in computerized learning environments*. Amsterdam: University of Amsterdam.

Veenman, M. V. J. (2007). The assessment and instruction of self-regulation in computer-based environments: A discussion. *Metacognition and Learning*, *2*(2–3), 177–183. doi:10.1007/s11409-007-9017-6

Veenman, M. V. J., van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*(1), 3–14. doi:10.1007/s11409-006-6893-0

Weijters, A. J. M. M., van der Aalst, W. M. P., & de Medeiros, A. K. A. (2006). Process mining with the HeuristicsMiner algorithm. *BETA Working Paper Series 166.* Eindhoven, Netherlands: Eindhoven University of Technology.

Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, *9*(2), 229–237. doi:10.1007/s11409-014-9113-3

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). New York, NY: Taylor & Francis.

Winne, P. H., & Hadwin, A. F. (2013). nStudy: Tracing and supporting self-regulated learning in the internet. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 293–308). New York, NY: Springer. doi:10.1007/978-1-4419-5546-3_20

Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 259–277). New York, NY: Routledge.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego, CA: Academic Press.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183. doi:10.3102/000283120731290

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments

Alexander Nussbaumer
Knowledge Technologies Institute, Graz University of Technology, Austria

Eva-Catherine Hillemann
Knowledge Technologies Institute, Graz University of Technology, Austria

Christian Gütl
Institute for Information Systems and Computer Media, Graz University of Technology, Austria
School of Information Systems, Curtin University, Perth, Western Australia

Dietrich Albert
Knowledge Technologies Institute, Graz University of Technology, Austria
Department of Psychology, University of Graz, Austria

alexander.nussbaumer@tugraz.at

ABSTRACT: This paper presents a conceptual approach and a Web-based service that aim at supporting self-regulated learning in virtual environments. The conceptual approach consists of four components: 1) a self-regulated learning model for supporting a learner-centred learning process, 2) a psychological model for facilitating competence-based personalization and knowledge assessment, 3) an open learner model approach for visual interaction and feedback, and 4) a learning analytics approach for capturing relevant learner information required by the other components. The Web-based service provides a technical implementation of the conceptual approach, as well as a linkage to existing virtual environments used for learning purposes. The approach and service have been evaluated in user studies in university courses on computer science to demonstrate the usefulness of the overall approach and to get an understanding of some limitations.

KEYWORDS: Self-regulated learning, competence-based knowledge space theory, learning analytics, personalization, reflection, learning environments

1 INTRODUCTION

This paper presents an approach and a service that support learners to learn in a self-regulated way. It has been shown in many studies that self-regulated learning (SRL) has many positive effects on the learning process, such as better learning in terms of being able to monitor, evaluate, and plan the learning process effectively, having better time-and-effort management, and demonstrating higher motivation for learning (e.g., Pintrich, 2000; Pintrich & De Groot, 1990; Zimmerman, 2008). Hence,

ISSN 1929-7750 (online).

learners able to learn in a self-regulated way can achieve better results. However, there are also studies suggesting that many learners have problems with this way of learning. Taking over control of one's own learning process and applying metacognitive strategies (i.e., monitoring, evaluating, and planning the learning) require specific metacognitive skills that not all students have (Mikroyannidis et al., 2013). Therefore, learners often need guidance on different levels for learning in a self-regulated manner (Law et al., 2012).

Though many technology-enhanced learning (TEL) solutions have evolved over the last decades, very few of them support SRL. Learning Management Systems (LMS) often used in educational settings, such as Moodle or Sakai, have become very popular (Paulsen, 2003). They focus primarily on distributing learning content, organizing the learning processes, and serving as an interface between learner and teacher. However, courses designed in LMS usually do not give learners much freedom, but rather provide a predefined learning trajectory. In contrast, Personal Learning Environments (PLE) (Henri, Charlier, & Limpens, 2008) strive for a more natural, learner-centric approach characterized by the freedom that individual learners have to select and control the services and tools they use. While this approach allows better opportunities for self-regulated learners, it still lacks guidance and help for learners with poor SRL skills.

Another category of TEL solutions include adaptive systems and intelligent tutoring systems that aim to tailor their content and behaviour to the needs and preferences of learners (Brusilovsky, Kobsa, & Nejdl, 2007). These systems are supposed to make learning more efficient by guiding the learner through the learning process. However, a self-regulated learner should not give control to a system, but take over the control on his own. The most promising approach towards SRL support is the concept of Open Learner Models (OLM). While adaptive systems make use of learner information (e.g., knowledge state) by using a learner model to adapt a course or system behaviour, the OLM approach pursues the idea of displaying learner information by making the model open to the learner and letting him choose the next steps. The main aim of the OLM approach is to support the learner's reflection process by providing formative feedback on the learning process, which has a positive effect on the learning outcome (Bull & Kay, 2010).

Having the OLM concept in mind, the approach and service presented in this paper aim at supporting SRL on a broader scale. In addition to reflection support, there is also support for planning, goal setting, self-monitoring, and self-evaluation. This is achieved by defining a self-regulated learning process that includes the aforementioned metacognitive activities. Individual tools with interactive visual interfaces for the diverse metacognitive support strategies provide technical support. On a conceptual level, this is achieved by combining the self-regulated learning approach with Competence-based Knowledge Space Theory (Albert & Lukas, 1999) and resulting personalization strategies. The technical infrastructure in the background manages learner models and domain models representing the subject domain. It also contains a learning analytics component for analyzing the learner's behaviour and providing visual information and recommendations.

This paper is structured as follows: the next section gives an overview of the relevant state of the art and theoretical background used for the conceptual approach, followed by research questions derived from a literature review. We then explain a new conceptual approach for SRL support, describe the technical design of the models and developed service, and explain the learner-centred approach with a use case example. An evaluation of this approach and service conducted in a university course on computer science is reported in the section following. Results, limitations, and opportunities are discussed in next. Finally, concluding remarks and future work is described in final section.

2 THEORETICAL BACKGROUND AND RELATED WORK

2.1 Self-regulated Learning

From a psychological and pedagogical point of view, self-regulated learning is a complex field of research that combines motivational as well as cognitive and personality theories. Components of SRL are cognition, metacognition, motivation, affects, and volition (Kitsantas, 2002). According to Zimmerman (2002), students can be described as self-regulated to the degree that they are metacognitively, motivationally, and behaviourally active participants in their own learning process. To define students' learning as self-regulated, they have to use specific strategies for attaining their goals and their learning behaviour has to be based on self-efficacy perceptions. In self-regulated learning, learners are active and able to control, monitor, and regulate their cognition, motivational state, behaviour, and context. Furthermore, learners set goals and try to achieve them through progress monitoring. These self-regulatory activities are mediators between personal characteristics, contextual features, and actual performance in the learning process. In a meta-analysis conducted by Hattie (2009), it turned out that self-regulated learning is one of the most effective methods to reach learning goals.

Zimmerman (2002) has developed a *cyclic SRL model* consisting of three phases: the forethought phase (i.e., goal setting or planning), the performance phase (i.e., self-observation processes), and the self-reflection phase (i.e., self-reflection processes). According to this model, learning performance and behaviour consist of both cognitive and metacognitive activities. Cognitive activities are related to dealing with subject domains; for example, acquiring domain knowledge through reading. Metacognitive activities are related to thinking about and regulating the cognitive activities; for example, making a plan about domain knowledge acquisition.

Boekaerts (1999), who developed the *three-layered SRL model*, pursued a similar approach. This model deals with cognitive and metacognitive activities, as well as with goals and resources. The first layer deals with regulation of the self, which is related to the choice of goals and resources that learners make. The second layer focuses on the regulation of the learning process, which relates to the use of metacognitive skills to direct the learning process. The third layer describes the regulation of the processing modes, which describes the choice of cognitive strategies.

According to Roberts & Erdos (1993), *metacognition* is a key concept in the study of cognition and it plays an important role in the transfer of cognitive skills and in problem solving. Often the term metacognition is simply defined as “thinking about thinking” or “cognition of cognition” (Flavell, 1976). This means that metacognition can be understood as the competence of reflecting on a mental task critically and of organizing the relevant learning and thinking processes in an efficient and effective way. Treier (2004) describes metacognition by the sub-components of self-monitoring, self-observation, and self-regulation related to cognitive and information processing. The usage of metacognitive strategies is an essential component of self-regulated learning and is very important for flexibility and personalization (Efklides, 2009).

A key aspect of SRL is the learner’s use of different cognitive and metacognitive strategies with the aim of controlling and regulating their learning (Pintrich, 1999). These strategies relate to being effective in learning, being able to self-regulate and control cognition (learning about learning), and being effective in applying resource management strategies. Dabbagh & Kitsantas (2004) summarized six key processes essential for SRL:

1. In the *goal-setting* process, the outcome of a learning process is defined and strategies are identified for how to reach these goals. Goal setting motivates the learner’s choice of and attention to the relevant tasks and motivates the learner towards higher effort and higher persistence over the course of time (Zimmerman, 2002). Furthermore, goal setting influences learning through affective reactions; for example, higher self-satisfaction when goals are reached.
2. *Self-monitoring* is defined as one’s reflected attention to an aspect of behaviour that directs the learners’ attention to the task and assists them in evaluating the outcomes of their efforts. Self-monitoring is important, because it helps learners attain their goals by adjusting their learning.
3. *Self-evaluation* is the process by which the learner compares the learning outcome with his own goals. It fosters better skill acquisition, self-efficacy beliefs, intrinsic interest, and self-satisfaction about performance.
4. *Task strategies* are defined as the processes of a learner who applies strategies that help reach his own goals. Studies indicate that students who applied strategies for learning had better performance than students who did not regularly apply them (Pintrich, 1990).
5. *Help-seeking* is taking place if a learner identifies and calls upon outside resources, not only human, but also analogue and digital resources.
6. *Time management* is the process by which learners manage their learning regarding time. Effective time budgeting highly correlates with academic achievement.

Supporting SRL in the right way is a crucial factor. On the one hand, it means providing enough freedom for the learner, in order to stimulate motivation. However, on the other hand, too much freedom may be overwhelming and appropriate guidance or even adaptation is usually needed to make the learning process effective and efficient. The concept of *guidance and freedom* is important, because it has been recognized that highly motivated learners attain better learning performance if they have more control over their learning and are more autonomous (Issing, 2002). On the other hand, some learners show difficulties in carrying out concrete metacognitive activities, such as planning, goal setting, monitoring,

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

evaluating, and as a result often perform less successfully than would be anticipated (Bannert, 2006). Such learners are in need of guidance. Furthermore, less motivated learners can also improve their performance if they receive more guidance. Keeping these reported findings in mind, individual support for learners should be tailored to suitable degrees of guidance and freedom. In this respect, the learner should be offered an optimal and balanced level of control and autonomy for his own learning process.

Motivation is a highly relevant aspect for achieving good learning outcomes and for performing self-regulated learning activities. Winne & Hadwin (2008) showed the positive impact of motivation on student attention to the learning progress, on the progress itself, and on the experience of satisfaction and positive affect. For the use of self-regulated learning activities, a learner has to be motivated, as these activities require additional time and effort. Ryan & Deci (2000) describe intrinsic motivation as one of the most important aspects regarding learning, because it is the prototypical manifestation of the human tendency toward learning and creativity. However, there is also a need for extrinsic motivation and especially a good balance between extrinsic and intrinsic motivation (Covington, 2000).

2.2 Learning Analytics

The Society for Learning Analytics Research (SoLAR) defines Learning Analytics (LA)[1] as “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.” Siemens (2010) describes learning analytics as “the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning.”

In the NCM Horizon Report, learning analytics is described as a rapidly developing trend in higher education, where learning is happening more and more within online and hybrid environments (Johnson, Adams Becker, Estrada, & Freeman, 2014). According to this report, learning analytics can potentially help transform education from a standard one-size-fits-all approach into responsive and flexible frameworks. Such frameworks are capable of adapting content and software behaviour to the needs of learners and provided tailored recommendations and visualizations. New kinds of visualizations and analytical reports are being developed to guide administrative bodies with empirical evidence because they help assess and improve the effectiveness of programs, schools, and entire school systems, which aids in the proper allocation of resources.

Duval (2011) outlines several possibilities about how learning analytics concepts and methods can be put into practice: 1) goal-oriented visualizations to impact the further behaviour of the user; 2) technical infrastructure to model the captured data; 3) dashboard applications to give a visual overview of the collected data and often relate that data to other learners. The captured data can also be used for learning recommendations of resources, activities, and people.

[1] See http://solaresearch.org/

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

In a learning analytics process model (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013), four stages of personal information and their applications are distinguished: awareness, reflection, sensemaking, and impact. Awareness is concerned with data represented as activity streams, tabular representations, or other visualization types. The reflection stage focuses on the user who asks questions regarding the provided data in order to investigate their usefulness and relevance. The sensemaking stage is about answering the posed questions to obtain new insights. Finally, the impact stage relates to the induction of new meaning or behaviour change. While most learning analytics approaches target reflection support, focusing on learning analytics infrastructure helps learners develop learning dispositions and transferable competencies, such as critical curiosity or creativity (Buckingham Shum & Crick, 2012).

2.3 Competence-based Knowledge Space Theory and Adaptation Approaches

Individually adapting learning methods to suit the learner's characteristics has a big influence on performance (Issing, 2002). The importance of the *adaptation* to the learner's characteristics (also called *personalization*) has been shown in several studies. For example, adaptive subject material combined with adaptive styles of presentation supports students to improve their learning achievements and increases learning efficiency (Tseng, Chu, Hwang, & Tsai, 2008). Through a requirement analysis, it has been found that the learner's knowledge, goals and tasks, language, and interests are important factors of personalization approaches (Höver & Steiner, 2009).

Knowledge Space Theory (KST) and its competence-based extensions (CbKST) are prominent examples of adaptation strategies grounded on theoretical framework (Hockemeyer, 2003). KST constitutes a psychological mathematical framework for both structuring knowledge domains and representing the knowledge of learners (Albert & Lukas, 1999). Due to dependencies between problems, prerequisite relations can be established. The knowledge state of a learner is identified with the subset of all problems this learner is capable of solving. By associating assessment problems with learning objects, a structure on learning objects can be established, which constitutes the basis for meaningful learning paths adapted to the learner's knowledge state.

Competence-based Knowledge Space Theory (CbKST) incorporates psychological assumptions on underlying skills and competences required for solving specific problems (Korossy, 1997; Heller Steiner, Hockemeyer, & Albert, 2006). In this approach, competences are assigned to both learning objects (taught competences) and assessment items (tested competences). Similar to the knowledge state, a competence state can be defined consisting of a set of skills that the learner has available. Furthermore, there may also be relationships between competences modelled in a prerequisite relation structure. CbKST provides adaptive assessment algorithms for efficiently determining the learner's current knowledge and competence state, which builds the basis for personalization purposes. Based on this learner information, personalized learning paths can be created. Goal setting can be done by defining skills to be achieved (competence goal) or problems to be capable of solving. The competence gap to be closed during learning is represented by the skills that are part of the goal, but not part of the competence state of a learner.

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

From a broader perspective, adaptation approaches have a long tradition in educational settings. For example, Intelligent Tutoring Systems (ITS) were supposed to bring intelligence to computer-based instruction, especially in the knowledge of the subject domain, as well as the tutoring principles and methods of their application (Anderson, 1988). This led to the development of four basic components: the domain model, the student model, the tutoring model, and the user interface model. Adaptive hypermedia systems in educational contexts are based on these models and aim at personalizing the learning experience (Brusilovsky et al., 2007). Finally, Open Learner Model (OLM) approaches (Bull & Kay, 2010) go one step further and make use of these models for pedagogical reasons by visualizing and presenting the information to the learner (see also the introductory section of this paper).

2.4 Summary and Resulting Research Questions

This section has presented three prominent learning concepts successfully applied in technology-enhanced learning. Each of them covers a specific field in the learning domain. Self-regulated learning focuses on metacognition and on how learners take over responsibility of their own learning process. Competence-based Knowledge Space Theory focuses on how the learning process can be structured in terms of the subject domain. Learning Analytics has its strengths in observing the learner and creating benefit from these observations. Personalization strategies and Open Learner Models are employed by the latter two concepts. However, each of them alone does not cover the full spectrum of the learning process. Learning is a complex process and ideally covers all these aspects; therefore, it makes sense to elaborate a solution that combines these learning concepts with their specific characteristics. An approach to integrating SRL and CbKST on a conceptual level was elaborated in previous work (Steiner, Nussbaumer, & Albert, 2009).

The next section presents a solution that combines the ideas and methods of the presented learning concepts. This leads to a new solution approach that smoothly integrates the methods and characteristics of these learning concepts by focusing on the benefit for the learner. This endeavour can be expressed in three research questions:

- **R1:** How can learners be supported in following a self-regulated learning paradigm and how can we ensure that they achieve their learning goals on the domain level at the same time?
- **R2:** How can Competence-based Knowledge Space Theory, Learning Analytics, personalization, and Open Learner Models be combined to design a holistic framework that supports self-regulated learning?
- **R3:** How should a system be designed that implements this conceptual approach?

3 CONCEPTUAL APPROACH AND TECHNOLOGY

3.1 Psycho-Pedagogical Framework

The central piece of the psycho-pedagogical framework is the ***domain model*** that formally covers the subject domain to be learned. Competence-based Knowledge Space Theory (CbKST) provides a methodology to structure a subject domain and also makes it usable for a technical application. The core

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

elements as used in this framework are learning objects, assessment items, and competences. Competences are structured through prerequisite relations, meaning that it can be assumed that, if competence A is a prerequisite of competence B, then a learner having competence B also has available competence A. Figure 1 depicts a prerequisite structure of five competences (A, B, C, D, E) in an acyclic directed graph. Competences below others are prerequisite for them, if connected through a path to them. This type of graph is also called Hasse diagram.

In order to structure a subject domain, the first step is to define competences necessary to master it. These competences are then structured according to their prerequisite relations into a competence map. In addition to competences, learning objects are created that convey these competences and assessment items are defined that test these competences. Thus, relations between learning objects, assessment items, and competences are established. The set of all elements and their relations is called the domain model (see Figure 1). Such a domain model is usually created by a domain expert or teacher.

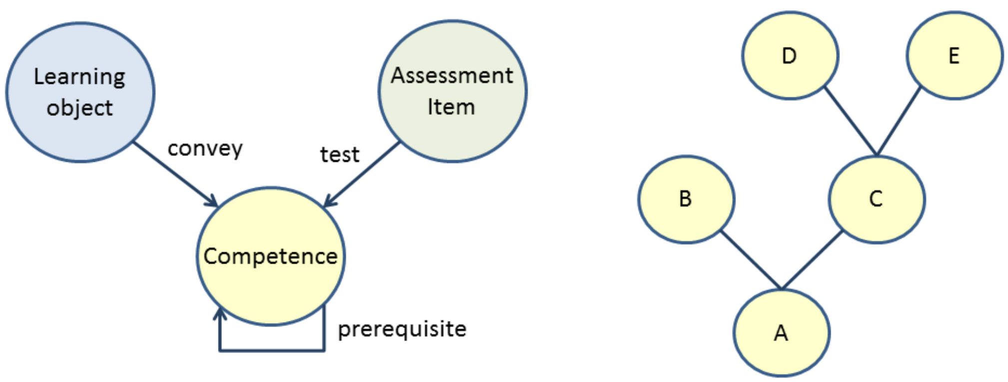


Figure 1. The diagram depicts the domain model structure on the left site and an example of five competences with prerequisite relations between them on the right side.

The ***user model*** is another core component of the framework. In general, the user model follows an overlay design (Brusilovsky et al., 2007) meaning that the user model elements are related to and defined by other models. In our case, the user model refers to the domain model and the tool interaction model (see below). The user model allows for defining information about the learner's knowledge, competences, goals, history, and used tools. It relates to both aspects of the domain model and aspects of the self-regulated learning, and thus connects them. Table 1 gives an overview of the information contained in the user model.

Further important components of the psycho-pedagogical framework are the ***learning process*** and ***learning cycle***. The learning process is a rather generic term that refers to all cognitive and metacognitive activities of the learner, as well as the interactions with the learning system. The learning process starts when the learner begins to learn and ends when the learner has finished learning (independent of the results). In order to operationalize the learning process, a self-regulated learning process model is defined that connects self-regulated learning concepts with elements of the domain model. The learning cycle is related to the iterations within this process (see below).

Table 1. Overview of the user model and the contained information.

Category	Information/Items	Description
Learning State	Knowledge State	The assessment items that a learner can solve
	Competence State	The competences a learner has available; each time a learner solves (or fails to solve) an assessment item, the related competences are added to or removed from the competence state
	Competence Goal	The competences that a learner wants to acquire
	Competence Gap	The competences needed to achieve the competence goal
Learning History/ Learning Behaviour	Learning Objects	The learning objects a learner has visited
	Assessment Items	Answers to assessment items including the information if solved or failed
	Learning Tool	The tools and their submenus that a learner has used
	Help	The help information that a learner has requested (related to a tool)

So far, the described concepts can be employed in different learning approaches. Traditional adaptive systems use domain and user models to create learning paths automatically through the learning objects and assessment items. Though such approaches are personalized, the learners just follow the paths without having alternative options. The other extreme would be an approach where all learning objects and assessment items are available and the learner is free to choose, but without any help, recommendations, or feedback from such a system. While the first type of approach might be efficient for weak learners by offering strict guidance, it does not provide the advantages of self-regulated learning. On the other hand, the second type of approach has its weaknesses, because it does not provide any support, which is also helpful for learners familiar with this type of learning. The goal of this framework is to establish an approach that offers freedom and support at the same time.

In order to facilitate support and guidance in a technical environment that does not restrict the freedom of the learner, a ***self-regulated learning process model*** is defined. This model follows the ideas of the cyclic SRL model of Zimmerman (2002) and SRL activities defined by Dabbagh & Kitsantas (2004). It describes learning as a cyclic sequence of four main phases: 1) planning and goal setting, 2) using learning resources, 3) knowledge and competence assessment, and 4) reflecting on learning behaviour and progress. For each of these phases, a visual tool that supports the respective cognitive and metacognitive activities is provided. In the *planning and goal-setting* phase learners set their short-term goals, in order to plan what they want to learn next. This phase is mainly related to the metacognitive activity of planning. In the next phase, the learners make use of learning resources fitting to the selected goals, in order to attain related domain knowledge. Learning in this phase mainly happens on the cognitive level, but is also related to the metacognitive level of self-observation. Then the learners undergo a knowledge assessment regarding the recently used learning content and current learning goal. This phase is also on a cognitive level and related to the self-monitoring activity. Finally, learners should reflect on the activities and outcomes of the last phases. The current goal, the visited learning objects, and the assessment results are visually displayed. *Reflecting on learning behaviour and progress* targets the metacognitive activities of self-reflection and self-evaluation. A **learning cycle** is defined as iteration through these four phases. An overview of these phases and their relations to cognitive and metacognitive activities is shown in Table 2.

Table 2. Overview on the cognitive and metacognitive learning activities in the SRL process.

SRL phase and related tool	Metacognitive activity	Cognitive activity
Planning and goal setting	Planning	Understanding the subject domain
Using learning resources	Self-monitoring	Attaining domain knowledge
Knowledge and competence assessment	Self-monitoring	Knowledge assessment
Reflecting on learning behaviour and progress	Self-reflection, self-evaluation	Knowing the learning progress

A system that features these phases by offering appropriate tools can be regarded as an environment that enables self-regulated learning. A major goal of this approach is to provide ***personalized scaffolds*** that assist learners in a self-regulated manner. These scaffolds are based on the domain model and the user model. Since the user model contains information on the individual learning history, the support can be adapted to the individual learner and thus be personalized. Personalized scaffolds are given to the learner in each SRL phase differently depending on the phase and related tool. For example, in the goal-setting phase the learner gets a visual representation of the competences, including visual clues of the current competence state, so that the goal-setting activity is guided by the prerequisite structure of the competences and the current competence state. More details on personal guidance, the user interface, and its tools are provided in the User Interface section below.

Learning Analytics methods are employed by exploiting the user model data and presenting relevant information in a graphical way. This representation should stimulate the learner's reflection and motivation. Instead of operating with pure log data, high-level information (based on competences, learning objects, and assessment items) is presented to the learner. Since the user model holds the history data in time sequences, calculations must be performed to extract information as described above (reflection phase). An ***Open Learning Model*** approach is employed by displaying information in a graphical way that allows the learner to self-regulate his own path through the learning resources. ***Personalization*** is implemented by giving visual recommendations in terms of selecting next goals and choosing next learning objects and by offering assessment items depending on the previous learning behaviour.

Summarizing the described approach, the framework consists of a domain model and a user model that cover the subject domain and provide the basis for supporting learning. The self-regulated learning process model allows the learner to navigate freely through the learning resources, but provides self-regulated learning scaffolding to help the learner follow the phases and related tools. Recommendations, visualizations, and monitoring mechanisms are used to provide personalized guidance. The overall approach is outlined in Figure 2.

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

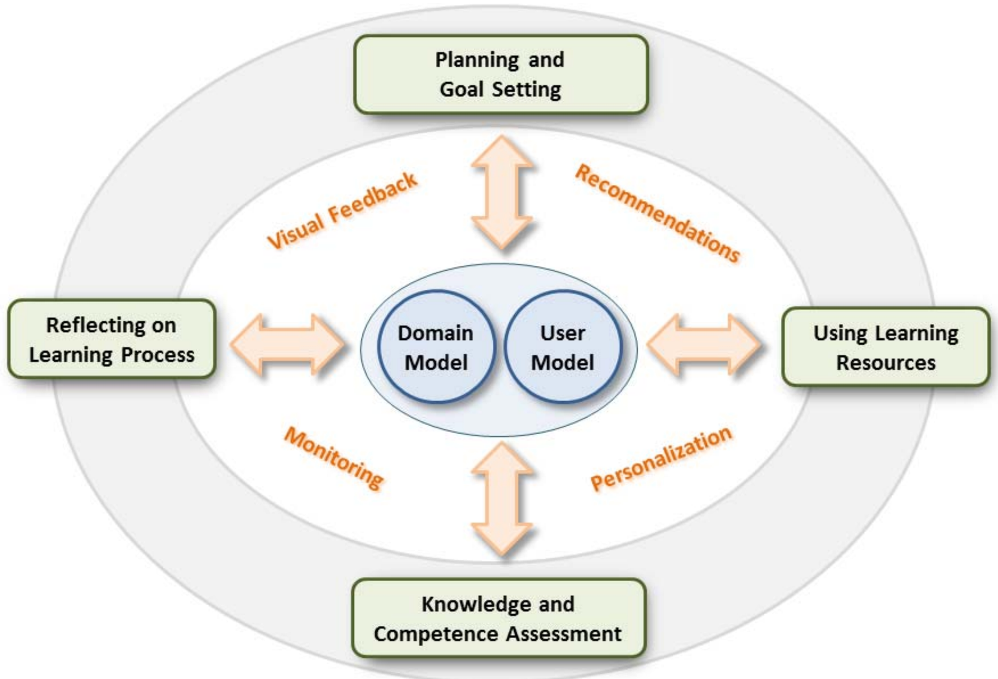


Figure 2. Overview of the psycho-pedagogical framework. The outer circle shows the metacognitive activities and the arrows indicate that these are supported in different ways based on user and domain models.

3.2 Technical Design and Compod Service

A core part of the technical design is the Compod service created to support SRL based on CbKST and learning analytics concepts as outlined above. This Web-based service uses a REST interface for functionality. It provides access to the user and domain model, as well as to the recommendation and assessment features. Furthermore, the service includes monitoring and tracking functionality to capture learner interactions, analyzing and reasoning features to model the learner state and behaviour, and access to this information. In order to offer this functionality to the learner, a technical component with a user interface is connected to the service (see the User Interface section below). The overall technical design is outlined in Figure 3.

The service holds and manages the domain models, which are represented in XML format and contain information regarding learning objects, assessment items, and competences, as well as the relations between these elements. Domain models can be added, removed, or edited. The REST interface exposes methods to retrieve whole domain models, but also offers reasoning on it. For example, learning objects related to a specific competence can be searched for and retrieved.

The user model component storage holds the learner information captured by the tracking and monitoring component, which captures the interaction data from the user interface and stores it in the user model as an extended semantic triplet. While the semantic triplet has the form <*subject, predicate,* *object>*, the extension adds contextual information about creation time and domain model. The subject is related to the learner who induced the information. The predicate describes the type of information; for example, *learning object visited*, *assessment item solved*, *assessment item failed*, *current competence goal*, or *current competence state*. The object contains the actual information according to the predicate type, which can be, for example, a learning object, an assessment item, or a competence. The domain model information is needed to specify the subject domain to which the triplet is related. For example, a student can learn two different subject domains at the same time, which has to be made distinguishable. This way of structuring user model information creates a flat structure that can be browsed and analyzed easily.

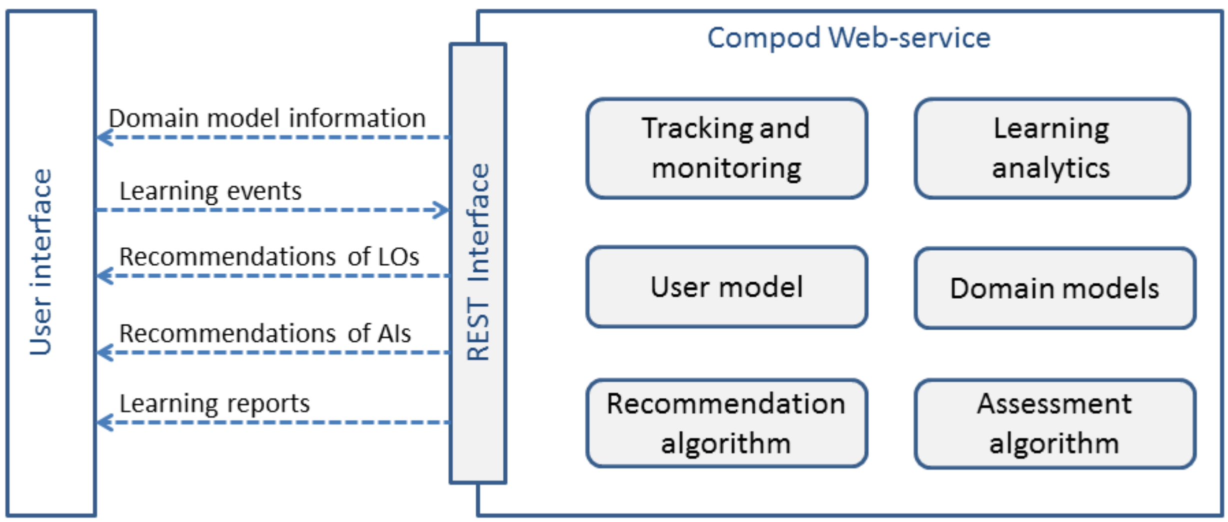


Figure 3. Overview of the technical architecture and main components.

The assessment component manages the knowledge and competence assessment. It recommends assessment items based on the current competence goal and previous answered assessment items. Furthermore, it processes the answers to the assessment items and calculates the current competence state. The recommendation component selects learning objects fitting the current competence goal and competence state. It also takes into account previously visited learning objects. These components form the basis for the personalization approach and personalized guidance.

The learning analytics component is responsible for analyzing the user model information and creating meaningful reports that help learners gain insight into their learning process and improve their learning. For example, it contains statistics on the visited learning objects and completed assessment items in relation to the competence goals. Furthermore, it includes information about how the competence state evolved over time throughout the learning process. This information is provided in XML format via the REST interface and is the basis for the visualizations in the user interface.

3.3 User Interface

This section describes a user interface designed to interact with the Compod service. This user interface consists of four tools representing the four phases of the SRL process model. Each tool supports the cognitive and metacognitive activities of the respective phase. Switching between these tools can be done by clicking on the tool name on top of the user interface. The tool names are the catchwords *Plan*, *Learn*, *Assess*, and *Reflect*. Following the sequence of the phases is suggested but not mandatory.

The first tool addresses the goal-setting activity (see Figure 4). This tool displays the competence map consisting of the competences of a domain model and the prerequisite relations between these competences. As overlay information, it also depicts the current competence state by drawing the contained competences as bigger green circles. Furthermore, it displays the current competence goal by drawing a red border around the circles. The user can add or remove competences by clicking on the respective ones. The prerequisite structure serves as guidance to navigate through the subject domain. The learner is free to choose any competences, but from a pedagogical point of view, it is meaningful to start with competences that have no other competence as prerequisite and then move up along the prerequisite relations. Thus, the prerequisite structure and the current competence state are scaffolds for the learner to choose goals and navigate through the subject domain.

The second tool is used to browse through the learning objects (Figure 5). All learning objects are listed on the left side, while the recommended learning objects are painted with a red border. Visited learning objects are marked with a blue line on the right side. Learners are free to choose, but it is recommended by visual scaffolds to follow the learning objects associated with the competence goal chosen by the learner.

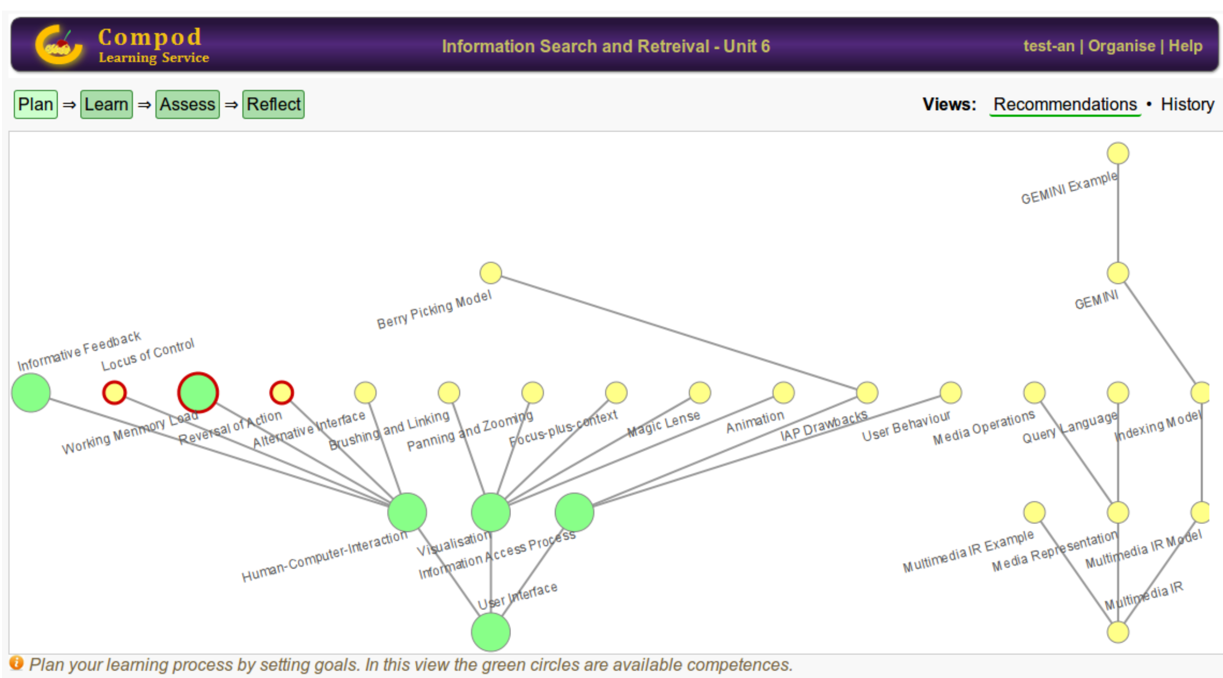


Figure 4. The goal-setting tool displays the competence map, the available competences (in green) and allows for selecting a competence goal (red border).

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

The assessment tool, the third tool, provides two options for assessment that can be freely chosen by the learner. The tool presents assessment items related either to the goal competences or to visited learning objects (Figure 5). Therefore, the learner can choose to use the goal-setting feature and do the assessment according to these goals or to omit the goal setting and do the assessment according to the visited learning objects.

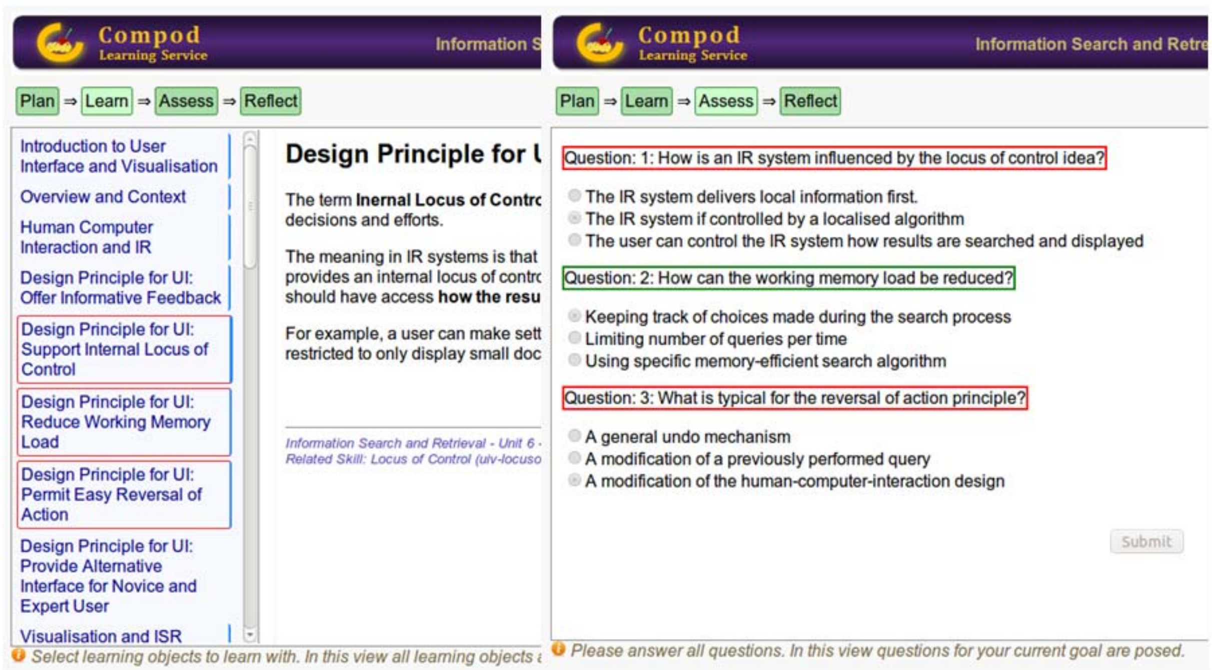


Figure 5. The learning and assessment tools are displayed in the screenshots.

The reflection tool presents the learning progress of the learner (Figure 6). It displays an overview of the learning progress over time, which is depicted as a green line representing the proportion of available competences after each learning cycle. The proportion is calculated as the number of available competences divided by the number of all competences multiplied by 100. Furthermore, it displays the proportion of correctly (green bars) and incorrectly (red bars) answered assessment questions. Absolute numbers are available to the learner through the tooltip. The tool can also display the number of visited learning objects in relation to the selected competence goal and assessment items, if the learner selects the "Learning Behaviour" view.

3.4 Case Study

From a self-regulated learning perspective, the proposed way of using this interface and thus the Compod system is to start with the goal-setting tool and then use the learning tool, the assessment tool, and the planning tool. This sequence should be reiterated until all competences have been acquired. For example, learner X adds three competences to her current learning goal. Then she navigates to the learning tool where she gets recommendations (highlighted with red borders) for learning objects that

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

teach these competences. The learner visits these learning objects and learns the content. After learning, she navigates to the assessment tool where she gets assessment items related to the three competences in her current goal. She answers these items and it turns out that one answer is wrong and the others are correct. This information is immediately shown by putting green borders on the questions with correct answers and a red border on the question with the wrong answer. The learner then navigates to the reflection tool where this first learning cycle (iteration) is shown. The average competence level (number of available competences divided by the number of all competences) is depicted as a progressing line. Further, the numbers of visited learning objects and results on the assessment items are shown for this first cycle. After completing this cycle, the learner navigates again to the planning phase. Now the competence map still shows the previous competence goal (highlighted with red circles), but also shows the available competences (green circles) because of mastering assessment items in the previous cycle. Therefore, the learner can modify her learning goal. For example, she will remove the available competences from the current goal, keeping the competence where she failed in the previous cycle and adding two new ones. The remainder of this cycle is processed as before.

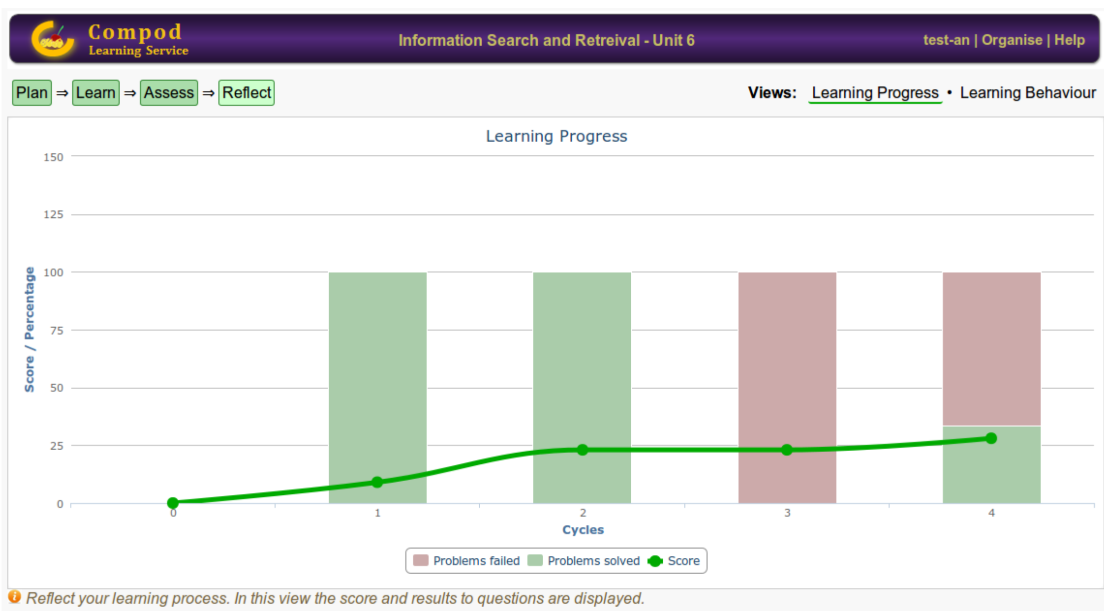


Figure 6. The reflection tool is displayed in the screenshot outlining the learning progress.

Though the case described above follows the self-regulated learning process model, learners can still choose to follow different learning paths and navigation behaviour. For example, learner Y prefers not to navigate through this whole cycle all the time so she navigates to the learning tools and starts learning with some of the learning objects. After a while, she navigates to the assessment tool and gets questions related to the visited learning objects. She completes the assessment with mostly correct answers and one wrong item. She then navigates back to the learning tool where she can see which learning objects teach the missing competences (this is a special view in the learning tool). The learner iterates this combination of the learning and assessment phase several times. From time to time, she also navigates to the goal-setting tool where she gets an overview of her current knowledge state (available competences).

Both cases describe a self-regulated learning behaviour, though these two learners pursue different strategies. Both are in control over their learning process and freely choose their learning behaviour. While learner X explicitly sets current goals and accepts respective personalized scaffolding strategies, learner Y omits the explicit goal setting and choses individual learning objects directly and creates an individual learning path through the learning objects. Learner X explicitly uses the reflection tool for feedback, but learner Y accepts feedback from the goal-setting tool where available competences are highlighted.

4 EVALUATION

In order to demonstrate the usefulness and applicability of the presented approach and service, the Compod system with its user interface was applied in the context of a university course held at Graz University of Technology. The main purpose of this evaluation was to answer the following questions. While the research questions outlined in the Summary and Resulting Research Questions section led to the solution approach and related system implementation described above, these evaluation questions intend to test these solutions:

- **E1:** How useable and acceptable is the overall Compod system and its user interface? This question addresses two aspects: usability and user acceptance. Usability refers to the issue of whether the system allows the user to achieve a specific goal effectively, efficiently, and satisfactorily. Concerning user acceptance, the main question of interest is whether users find the system acceptable and intend to use it. Research has shown that although a system is technically sound, users often do not intend to use the system because they lack a positive attitude towards system's usage (e.g., Davis, 1986; Hirschheim, 2007). Thus, answering this question is an indicator of whether learners will adopt the self-regulated way of learning in future.
- **E2:** Do learners accept and follow the learning approach provided by the Compod system? Do they adopt the self-regulated way of learning? This question refers to whether learners actually use the system and its different self-regulated learning functionalities.
- **E3:** Does use of the Compod system and its underlying learning approach benefit the acquisition of domain knowledge? This evaluation question addresses the learning effectiveness that can be achieved through self-regulated learning with the Compod system.
- **E4:** How do users feel about the visualizations (i.e., goal-setting tool and reflection tool) provided by the system? This question refers to users' perceived benefit of the visualizations provided by the system. This includes questions like whether the visualizations are understandable and suitable for their given tasks, or whether the visualization types help them to plan or reflect on their learning process in terms of learning effectiveness.

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

4.1 Method

This section presents the method and results of the evaluation study, consisting of two evaluation rounds conducted in 2013 and 2014. We employed a descriptive evaluation study design with a rather formative character aimed at improving the software and the learning approach. In the following sections, the methodology of the evaluation study is first described, and then the analysis of the data collected is given.

4.1.1 Setting

The study was conducted in the context of a university course on "Information Search and Retrieval (ISR)" held at the Graz University of Technology, Austria, in November 2013 and November 2014. This university course is a typical mandatory computer science course in the Master's Program, which guarantees that enough students are available and motivated to participate in research studies. The whole lecture consists of different topics, such as Web Retrieval, Query Languages, or Information Retrieval Models. For one topic, namely "User Interface and Visualization" about multimedia information retrieval and user interfaces in information retrieval, the Compod system has been used. In order to use the Compod system and its functionalities adequately, the domain model was created by the authors. The domain model consisted of 26 competences structured through prerequisite relations (see Figure 4). Furthermore, 32 learning objects and 26 assessment items related to the competences were created. The learning objects included the content conveyed in the lecture on this unit in previous years.

4.1.2 Participants

The study was carried out with two groups of Master's students at two points in time. In the first group (winter 2013), 28 students (6 female, 22 male) took part and filled in the evaluation survey. The sample consists of students of Computer Science (20 students), Software Development and Business Management (5 students), and Telematics (3 students). Students indicated that they had a medium pre-knowledge of the subject taught. In the second group (winter 2014), 22 students (4 female, 18 male) participated in the study and completed the evaluation survey. Participants were students of Computer Science (18 students), Software Development and Business Management (2 students), and Telematics (2 students). They indicated that they had an average knowledge in the field taught.

4.1.3 Material

In order to collect and analyze quantitative and qualitative data, a multi-method approach consisting of user model data sources and questionnaire data sources was applied in this study.

User Model Data: For investigating how and in which way students use and apply the system and its underlying learning approach, data on the learning and navigational behaviour of students was recorded as useful descriptive information. Learning behaviour data contains information on the visited learning objects, the responded assessment items, selected goals, and the acquired competences. Navigational data contains log data recording user interaction with the Compod system, or more concretely, which tools were used in which sequence and with what frequency.

Questionnaire: An online survey was created consisting of five short questionnaires allowing for capturing quantitative as well as qualitative feedback from students. The survey was realized and administered using an online survey system. The following main aspects were addressed by the online survey: usability, user acceptance, learning approach and guidance, usability and benefit of the goal-setting tool, and usability and benefit of the reflection tool.

For answering the first evaluation question referring to the general usability of the system (E1), the System Usability Scale (SUS; Brooke, 1996) covering 10 items was used. With respect to *user acceptance,* which refers to the first evaluation question (E1), a scale of three items covering the main aspects (i.e., perceived ease of use, perceived usefulness, and intention to use) according to the technology acceptance model (Davis, Bagozzi, & Warshaw 1989) was adapted.

The aspects *learning approach* and *guidance* were captured by three items each. These newly created items ask for general level feedback on the usefulness and applicability of the learning approach used by the system. To collect qualitative feedback on these aspects, this section was completed with one open question where participants had the opportunity to give additional feedback and comments. Gathering data on this aspect allows for answering the second evaluation question (E2).

In order to evaluate *visualization types* used for the *planning* (i.e., goal-setting tool) and *reflection* (i.e., reflection tool) phases and in order to answer the fourth evaluation question (E4), a questionnaire developed in order to evaluate visualizations in the context of digital libraries (Steiner et al., 2014) was adapted (see Table 3). Overall, the questionnaire consists of two scales: one (4 items) assessing *usability*, and the other (6 items) investigating the *perceived benefit*. Usability consists of the subscales *suitability for the task* and *self-descriptiveness*. The perceived benefit scale consists of the subscales *metacognition*, *cognitive load*, and *learning effectiveness*. At the end of the study, participants were asked to provide qualitative feedback on the strengths and weaknesses of the goal-setting tool and the reflection tool.

Table 3. Subscales on *usability* (i.e., suitability for the task and self-descriptiveness) and *perceived benefit* (i.e., metacognition, cognitive load, and learning effectiveness) to evaluate the goal-setting tool as well as the reflection tool.

Subscale	Items
Suitability for the Task	I find this visualization suitable for getting an overview of the current status in the learning process. I think the visualization provides irrelevant information.
Self-Descriptiveness	It is easy to understand this visualization. I find this visualization unnecessarily complex.
Metacognition	I think this visualization can help learners reflect on their learning process. I think this visualization supports learners in better planning their learning goals.
Cognitive Load	I think interpreting this visualization would put additional cognitive effort on the learner. I think this visualization is able to leverage the mental workload.
Learning Effectiveness	I think this visualization can help learners in accomplishing their goals. I think the use of this visualization will not make a difference for learning performance.

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

Each aspect (i.e., user acceptance, learning approach and guidance, goal-setting tool, and reflection tool) — except usability — covered by the survey was assessed with items or statements answered on a seven-point rating scale ranging from strongly disagree (1) to strongly agree (7). Negatively poled items have been recoded for further calculations. Then, for each aspect, a mean score averaging across the rating scale can be calculated with higher values indicating a better result. For assessing overall usability, we applied the SUS, including statements rated on a five-point scale ranging from strongly disagree (1) to strongly agree (5). The raw data generated from the survey was then computed to an overall SUS score ranging from 0 to 100, with higher values indicting a better result.

4.1.4 Procedure

Master's students were asked to use the system for one week, and were completely free to decide when and for how long to use the system for learning. However, it was mandatory for them to use and to answer enough assessment items correctly in order to collect the required points used for their marks. Of course, they could also use other resources for their learning in addition to the material provided by the Compod system. Previously, accounts for each student had been created separately and sent out to them. Furthermore, students got a general explanation of how to use the system and the evaluation study in a lecture before starting to work with the system. After one week, the system was closed so that the students could not use it anymore. After working with the system, the students were asked to fill in the online evaluation survey.

4.2 Results

4.2.1 Results of the learning progress and the tool usage

Since the learning behaviour was tracked by the Compod Service, user model data for each student is available. Though this data was used to support the students during their SRL processes, a post-analysis revealed interesting information about not only the learning effectiveness of the system, but also whether students followed the proposed learning approach. Overall, the data consisted of two types of information: learning progress and navigational behaviour.

Learning progress and knowledge level: The goal of the students was to achieve a score of at least 50 out of 100 points. In order to earn points, they had to answer assessment items correctly. Having available all competences leads to 100 points and a subset of the available competence leads to the respective proportion (score = (number of available competences / number of all competences) * 100). There was no time limit, so all students had the chance to use the system until they achieved as many points as they wanted.

In the first evaluation round (n=28), 27 students achieved 100 points; one only 62 points. In the second evaluation round (n=22), 20 students achieved 100 points, one achieved 92 points, and one dropped out after a few minutes of using the system. The positive learning result is an indicator that the students were highly engaged in using the system and did not get frustrated (except one or two). This result also positively answers the third evaluation question (E3) about whether the learning approach is suitable to learning effectively and to acquiring domain knowledge.

Navigational behaviour: The user interface consists of four main tools (i.e., goal-setting tool, learning tool, assessing tool, and reflection tool) related to different phases of self-regulated learning (see Figure 7). In principle, students could freely decide which tools they used, how often, and in which sequence. They had to complete all assessment items provided by the assessment tool and therefore had to select all competences at least once. Analysis of the navigational log data revealed that students had a strong tendency to follow the learning cycle "planning–learning–assessment–reflecting." A second smaller cycle "learning–assessing" could be observed, which is not surprising since it constitutes a more direct way of achieving the required learning score and is a common learning practice in the Austrian educational system. An overview of the navigational behaviour is depicted in Figure 7. This diagram shows the relative frequencies of the transitions from one phase to another. For example, the transitions from the planning phase to the learning phase are 0.85 (or 85%) of all transitions from the planning phase. The sum of the relative frequencies from one phase to the other is always 1 (or 100%). Additionally, the diagram shows how often the single tools were used by indicating the relative frequency of the overall number of visits to the tool. When looking at these relative frequencies in more detail, it becomes clear that the learning and assessment phases were visited more often than the planning and reflection phase. Data from both evaluation rounds are presented together and separately in brackets.

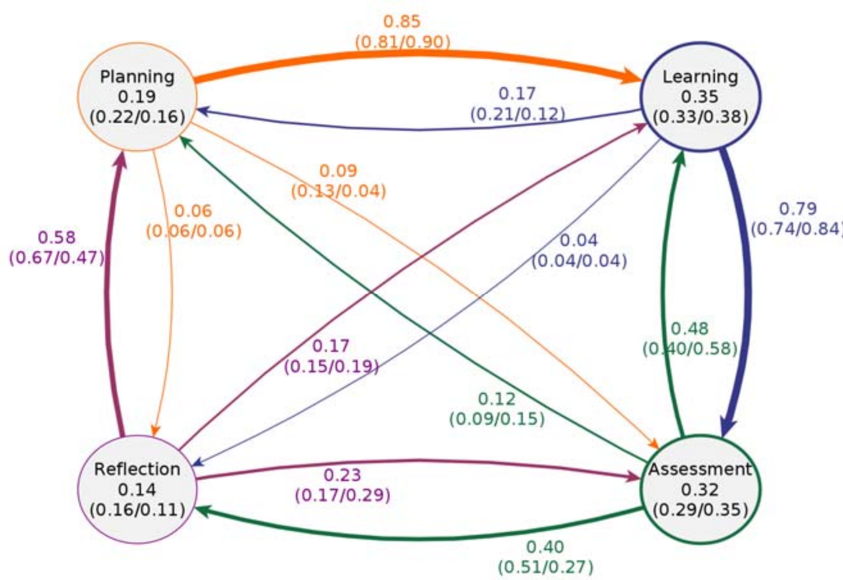


Figure 7. Overview of the navigational behaviour between the SRL tools. The arrows indicate navigation from one SRL tool to another and the values indicate the relative frequencies.

These results also answer the second evaluation question (E2): do learners follow the proposed learning approach and do they adopt the self-regulated way of learning? The navigational data shows that learners used all the tools, including those related to metacognition (planning and reflection), without being forced to and that they freely followed the learning cycle, at least to some extent. Main activities shown between the learning and assessment phase might be explained by the fact that students are familiar with this kind of learning, as it is common in the educational context. Interestingly, when additionally using the planning and/or reflecting functionality for their learning, students follow the SRL cycle, meaning that they first plan, then learn, then assess, and finally reflect on their learning before starting again with planning. This behaviour indicates that students adopted the self-regulated way of learning by following the proposed learning approach.

4.2.2 Results of questionnaires

Usability: The system's usability scored well, with an overall average score of *M* = 76.01 (*SD* = 12.85, *Mdn* = 75.00) on a scale ranging from 0 to 100, where higher values indicate a better result. In the first evaluation round, a mean usability score of 79.30 (*SD* = 12.63; *Mdn* = 80.00) resulted, which indicated a good to excellent usability of the Compod environment. In the second evaluation round, a slightly lower score resulted with 72.89 (*SD* = 12.59; *Mdn* = 72.50). Results are presented in Figure 8. Looking at individual items, participants assessed the learnability of the system's functionalities as appropriate, meaning that participants need no additional support in order to work effectively in and with this environment. The lowest result was for the potential future use of the system; however, the resulting score is slightly above the mid-point (*M* = 2.08; *SD* = 1.30; *Mdn* = 2.00), arguing for a satisfactory assessment. This result is also reflected in the qualitative feedback given by participants, where most students appreciated the system's usefulness for learning. Despite some minor spelling errors, they mostly found the system very usable.

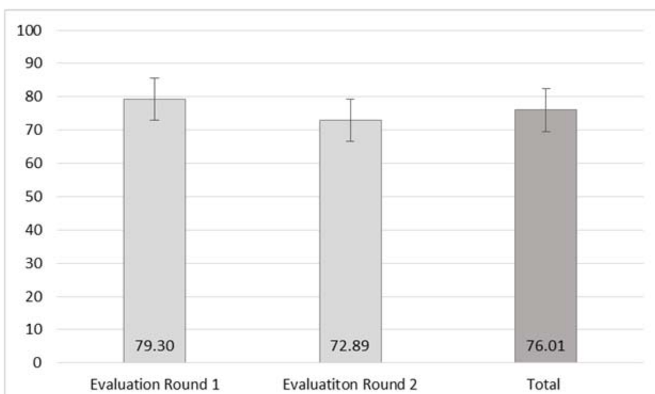


Figure 8. Overview of results (mean scores and SDs) on the aspect usability for both evaluation rounds and total usability score.

User Acceptance: For assessing user acceptance, ratings on the subscales *perceived ease of use*, *perceived usefulness*, and *intention to use* were collected. The results of the mean scores for each individual aspect, as well as the overall score for both evaluation phases, are depicted in Figure 9. The best result was for *perceived ease of use*, with *M* = 6.33 (*SD* = 0.84, *Mdn* = 7.00) in the first evaluation round and *M* = 5.58 (*SD* = 1.50; *Mdn* = 6.00) in the second. Students found the system generally easy to use. The overall score for ease of use, taking into account both evaluation rounds (*M* = 5.95; *SD* = 1.27; *Mdn* = 6.00), was strongly correlated with usability: $r = 0.65$ ($p = 0.00$). This means that students who gave the system high marks for usability also gave it high marks for ease of use. For *perceived usefulness* (*M* = 4.57; *SD* = 2.03; *Mdn* = 5.00) and *intention to use* (*M* = 4.35; *SD* = 2.09; *Mdn* = 4.50), slightly lower values could be identified. Perceived usefulness was rated at *M* = 4.89 (*SD* = 2.00; *Mdn* = 5.5) in the first evaluation round and *M* = 4.26 (*SD* = 2.08; *Mdn* = 5.00) in the second. Intention to use resulted in a mean score of 4.69 (*SD* = 2.09; *Mdn* = 5.25) in the first and 4.03 (*SD* = 2.10; *Mdn* = 4.00) in the second. Overall, this indicates a medium to good result and is completely in line with the qualitative feedback given by participants, who highlighted the support for learning that the system can provide.

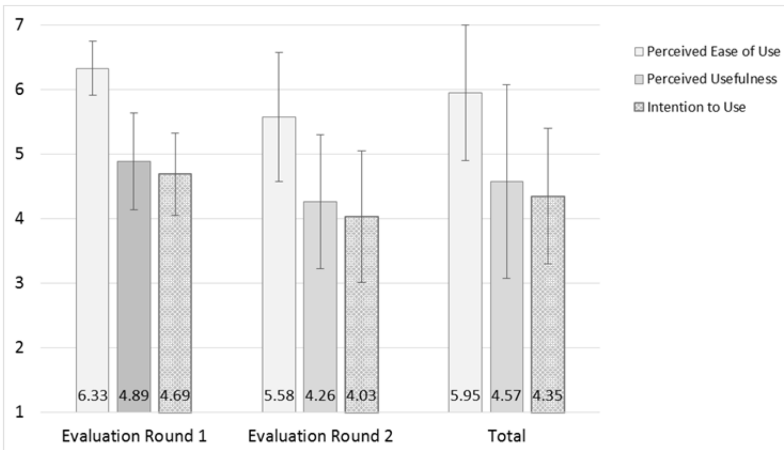


Figure 9. Overview of results (mean scores and *SD*s) on the user acceptance aspects: perceived ease of use, perceived usefulness, and intention to use for both evaluation rounds and in total.

Learning Approach and System Guidance: In the third questionnaire, students were asked to assess the overall learning approach and the guidance facilities provided by the system. Overall, the learning approach was rated as moderately good, with a mean score of 5.35 (*SD* = 1.30; *Mdn* = 5.67) on a scale ranging from 1 to 7 where higher values meant a better result. In the single evaluation rounds, a mean score of 5.65 (*SD* = 1.19; *Mdn* = 5.67) was obtained in the first round and 5.03 (*SD* = 1.35; *Mdn* = 5.33) in the second. Looking at the items for both evaluation phases, it became obvious that students found the learning approach not only supportive for learning generally, but also provided them with a better learning experience than other learning systems. The detailed results of the individual items are displayed in Figure 10. These generally positive results were also confirmed by the students when explicitly asked about the strengths and weaknesses of the system and its underlying approach. Most of them liked the cyclical learning approach, especially the feedback function because it provided a good overview of which questions had not been answered correctly. However, they also remarked critically that it was not clear at the beginning how to work through these different learning cycles (i.e., planning, learning, assessing, and reflecting) and how to use the different functionalities provided by the system, especially the visualizations (i.e., clicking on bubbles and marking them with different colours). With this in mind, a small tutorial would facilitate dealing with the environment.

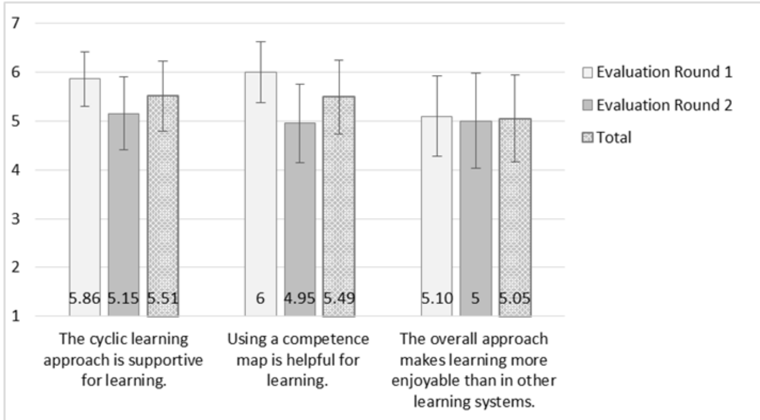


Figure 10. Results on items assessing the learning approach used by the system.

Similar results were obtained when assessing the guidance functionality provided by the system. In the first evaluation, this aspect was rated at *M* = 5.56 (*SD* = 1.12; *Mdn* = 5.67) and the second at M = 5.25 (*SD* = 1.57; *Mdn* = 5.67). This resulted in an overall score of *M* = 5.41 (*SD* = 1.35; *Mdn* = 5.67) as depicted in Figure 11. These positive results indicate that participants see the additional support this guidance functionality provides to their learning. They also highlighted the freedom to plan and organize learning on their own. However, they also pointed out that a primer on how to use and apply the learning approach and guidance functionality would be helpful.

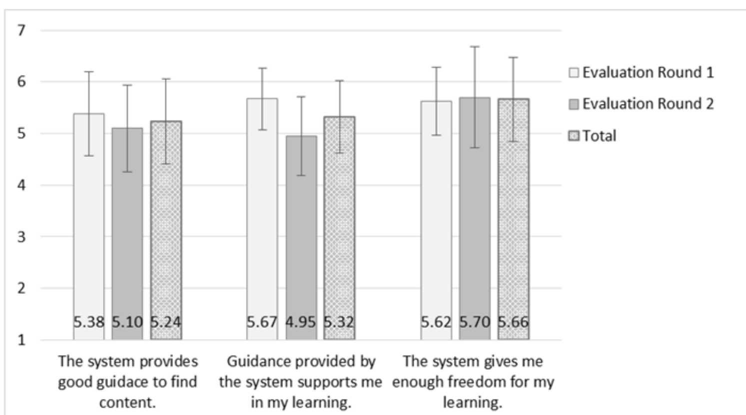


Figure 11. Results on items assessing the guidance functionality provided by the system.

Visualizations: Goal-Setting Tool: Concerning the goal-setting tool, good to medium results were obtained for both *usability* and *benefit*. For *usability*, a mean value of *M* = 5.35 (*SD* = 0.84; *Mdn* = 5.60) in the first evaluation round and *M* = 4.79 (*SD* = 1.12; *Mdn* = 4.80) in the second were found. *Benefit* was rated slightly lower with *M* = 4.36 (*SD* = 1.12; *Mdn* = 4.29) in the first evaluation and M = 4.43 (SD = 0.88; *Mdn* = 4.29) in the second. When looking at the single subscales, in the first evaluation round scores ranged from 4.00 (*SD* = 1.08; *Mdn* = 4.00) in the cognitive load scale to 5.47 (*SD* = 1.05; *Mdn* = 5.50) in the self-descriptiveness scale. In the second evaluation round, similar results were obtained, with mean scores ranging from 3.79 (*SD* = 0.90; *Mdn* = 3.50) in the cognitive load scale to 4.79 (*SD* = 1.17; *Mdn*

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

=4.50) in the suitability scale. Table 4 shows all results in detail.

Table 4. Results of the subscales, total usability, and total benefit for the goal-setting tool.

	Evaluation Round 1	**Evaluation Round 2**	**Total**
	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)
Suitability for the task	5.03 (1.18)	4.79 (1.17)	4.91 (1.17)
Self-descriptiveness	5.47 (1.05)	4.63 (1.34)	5.05 (1.26)
Total Usability	**5.35 (0.84)**	**4.79 (1.12)**	**5.07 (1.02)**
Metacognition	4.61 (1.42)	4.77 (1.08)	4.69 (1.12)
Cognitive Load	4.00 (1.08)	3.79 (0.90)	3.89 (0.99)
Learning-effectiveness	4.26 (1.42)	4.21 (1.22)	4.24 (1.30)
Total Benefit	**4.36 (1.12)**	**4.43 (0.88)**	**4.39 (0.99)**

Reflection tool: Similarly to the goal-setting tool, this type of visualization also received good results in all aspects (see Table 5). In the first evaluation round, the scores ranged from 4.03 (*SD* = 0.87; *Mdn* = 5.50) for cognitive load to 5.41 (*SD* = 1.19; *Mdn* = 5.00) for self-descriptiveness, resulting in an overall usability score of 5.25 (*SD* = 1.22; *Mdn* = 5.50) and an overall benefit score of 4.19 (*SD* = 1.21; *Mdn*= 4.57). In the second evaluation round, similar results were identified: scores ranged from 3.86 (*SD* = 0.92; *Mdn* = 3.50) for cognitive load to 4.92 (*SD* = 1.48; *Mdn* = 4.50) for suitability for the task. The overall usability score was *M* = 4.82 (*SD* = 1.33; *Mdn* = 4.50) and the benefit score was *M* = 4.51 (*SD* = 1.05; *Mdn* = 4.43), also satisfactorily good results. These results indicate that students find the tool suitable for reflecting on their learning and see the benefit it can provide.

Table 5: Results of the subscales, total usability, and total benefit for the reflection tool.

	Evaluation Round 1	**Evaluation Round 2**	**Total**
	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)
Suitability for the task	5.09 (1.51)	4.92 (1.48)	5.00 (1.48)
Self-descriptiveness	5.41 (1.19)	4.72 (1.34)	5.06 (1.30)
Total Usability	**5.25 (1.22)**	**4.82 (1.33)**	**5.03 (1.28)**
Metacognition	4.41 (1.50)	4.80 (1.35)	4.61 (1.42)
Cognitive Load	4.03 (0.87)	3.86 (0.92)	3.94 (0.89)
Learning-effectiveness	4.29 (1.51)	4.39 (1.40)	4.34 (1.43)
Total Benefit	**4.19 (1.21)**	**4.51 (1.05)**	**4.36 (1.13)**

Summing up the results obtained for both types of visualizations used by the Compod system, generally students agreed that visualizations provide useful and relevant information in terms of better planning and reflecting on their learning. These results were confirmed in the qualitative feedback. Most participants found that the goal-setting visualization provides useful information and is thus suitable for planning future learning activities, especially the opportunity to see the dependencies between topics (i.e., main topic and sub-topic). Additionally, students appreciated being able to organize their own learning process by simply choosing and clicking on their own learning goals. On the other hand, however, they pointed out the complexity of the visualization type (i.e., the Hasse Diagram) and consequently the need for a short tutorial (e.g., video or animation) explaining the functions (e.g., choosing a bubble, which consequently changes his colour). Regarding the reflection tool, participants found it clear, easy to understand, and that it provided good and useful information.

On a nominal level, the Compod system and services were assessed slightly higher in most qualities in the first evaluation study than in the second. The two exceptions were the goal-setting tool and the reflection tool, which were rated lower in the first study. However, statistical comparisons within the sample (t-tests for independent samples or, respectively, non-parametric tests where necessary) between the evaluations of the two student cohorts yielded no significant differences.

5 DISCUSSION, LIMITATIONS, AND OPPORTUNITIES

5.1 Discussion

Referring to our original research questions, the main aim was to create a framework composed of different learning methodologies (R2) and system implementing this framework (R3) that supports learners to follow the self-regulated learning paradigm and to acquire knowledge of the subject domain (R1). The proposed solution to these research questions was described in our psycho-pedagogical framework that incorporates Competence-based Knowledge Space Theory, personalization approaches, the Open Learner Model, and learning analytics methods, in order to provide personalized guidance in a self-regulated learning process. In order to validate the proposed solutions, a study, guided by four evaluation questions was undertaken to give insight into the extent to which this framework and system answer the research questions. In this evaluation, a descriptive study design using a multi-method approach for gathering data was used in order to gain in-depth data on the usefulness and applicability of the Compod system and its different functionalities. To ensure the same learning conditions for each learner in the context of a university course where students finished with a certificate, no control group was established. Because of the chosen design, the study also did not include a pre-test–post-test condition. This might be problematic, especially when making conclusions about the learning effectiveness of the system. However, the focus of this formative evaluation, as well as discerning what was good and useful to students, was on identifying any issues or potential problems with the technology and the underlying learning approach from a user-centred perspective. Such a procedure should ensure that valuable information for further improvement of the software is provided to the developer.

The first evaluation question (E1) addresses the usability and user acceptance of the overall system and approach. The results from questionnaires were rather good on both user acceptance and usability. This indicates that, in general, the system and its approach are acceptable. The second evaluation question (E2) specifically targets the learning approach and thus the framework. The results from these questionnaires indicated that students like this way of learning. The third evaluation question (E3) asked if learners actually acquire domain knowledge with this learning approach. Results retrieved from the user-model data show that learners achieved the goal of mastering the subject domain. Almost all students achieved the maximum score even though this was not required. Finally, the forth evaluation

(2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, 2(1), 101–133.

question (E4) specifically targets the goal setting and reflection tools and their visualizations. Results indicate that usability and the benefit for the learning process is above average, and students explicitly pointed out the positive effects in the qualitative feedback.

Overall, the results of the study are quite promising, indicating that the elaborated framework and developed system fulfil their purpose. Individual results on the evaluation questions point out that students accepted the system with its functions, user interface, and pedagogical framework. Additionally, the learning results in terms of domain knowledge were successful. If the system (R3) and the framework (R2) are suitable for self-regulated learning and achieving the learning goals (R1), then the Compod system passed this test.

Another discussion point addresses the system's relation to similar systems in the educational area, which already has a long tradition. The Compod system and approach follow the tradition of adaptive systems. It uses the same types of models (learner, domain, tutoring, and user interface model) and follows the Open Learner Model (OLM) approach by making these models visually accessible to the learner. Some similar systems deal with concepts, concept hierarchies, and concept-learning object relations. For example, both Interbook and AHA! present learning objects and concepts and give the user some freedom to navigate through the concept space (De Bra, Santic, & Brusilovsky, 2003). However, Compod has at least two distinguishing aspects. First, Compod is built upon CbKST, which not only includes concepts and learning objects, but also assessment items and probabilistic methods for adaptive testing and learner model updates, and brings them together in a holistic theory. Though not all features of CbKST are now used in Compod, there is still an opportunity to integrate them into the system. Second, Compod uses explicit support of self-regulated learning in terms of a phased approach that stimulates metacognitive activities (e.g., goal setting, planning, reflecting, and self-evaluation). A prominent example for an adaptive system using Knowledge Space Theory is the ALEKS system (Falmagne, Cosyn, Doignon, & Thiéry 2006). ALEKS uses an adaptation strategy based on Knowledge Space Theory (without the competence-based extension) to perform efficient, adaptive knowledge tests and personalized learning paths. However, the adaptation strategy and user models are not open to the learners.

5.2 Limitations

Though the overall feedback from the user study was rather positive, there are also some limitations to this approach. Self-regulated learning is a complex process and it is very unlikely that students can increase their SRL capabilities in a course unit during one week. Though this was not the aim of the study, it is still worth a closer look. Despite the fact that self-regulated learning is a complex process, it can also be seen as an integrated process involving the active construction of new behaviour affecting learning. In this way, self-regulated learning can be seen as a constructive learning process. Research has shown that the active application and development of self-regulatory processes is essential for enhancing self-regulatory skills (e.g., Pintrich & Zusho, 2002). Such a constructive view of learning understands the learner as being actively, mindfully, and effortfully involved in the learning process, or more concretely, in the process of knowledge and skill acquisition. With this constructive learning theory in mind, it becomes clear that even if students did not significantly enhance their self-regulated learning skills within the week of the study, applying them was still valuable training. However, in this study, the research focus was not on investigating whether self-regulated learning skills are enhanced by the Compod system — the research focus was more on investigating whether the learning approach itself applied. For future work, having a concrete focus on the development of students' SRL skills and their development by utilizing a standardized strategy to evaluate an increase of these competences would be useful (e.g., Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; Pintrich et. al, 1987).

Another problem with the evaluation was that a few learners mastered the unit for the evaluation not alone but in a small group. By analyzing the time stamps of the navigation events, it seems that some learners did the learning process together. However, doing assessment items together and probably not learning the content as suggested brings a slight distortion of the results. Since this is an online course, it is hard to prevent such effects. On the other hand, this also indicates that at least some students prefer to learn in groups, which reveals a general shortcoming in the presented approach, which is the missing aspect of collaborative learning. As known from the literature on SRL, collaborative learning provides opportunities for metacognition, motivation, and self-regulated learning in general so there is much room for extending the presented approach with collaborative aspects. For example, groups could be formed automatically with the same competences or peers could be recommended who have already mastered competences with which other students have problems.

Finally, using complex visualizations as a basis for planning, goal setting, and reflecting can cause problems if a learner does not understand such visualizations. In our user study this problem did not occur but this is not surprising since computer science students understand a graph-based view. However, other learners might have difficulties understanding graph visualization, which would lead to completely different results. One solution would be a short training session for using these visualizations.

5.3 Opportunities

Self-regulated learning is a complex psychological and pedagogical matter and appropriate technological support is a huge challenge. The presented Compod system supports self-regulated learning to some extent, but still leaves room for much improvement and further in-depth research. This section discusses some of the further research and development opportunities.

A key aspect of self-regulated learning is related to the cognitive and metacognitive activities and the sequences in which a learner performs them. When using technology for learning, these activities are manifest in interactions with the user interface. The sequences of these interactions can be subsumed as navigation behaviour and thus become an interesting and valuable resource for further research. Data collected in the user model contains information on the selected goals in terms of competences, the visited learning objects, the answered assessment items, and the selected tools. This information brings further research opportunities by analyzing in depth the navigation behaviour through these elements; for example, a detailed analysis of the extent to which a learner has followed the prerequisite structure provided in the competence map. This would provide insight, if this type of scaffold were accepted. Relatedly, analyzing how many competences per goal and how much time per goal (visited learning objects, assessment items) would facilitate investigating any correlations between the navigation behaviour and the learning progress. An analysis of the navigation behaviour through the tools has already been made (see Figure 7). However, there are more possibilities to relate the tool navigation behaviour with the goals, learning objects, and assessment items; for example, if there is any correlation between the navigation patterns and the learning outcomes.

These further analyses could in turn be used for live feedback on the reflection tool (Figure 6). Currently the reflection tool just shows information on the learning progress in terms of mastered assessment items and acquired skills. However, information on the navigation patterns could be used as a basis for further feedback. For example, the degree to which extent the learner follows the prerequisite structure of the competences could be visually displayed. We argue that it is pedagogically meaningful to follow this structure, as this feedback would reinforce the adoption of respective learning paths. In addition, whether the navigation behaviour correlates with the learning progress could be interesting for learners. This would require further evaluation and analysis to see if it affects learning.

The domain model (competences, learning objects, assessment items) used in the evaluation was rather small; larger domain models with more learning objects and assessment items would be beneficial for the learner. In the current domain model, only one learning object and one assessment item are available for each competence. However, the Compod approach allows for defining many more learning objects and assessment items for each competence. Learners might benefit from multiple perspectives, explanations, and examples to train a single competence. Also testing competences would benefit from multiple assessment items, which can foster the accuracy of the result.

The evaluation itself mainly addressed the question of whether the Compod system and its pedagogical approach would be accepted by learners; however, many more aspects can be evaluated in further studies. While the present evaluation included only one lecture and the average use of the system was a few hours, a longer study over the whole semester of following an experimental design could reveal more insight into self-regulated learning behaviour and its effect on learning outcomes. The whole course studied consists of about ten lectures and each of the lectures could be adapted for the Compod system. A test on self-regulated learning skills could be made at the beginning and end of the semester, which would show whether there were improvements over this time. Furthermore, an analysis could be made of whether learners change their learning behaviour over time in terms of their navigation behaviour. Using pre- and post-knowledge tests could also reveal if the frequent use of Compod and self-regulated learning approaches has a positive effect on learning outcomes.

Finally, due to the modular technical approach (separation of front-end and back-end), there is the potential to try out different user interfaces and thus modified SRL approaches. This might lead to designs with different levels of guidance, different amounts of feedback and recommendations, and different types of personalization. There is already an alternate user interface design available that reduces the amount of freedom in the goal-setting phase (Kopeinik et al., 2014). This modification was designed for vocational training with workers who are not used to setting their own learning goals.

6 CONCLUSION AND OUTLOOK

This paper has presented our framework for self-regulated learning. We identified a self-regulated learning process model and developed scaffolding techniques on a conceptual and technical level to support cognitive and metacognitive activities. We evaluated the learning approach and the technology in a descriptive evaluation study with computer science students. Students reacted positively and tended to accept both the learning approach and the system.

In the authors' opinion, the most innovative aspect is the combination of different learning concepts, which leads to a new framework that addresses learning support on both cognitive and metacognitive levels. While navigating the learning resources and attaining respective domain knowledge is related to the cognitive level, control of the overall learning process, including goal setting and reflecting, is related to the metacognitive level. The framework has means to capture both cognitive and metacognitive learning activities and uses them for personalized support, including goal setting, navigation, and reflection, on both the cognitive and metacognitive levels. In this way, broader, more holistic support for self-regulated learning is provided. To the best of our knowledge, a system that combines self-regulated learning with its metacognitive aspects and learning on the domain level is not available so far.

This paper contributes to the learning analytics field, as the presented framework and learning support strategies are based on traced and monitored learner data. Captured learner data include selected learning goals, visited learning objects, solved assessment items, or the navigation behaviour through the learning cycle. This information is used to provide recommendations and visual feedback following the idea of OLMs, in order to support self-regulated learning. In this way, learning analytics techniques together with OLM are used to assist a learner pedagogically in controlling his own learning process.

Future work includes more and deeper analysis of the learner interaction data in order to use it for further guidance mechanisms; for example, the navigation behaviour through the competence structure and related learning objects could be further analyzed. Having the nature of prerequisite relations in mind, the meaningfulness of the learners' navigation behaviour can be seen on different levels. As well, the reflection tool can benefit from further analysis in terms of learning behaviour. For example, analysis of the navigation behaviour presented in Figure 7 could be made visible in the reflection tool. This would feedback on one's own learning process to exploit learning analytics methods further while using the system. Another type of improvement includes collaborative support; for example, information can be provided about peer learners, so that learners can compare themselves with others regarding learning progress or navigation behaviour.

In future research, a specific focus could be on further investigating whether the Compod system with its proposed learning approach has a positive effect on learning in general, and on the enhancement of competences in self-regulated learning in particular. Thus we are planning to alter our experimental design to including a pre-test–post-test and a control group in order to investigate the cognitive processes involved in the learning experience in more detail.

REFERENCES

Anderson, J. R. (1988). The expert module. In M. C. Polson & J. J. Richardson (Eds.), *Foundations of intelligent tutoring systems* (pp. 21–53). Mahwah, NJ: Lawrence Erlbaum.

Albert, D., & Lukas, J. (Eds.) (1999). Knowledge spaces: Theories, empirical research, and applications. Mahwah, NJ: Lawrence Erlbaum.

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research, 35*(4), 359–375.

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research, 31*(6), 445–457.

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London: Taylor & Francis.

Brusilovsky, P., Kobsa, A., & Nejdl, W. (2007). The adaptive web. *Lecture Notes in Computer Science 4321*. Berlin/Heidelberg: Springer.

Buckingham Shum, S., & Crick, R. D. (2012). Learning dispositions and transferable competencies: pedagogy, Modelling and Learning Analytics. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 92–101). doi: 10.1145/2330601.2330629

Bull, S., & Kay, J. (2010). Open learner models. *Advances in Intelligent Tutoring Systems, 207*, 301–322. Berlin/Heidelberg: Springer. doi: 10.1007/978-3-642-14363-2_15

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology, 51*, 171–200. doi: 10.1146/annurev.psych.51.1.171

Dabbagh, N., & Kitsantas, A. (2004). Supporting self-regulation in student-centered web-based learning environments. *International Journal on e-Learning, 3*(1), 40–47.

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Doctoral dissertation, Sloan School of Management, Massachusetts Institute of Technology.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003.

De Bra, P., Santic, T., & Brusilovsky, P. (2003, November). AHA! meets Interbook, and more... *Proceedings of the AACE ELearn 2003 Conference* (pp. 57–64), Phoenix, Arizona.

Duval, E. (2011). Attention please!: Learning analytics for visualization and recommendation. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 9–17). doi: 10.1145/2090116.2090118

Efklides, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema, 21*(1), 76–

82.

Falmagne, J.-C., Cosyn, E., Doignon, J.-P., & Thiéry, N. (2006). The assessment of knowledge, in theory and in practice. In R. Missaoui & J. Schmidt (Eds.), *Lecture Notes in Computer Science, 3874* (pp. 61–79). Berlin/Heidelberg: Springer. doi:10.1007/11671404_4

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231–325). Hillsdale, NJ: Erlbaum.

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning, 2*(2–3), 107–124.

Hattie, J. A. C. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. London/New York: Routledge.

Heller, J., Steiner, C., Hockemeyer, C., & Albert, D. (2006). Competence-based knowledge structures for personalised learning. *International Journal on E-Learning, 5*(1), 75–88.

Henri, F., Charlier, B., & Limpens, F. (2008). Understanding PLE as an essential component of the learning process. *Proceedings of ED-Media 2008 Conference* (pp. 3766– 3770). Vienna, Austria.

Hirschheim, R. (2007). Introduction to the special issue on “Quo Vadis TAM: Issues and Reflections on Technology Acceptance Research.” *Journal of the Association for Information Systems, 8*, 203–205

Hockemeyer, C. (2003). Competence based adaptive e-learning in dynamic domains. In F. W. Hesse & Y. Tamura (Eds.), *The Joint Workshop of Cognition and Learning through Media-Communication for Advanced E-Learning* (pp. 79–82).

Höver, K. M., & Steiner, C. M. (2009). Adaptive learning environments: A requirements analysis in business settings. *International Journal of Advanced Corporate Learning, 2*(3), 27–33. doi: 10.3991/ijac.v2i3.956

Issing, L. J. (2002). Instruktions-design für multimedia. In J. Issing & P. Klimsa (Eds.), *Informationen und Lernen mit Multimedia und Internet. Lehrbuch für Studium und Praxis*, 3rd ed., Beltz, Weinheim, 151–178.

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2014). NMC horizon report 2014: Higher education edition. The New Media Consortium. Retrieved from http://www.nmc.org/publications/2014-horizon-report-higher-ed.

Kitsantas, A. (2002). Test preparation and performance: A self-regulatory analysis. *The Journal of Experimental Education, 70*(2), 101–113.

Kopeinik, S., Nussbaumer, A., Winter, L.-C., Dimanche, A., Albert, D., & Roche, T. (2014). Combining self-regulation and competence-based guidance to personalise the learning experience in Moodle. *The 14th IEEE International Conference on Advanced Learning Technologies* (ICALT 2014). doi: 10.1109/ICALT.2014.28

Korossy, K. (1997). Extending the theory of knowledge spaces: A competence-performance approach. *Zeitschrift für Psychologie, 205*, 53–82.

Law, E. L. C., Schmitz, H. C., Wolpers, M., Klamma, R., Berthold, M., & Albert, D. (2012). Responsive and open learning environments (ROLE): Requirements, evaluation and reflection. *Interaction Design and Architecture(s) Journal* (IxD&A), *15*, 87–101. Retrieved from:

http://www.mifav.uniroma2.it/inevent/events/idea2010/doc/15_7.pdf

Mikroyannidis, A., Connolly, T., Law, E., Schmitz, H.-C., Vieritz, H., Nussbaumer, A., Berthold, M., Ullrich, C., & Dhir, A. (2014). Self-regulated learning in formal education: Perceptions, challenges and opportunities. *International Journal of Technology Enhanced Learning, 6*(2), 145–163. Inderscience.

Paulsen, M. (2003). Experiences with learning management systems in 113 European institutions. *Educational Technology & Society, 6*(4), 134–148

Pintrich, P. R. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40. doi: 10.1037/0022-0663.82.1.33

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*, 459–470.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33–40.

Pintrich, P. R., McKeachie, W. J., Smith, D. A., Doljanac, R., Lin, Y. G., Naveh-Benjamin, M., et al. (1987). *The Motivated Strategies for Learning Questi*onnaire (MSLQ). Ann Arbor: NCRIPTAL, University of Michigan.

Pintrich, P. R., & Zusho, A. (2002). The development of academic self-regulation: The role of cognitive and motivational factors. In A. Wigfield & J. Eccles (Eds.), *Development of Achievement Motivation* (pp. 249–284). San Diego, CA: Academic Press.

Roberts, M. J., & Erdos, G. (1993). Strategy selection and metacognition. *Educational Psychology, 13*(3–4), 259–266.

Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. doi: 10.1006/ceps.1999.1020

Siemens, G. (2010). What are learning analytics? Retrieved from http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

Steiner, C., Nussbaumer, A., & Albert, D. (2009). Supporting self-regulated personalised learning through competence-based knowledge space theory. *Policy Futures in Education, 7*(6), 645–661.

Steiner, C. M., Agosti, M., Sweetnam, M. S., Hillemann, E.-C., Orio, N., Ponchia, C., Hampson, C., Munnelly, G., Nussbaumer, A., Albert, D., & Conlan, O. (2014). Evaluating a digital humanities research environment: The CULTURA approach. *International Journal on Digital Libraries, 15*(1), 53–70. Berlin/Heidelberg: Springer. doi: 10.1007/s00799-014-0127-x.

Treier, M. (2004). Personale voraussetzungen für das lernen mit neuen medien. Evaluation und gestaltung im zusammenhand mit der implementierung einer bildungsplattform in einem konzern. Hamburg: Dr. Kovac.

Tseng, J., Chu, H., Hwang, G., & Tsai, C. (2008). Development of an adaptive learning system with two sources of personalization information. *Computers & Education, 51*(2), 776–786. doi: 10.1016/j.compedu.2007.08.002

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500–1509. doi: 10.1177/0002764213479363.

Winne, P., & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Mahwah, NJ: Lawrence Erlbaum.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. doi: 10.1207/s15430421tip4102_2

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166–183.

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting

Kay Colthorpe, Kirsten Zimbardi, Louise Ainscough, and Stephen Anderson
School of Biomedical Sciences, The University of Queensland, Australia
k.colthorpe@uq.edu.au

ABSTRACT: It is well established that a student's capacity to regulate his or her own learning is a key determinant of academic success, suggesting that interventions targeting improvements in self-regulation will have a positive impact on academic performance. However, to evaluate the success of such interventions, the self-regulatory characteristics of students need to be established. This paper examines the self-regulatory characteristics of a cohort of second-year allied health students, using the evaluation of responses to "meta-learning" assessment tasks supported by access data from the learning management system. Students primarily report using learning strategies from the performance and self-reflection phases. Although few reported using forethought strategies, access to preparatory course materials suggests that these were under-reported. Students who reported reviewing lectures as a learning strategy were more likely to access the online lecture recordings; however, higher access was associated with poorer academic performance. Cluster analysis of all available data showed high academic performance was positively associated with early submission of intra-semester assessment tasks but negatively associated with both use of, and reported of use of lecture recordings by students. These findings suggest that early submission of intra-semester assessment may be useful as a predictor of academic achievement.

KEYWORDS: meta-learning, self-regulation, allied health

1 INTRODUCTION

As knowledge increases with often overwhelming complexity, the development of lifelong learning skills is an imperative for graduates to excel in a global society. These skills are particularly important in professions such as science and the allied health sector, where the pace of new knowledge generation is rapidly accelerating. While it is well established that a student's capacity to self-regulate learning is a key determinant of academic success and the ability to overcome academic adversity (Turner & Husman, 2008; Zimmerman, 2002), the development of self-regulatory skills is also considered critical to the development of lifelong learning (Schunk, 2005).

Self-regulated behaviour can be defined as the thoughts, feelings, and actions planned and adapted by an individual in order to attain a self-selected goal (Zimmerman, 2000). Self-regulated learning involves more than just detailed knowledge; it encompasses the self-awareness, motivation, and behavioural adjustments made in order to implement knowledge. There are various theories of self-regulated learning with differing levels of complexity (Winne, 1996; Boekaerts, 1997; Zimmerman, 2000; Bannert, Reimann, & Sonnenberg, 2013). Most theories agree that self-regulated learning is adaptive and is therefore cyclical in nature. Learners modify their learning strategies to suit the task at hand and through critical appraisal of past learning events. The current study utilizes Zimmerman's three-phase cyclical model of self-regulation (Zimmerman, 2000; Nota, Soresi, & Zimmerman, 2004). This well-established model proposes that self-regulation occurs through three strategic phases: forethought, performance, and self-reflection (Zimmerman, 2000).

The forethought phase involves processes, beliefs, and thoughts that occur prior to learning. Processes in this phase include task analysis, involving goal setting and strategic planning, and self-motivation (Zimmerman, 2002; Cleary & Zimmerman, 2004; Schunk, 2005). The performance phase involves behaviours implemented in response to the learning process (Postholm, 2011), with the main processes being self-control and self-observation (Zimmerman, 2000). Finally, the self-reflection phase occurs after learning has taken place. It involves self-reaction and self-judgement processes, with self-evaluation and causal attributions being part of the latter process (Dunlosky & Rawson, 2012; Zimmerman, 2000). Together, these phases represent the various processes undertaken within self-regulation, and are cyclical, as feedback from prior efforts informs adjustments to both current and future learning attempts (Zimmerman, 2000).

In order for students both to recognize and to modify their self-regulatory behaviour, they must develop an awareness of their own learning and use this knowledge to control it (Winters, 2011). When students learn to determine what causes the performance of a task to be either successful or unsuccessful, they are able to go beyond the goal of just being accurate learners as they gain an understanding of the conditions under which certain learning strategies are most effective (Vilalta, Giraud-Carrier, & Brazdil, 2010). The awareness of one's own learning is referred to as meta-learning (Winters, 2011). Essentially, meta-learning involves various metacognitive aspects of learning, whereby students are aware of their motives, abilities, and the demands of a learning task, and are able to control their behaviour to achieve desired outcomes (for review see Jackson, 2004). In an extensive synthesis of over 800 meta-analyses, Hattie (2009) found that meta-cognitive strategies have some of the most powerful effects on improving student learning. For these reasons meta-learning is often associated with the theory of self-regulation (Winne, 2001; Zimmerman, 2006).

Metacognitive processes are inherently personal and therefore difficult to observe in students. In this study of allied health university students, self-reports through "meta-learning" assessment tasks were utilized to collect information about self-regulatory behaviour and to evaluate student use of different categories of self-regulatory strategies from each phase of the self-regulation cycle. However, as self-reported data has potential limitations in accuracy (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; Schmitz & Wiese, 2006), it has been combined with learning analytics of data available from the course learning management system to enhance this understanding of student self-regulatory behaviour and processes in a non-research setting.

The aims of the study were to 1) identify the self-regulatory learning strategies employed by second-year allied health students; 2) characterize student behaviour in interacting with course materials, in view of their self-reported planning and use of strategies; and 3) compare and contrast the use of these different sources of information for evaluating the self-regulatory behaviour of students in an authentic setting.

2 METHODS

2.1 Institutional and Course Context

The University of Queensland is a large, research-intensive Australian university, with over 40,000 undergraduate and 8,000 post-graduate students. The participants for this study were second-year undergraduate students in the Bachelor of Physiotherapy (n=121) or Bachelor of Speech Pathology (n=95) programs or post-graduate students in the Master of Speech Pathology (n=11) program. The entrance requirements for these programs are very high, although the academic backgrounds of students tends to differ between programs (Ernst & Colthorpe, 2007). All the students took a Human Physiology course, which covered cell, nerve, and muscle physiology, and the physiology of the cardiovascular, respiratory, and renal systems. Students enrolling in the course had an average age of 20.2 years, 75% were female, and 12% were international students. All enrolled students had access to a course Blackboard™[1] site through which they were invited to participate in the study. No incentive for participation was offered. Of the cohort, 99 students (44%) provided informed consent for their inclusion in this study, which was approved by the University of Queensland Human Experimentation Ethical Review Committee.

The course consisted of three hours of lectures every week together with three-hour practical classes in eight of the thirteen weeks of semester. Lectures were recorded through Echo360™[2] and were made available to students through a folder on the course Blackboard site 1–5 days after the lecture took place. Lectures were scheduled twice a week with a 1-hour lecture on Tuesdays and a 2-hour lecture on Wednesdays, giving a total of 26 lecture recordings. Seven practical classes were laboratory based and one practical class was an online scenario-based learning task completed during scheduled class time. In the laboratory-based classes students undertook short experiments, demonstrating physiological concepts such as osmosis or excitation-contraction coupling of skeletal muscle. Each laboratory-based practical class had an introductory video to aid students' preparation for the class, with one video covering a combination of topics from two classes, so that there were a total of six videos. These videos

[1] Blackboard Inc., Washington, DC, USA.
[2] Echo360 Inc., Dulles, VA, USA.

were also contained in a folder on the course Blackboard site and were available from the commencement of the semester; student access of that folder was tracked.

Assessment for the course consisted of a mid-semester exam worth 20%, a written assignment worth 13%, and an end-of-semester exam worth 55%. Further four equally weighted meta-learning tasks, collectively worth 12%, were used to assess student management of learning and promote self-regulatory behaviour.

2.2 Meta-Learning Assessment Tasks

Students completed four meta-learning assessment tasks during the semester at approximately three-week intervals. Each task comprised six questions designed to help students articulate their own learning and engage in learning strategies from all the self-regulatory phases (Zimmerman, 2000). The first meta-learning task was developed with the purpose of determining the goals and motivations of students; it asked students to articulate the study strategies they had used in the past and identify hindrances to their learning. The second task was completed two weeks prior to the mid-semester exam and was developed with the aims of increasing student awareness of their understanding of course content, articulating strategies they may use to improve their learning, and promoting effective study for the mid-semester exam. Students completed the third meta-learning task after the mid-semester exam. It aimed to encourage students to reflect on the strategies they had used for mid-semester exam study, to determine the effectiveness of these strategies, and to consider how they could improve their study for future exams. The final meta-learning task completed the self-regulatory cycle by allowing students to reflect once again on their learning and propose strategies they could use to assist their study for the upcoming end-of-semester exam (Zimmerman, 2000). Students were awarded a small number of marks (0.5% per answer) if the answer was appropriate and relevant, with the vast majority of students who completed the task receiving full marks. Each meta-learning task was available for one week on the course learning management system, and students could access and submit their completed task at any time during that week. The date and time of each of their submissions was recorded.

2.3 Qualitative and Quantitative Analysis

To identify the self-regulatory strategies students reported using, the first meta-learning task included a question asking students to list all the learning strategies they employed. The responses of consenting students were then categorized, based on the classification system of Nota et al. (2004) adapted to suit the university setting (Table 1). These adaptations included the following: 1) splitting the "goal setting and planning" category to identify student goals and the strategies they planned to use to achieve those goals; 2) expanding the "seeking social assistance" category to include collaborative learning; and 3) splitting the "organization and transforming" category into two. "Organization" was defined as time and resource management. "Transforming" was expanded to include actively reappraising records and summarizing them into various forms of information, such as concept maps, lists of key points, and

diagrammatic representations. Where a student response encompassed more than one category, that response was coded to all relevant categories. Blind coding was initially performed by one researcher; a subset (33%) of responses was then blindly coded by a second researcher, demonstrating an inter-rater coding reliability of 82%. Once coding was complete, the number and type of strategies reported by each student were used for comparison to academic performance and learning analytics data.
Across the four meta-learning tasks, the students were asked a series of questions probing their self-regulation strategies. Students reported on 1) strategies they had used in the past; 2) strategies they intended to use to study for the mid-semester exam; 3) strategies they had used to study for that exam; and finally 4) strategies they intended to use to study for the end-of-semester exam. Students were categorized on whether they did or did not mention using lecture recordings in their study in at least one meta-learning task.

Overall, students had five intra-semester assessment tasks submitted through the learning management system. Four of the meta-learning tasks were open for submission for one week. The fifth item was an assignment, with topics provided in week 3 of semester, and submissions due in week 11. The submission date and time for each student was collected from the learning management system.

Throughout this study, quantitative analyses were performed using the GraphPad Prism 6™[3] or R 3.1.1[4] programs. The results were expressed as mean and standard error of the mean (SEM), and were considered significant if $p<0.05$. Where possible, anonymous aggregate data from the whole cohort was used for analyses, but where comparisons of qualitative data to academic performance or access to resources were performed, only de-identified data from consenting students were used. The end-of-semester exam results of consenting (n=99, 44% of the cohort) and non-consenting students (n=127) were subjected to an unpaired t-test. The mean end-of-semester exam performance for the consenting students (73.4 $\pm$ 1.2) was not significantly different from that of the whole cohort (71.7 $\pm$ 0.8), suggesting that the consenting students are representative of the range of academic performance levels present in the cohort.

Exploratory cluster analysis was used to determine if there were differences in academic performance (course grade) between groups of consenting students who shared approaches to the regulation of their learning, extent and timing of lecture recording use, and timing of assessment submission. In stage 1, Ward's (1963) hierarchical method was used to produce a cluster dendrogram (Figure 1). The clearest demarcation partitioned students into two clusters. In stage 2, the k-means method with Euclidean distances was used to explore the characteristics of partitioning students into 5, 3, and 2 cluster groupings. Reducing k-means from 5, to 3, to 2 dramatically improved the clarity of the relationship between cluster allocation and academic performance. In the final k=2 clustering, the cluster of high performers scored 80.2+/–0.9% (n=81 students) while the cluster of low performers scored 68.3+/–1.8% (n=13 students).

[3] San Diego, CA, USA.
[4] R Development Core Team, Auckland, NZ.

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

Table 1. Strategy classifications, definitions (adapted from Nota et al., 2004), phases (Zimmerman, 2000), and example student responses from meta-learning tasks.

Self-Regulation Phase	Strategy	Definition	Examples from Student Responses
Forethought	Goal Setting	Setting goals or sub-goals (student initiated).	"Set smaller and more manageable goals over a long period of time (go through a few concepts a day, not the whole 3 or 4 weeks of lectures in a few hours)."
	Strategic Planning	Developing plans to use specific learning strategies or behaviours.	"In the past, I would review lecture notes and take notes in lectures. I plan on doing the same this semester with an even greater emphasis on taking notes."
Performance	Environmental Structuring	Adapting physical or virtual surroundings to be conducive to learning.	"I definitely use the private study section of the library to remove distractions."
	Keeping Records	Recording events or results (student initiated).	"Listening attentively in lectures and taking down relevant points which would aid my revision."
	Organisation	Managing time and resources appropriate to the task.	"Organise set times to complete everything by a set time or date (like study for a set amount of time, and then if done I allow myself to go out with friends the next day etc.)"
	Reviewing Records	Re-reading notes or supplied resources, or accessing lecture recordings.	"Regularly reviewing my lecture notes so that I'm not learning it all for the first time just before exams."
	Seeking Information	Securing further relevant information from non-social sources.	"Watching YouTube videos or researching concepts to get concepts explained in different ways."
	Seeking Social Assistance	Learning collaboratively with others or seeking help from peers, lecturers, and others	"Discussing with friends and asking them questions, explaining to each other all the different concepts and principles."
Self-Reflection	Active Reappraisal or Transformation of Records	Appraising and rearranging records or resources to improve learning.	"I write notes and summaries for each lecture, lastly I draw concept maps, diagrams, or flow charts to better visualise the information learnt during the lectures."
	Self-Evaluation	Evaluating quality or progress of learning or effectiveness of strategies used.	"Doing past year exam questions which reinforces my knowledge and alerts me to what I do not know."

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

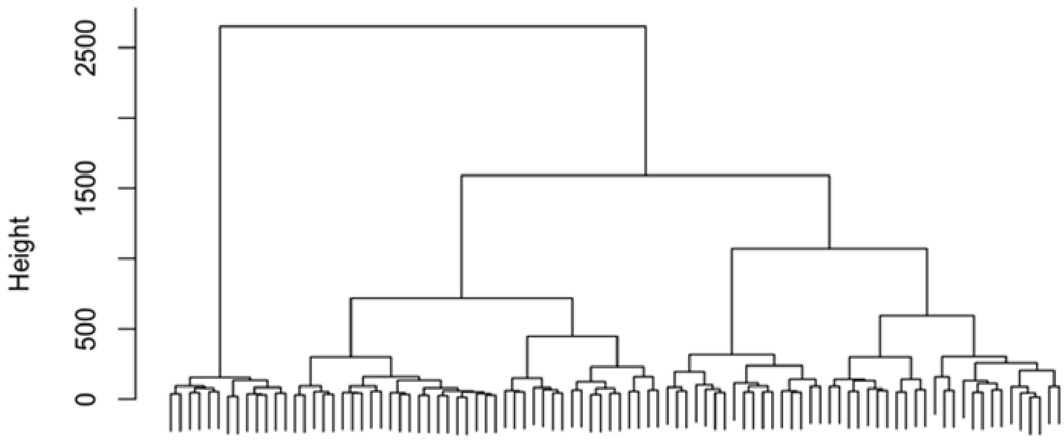


Figure 1. Cluster dendrogram of consenting students (n=99) based on access of lecture recordings, meta-learning and assignment submission dates, and assignment and course grades. Height represents the degree to which clusters of students can be differentiated based on the variables used for analysis.

3 RESULTS

Students (n=97) reported that they previously used an average of 5.3 $\pm$ 0.2 learning strategies, with students reporting between 1 and 15 strategies each. A total of 511 responses were coded, with reports of strategies from across all 10 categories. The most commonly reported category was "reviewing records" (71% of students), which included reviewing lecture recordings. The least reported category was "goal setting," reported by only 1% of students. All students reported using strategies from the performance phase of the self-regulatory cycle, 68% also used strategies from the self-reflection phase, but only 8% reported strategies from the forethought phase (Figure 2). Analysis by unpaired t-test with Welch correction showed that students who reported the use of strategies from more than one phase of the self-regulatory cycle (n=73) performed significantly better in the course overall (79.1 $\pm$ 1.08%) than students who reported using only strategies from the performance phase (74.3 $\pm$ 1.68%). Further analyses using Pearson's correlation were performed to determine the relationship between individual self-regulatory strategies and overall achievement in the course. There was a small, but significant, negative correlation between environmental structuring and academic performance ($r = -0.2019$, $p<0.05$) indicating that students who reported undertaking environmental structuring more frequently were likely to achieve lower academically than those who did not. Surprisingly, there were no other significant correlations between the use of individual strategies and overall performance.

The most frequently reported strategies of self-regulation (90% of students) were reviewing records and active reappraisal of records, which included references to using lecture recordings. Across all four meta-learning tasks, 48% of students specifically reported viewing lecture recordings as a learning strategy, a further 18% said they reviewed lectures but did not specify whether these were lecture notes

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

or recordings, whilst 34% of students did not specifically mention the lecture recordings but did mention reviewing lecture notes. Students who mentioned using lecture recordings were twice as likely to report using lecture recordings rather than planning to use them. This "reactive" pattern is consistent with the low level of planning strategies that students reported (Figure 2).

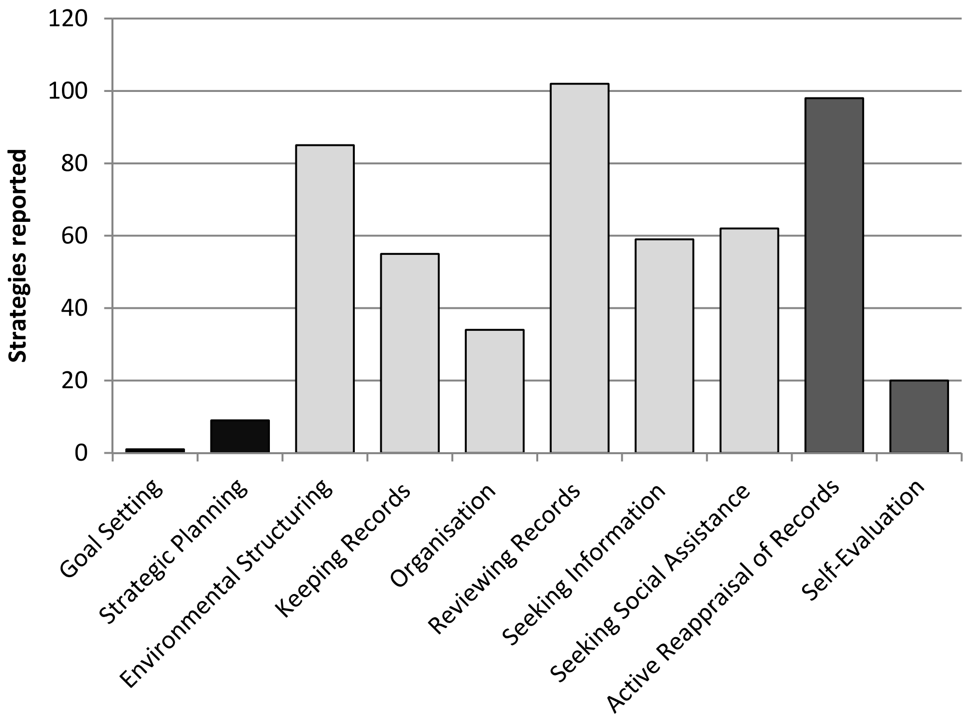


Figure 2. Frequency distribution of the categories of self-regulatory strategies (Nota et al., 2004) consenting students (n=97) reported they had previously used. Students used an average of 5.3 ± 0.2 strategies each, ranging between 1 and 15 strategies from the forethought (black bars), performance (light grey bars) and self-reflection (dark grey bars) phases of the self-regulatory cycle (Zimmerman, 2000).

Although only half of the students explicitly stated in meta-learning assessment tasks that they used lecture recordings to support their learning, every student in this study accessed the folder containing the lecture recordings at some point during the semester. There was an extremely large range in the extent of lecture recording use, from a student who accessed lecture recordings on just 3 days of the 144-day semester, to a student who accessed lecture recordings on 41 days over the semester. Lectures for this course were held every Tuesday and Wednesday, and there was a general pattern of increased access to lecture recordings between Tuesday and Thursday for every week of semester (Figure 3), indicating a consistent use of lecture recordings around the time of each recording. However dramatic peaks in access occurred in the days prior to the mid-semester and end-of-semester exams (Figure 3).

On average, individual students accessed the lecture recordings folder 48.1 ± 2.4 times, but the variation between students was very large, with individuals accessing the folder anywhere between 4 and 194

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

times. Students who reported reviewing lectures as a learning strategy in meta-learning were significantly (p<0.0001) more likely to access the folder more frequently, and on more days, than students who did not report using lecture recordings (Table 2).

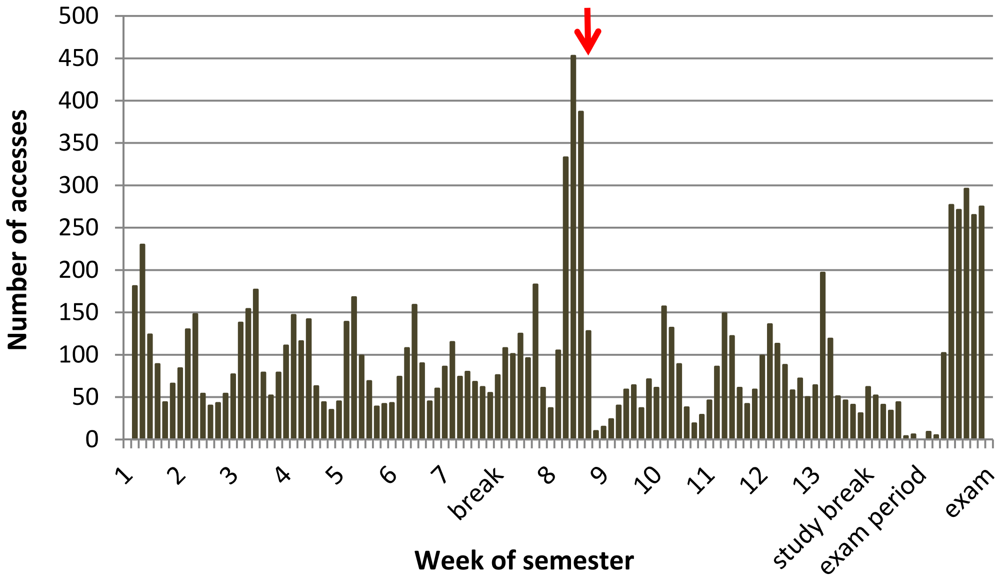


Figure 3. Student (n=227) access of the folder containing lecture recordings. Lectures commenced on Tuesday of week one of the semester, and were added sequentially as the semester progressed. The folder ultimately contained 26 recordings; two for each week of the semester. The mid-semester exam was conducted at the end of week 8 (see arrow).

Table 2. Comparison of consenting student (n=99) actual access of the lecture recordings and their reported use or intention to use lecture recordings as a learning strategy in meta-learning analysis. Data were analyzed using a non-parametric Mann Whitney *U* test.

Meta-Learning Analysis *(did and/or intended to use lecture recordings)*	***n***	***Frequency***	***Days***
Yes reported	47	66.1 ± 5.6	21.3 ± 1.3
Never reported	52	28.2 ± 2.5***	13.4 ± 0.9***

*** Significant p<0.0001 difference between groups

The majority of students (71%) reported using learning strategies that involved organizing their learning. Some examples of such organizational strategies included modifying their environment, either physically ("*I definitely use the private study section of the library to remove distractions*") or virtually ("*disconnecting the Internet so I don't suddenly find myself online shopping or on Facebook etc.*"); managing their time ("*Studying for shorter 1-hour blocks, rather than 3-hour blocks. This is to aid my focus and attention*"); or their behaviour ("*making lists of things I need to do*"). These represent strategies of self-control and self-observation (Zimmerman, 2000) carried out as part of the performance phase of the self-regulatory cycle (Figure 2).

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

Of the whole cohort, 65% of students accessed the online folder containing the practical introduction videos, with the majority of them accessing it one to three times. The access occurrence was highest just prior to the first practical class, in week 3 or 4 of semester, with 56% of students having accessed the folder by the time their first class was scheduled (Figure 4), and access declining thereafter. Higher frequencies of access tended to occur on Thursdays and Fridays, which corresponds to the scheduled time of the classes. While only 7.4% of students accessed the folder six or more times, it was possible to watch or download more than one video at any given access, so use is probably greater than this percentage indicates.

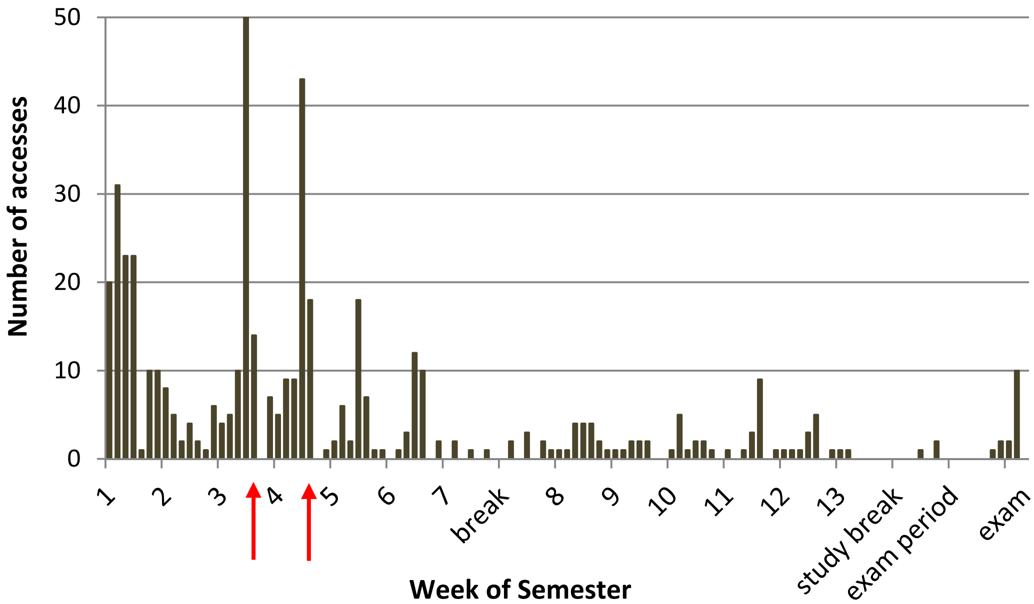


Figure 4. Student (n=227) access of the online folder containing introductory videos for the laboratory-based practical classes. Students commenced practical classes on Thursday afternoon or Friday morning of either week 3 or week 4 (at red arrows). The folder contained six videos; five specifically for a single class, with a further video containing a combined introduction for two of the classes.

For each meta-learning assessment task, the date and time of submission was recorded, with the average time of submission prior to the due date varying for each task (Figure 5). Students submitted the fourth meta-learning task significantly ($p<0.05$) earlier than the second meta-learning task (Figure 5). There were significant ($p<0.01$), modest, positive correlations between the times students submitted all four meta-learning tasks, indicating that the time at which students submit one meta-learning task explains 37–53% of the variation in the timing of their submission of their other meta-leaning tasks relative to the release data and the due date (Figure 5). In addition, the earliness of a student's meta-learning submission was significantly positively correlated with his or her course grade ($r = 0.25$, $p<0.05$). There was also considerable variation in the time that students submitted the assignment: 44% of students submitted within the 24 hours prior to the deadline, with a further 30% submitting 24–48 hours prior. Only a small proportion of students (1.3%) submitted late, but these were within 24 hours of the submission deadline (Figure 6).

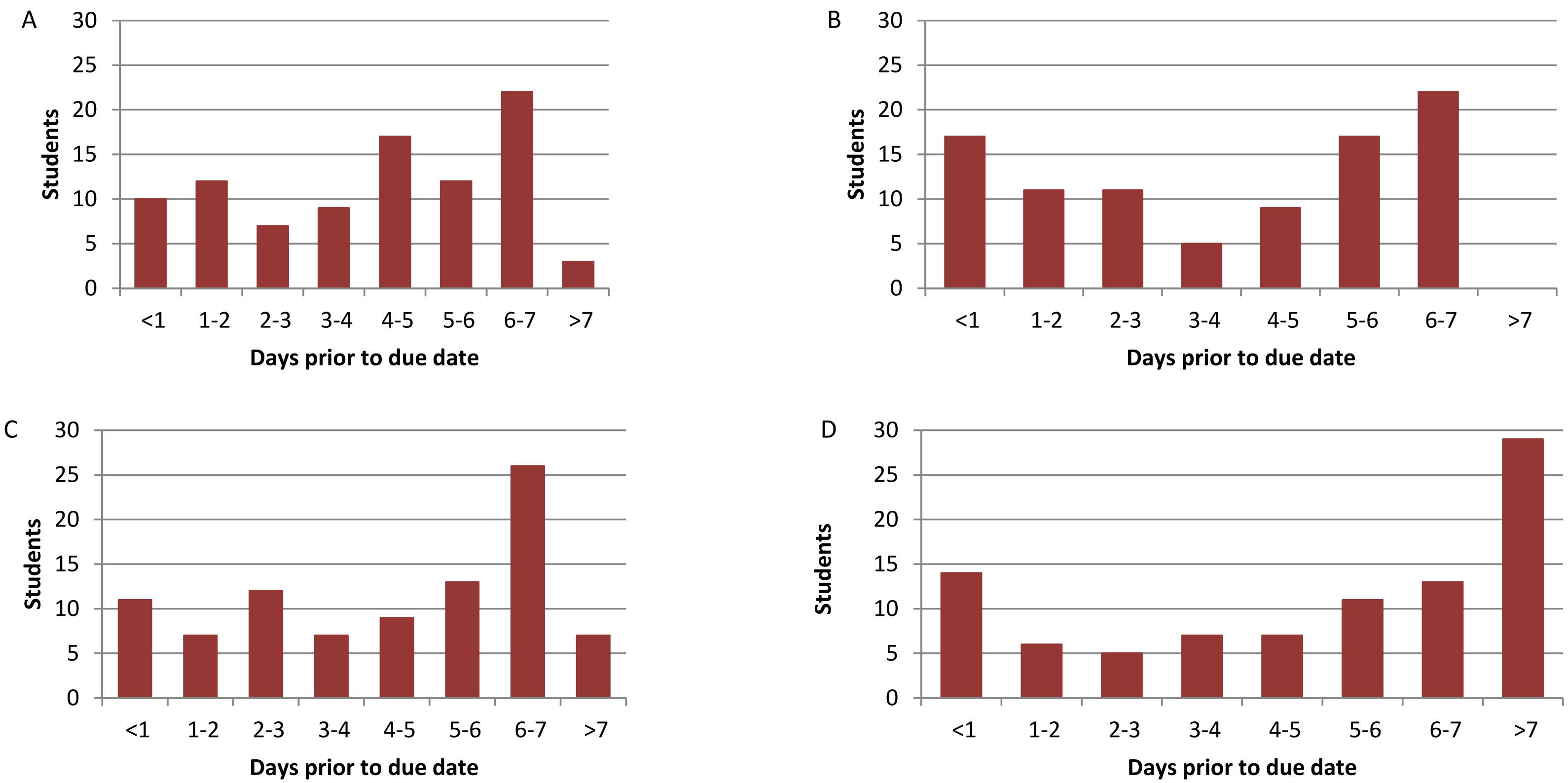


Figure 5. Submission times for meta-learning tasks 1–4 (A, ML1; B, ML2; C, ML3; D, ML4) prior to their respective deadline. Tasks were due at approximately three-week intervals, with tasks due on Wednesday of weeks 4, 7, 10, and 12 respectively. Submission time for individual students showed significant (p<0.01), modest, positive correlations.

ISSN 1929-7750 (online).

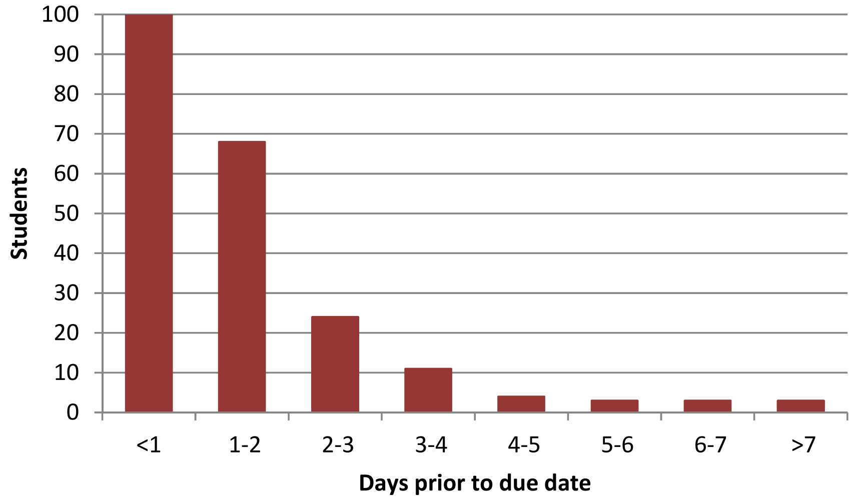


Figure 6. Time of submission of assignment prior to the due date by students (n=216). Three students submitted after the deadline (two within five minutes of it, and the other 15 hours late). The remaining seven students had extensions granted. The maximum time prior to the due date that a student submitted was 9.5 days.

Analysis of student access of the lecture recording data revealed two patterns of access behaviour, high access and low access. The access patterns are represented as a heat map on the calendar of semester (Figure 7A–B).

Table 3. Characteristics of high and low performing clusters of students (n=94).

Cluster Analysis	*High performers (n=81)*	*Low performers (n=13)*
Course grade (%)	80.2 ± 0.9	68.3 ± 1.8***
Lecture recording use (number of accesses across semester)	45 ± 4	62 ± 7*
Self-reported lecture recording use (number of meta-learning tasks)	0.77 ± 0.12	1.70 ± 0.35**
Meta-Learning task submission (days and hours before deadline across four tasks)	4 days 19 hours ± 4 hours	21 hours ± 3 hours***
Assignment submission (hours prior to deadline)	32.2 ± 4.7 hours	23.8 ± 17.7 hours*

*p<0.05, **p<0.01, *** p<0.0001 significantly different between clusters

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

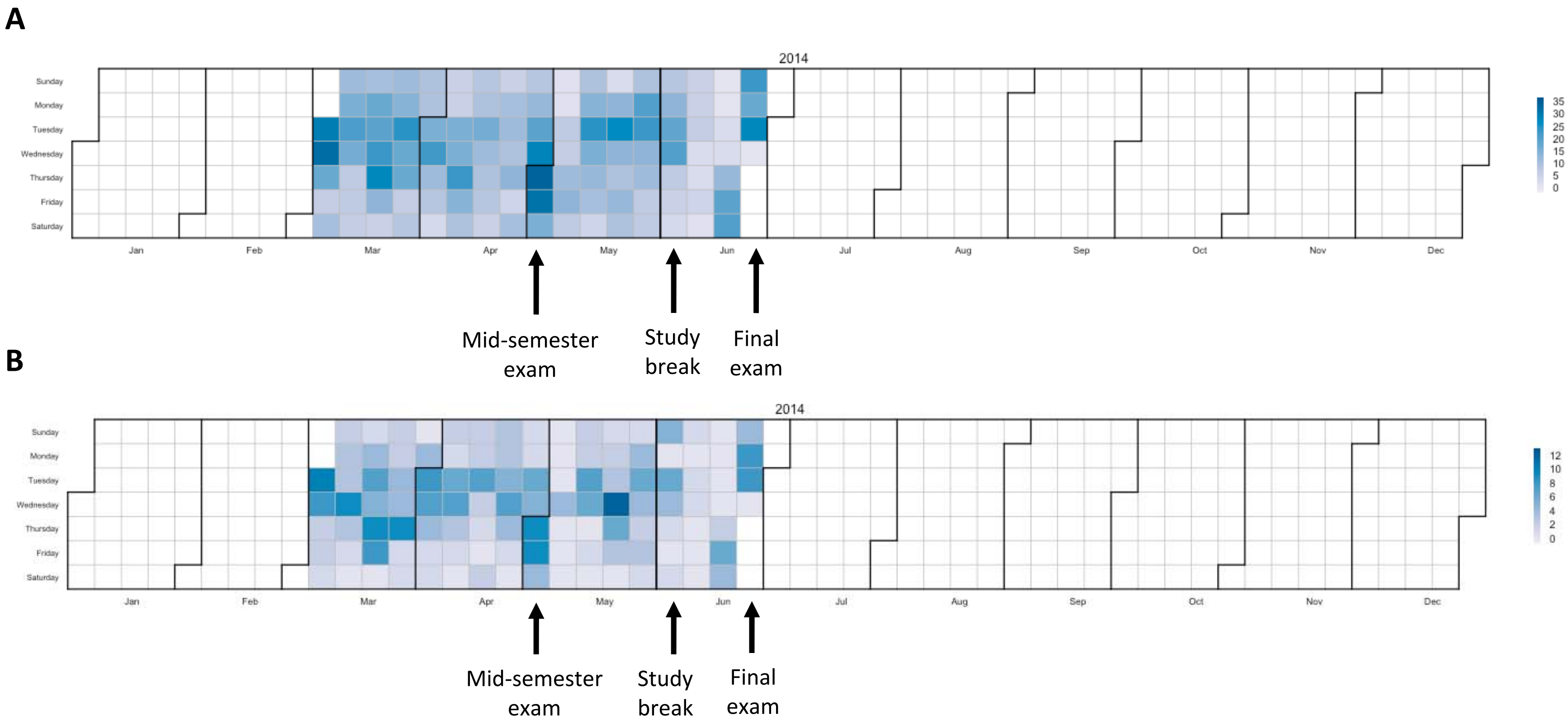


Figure 7. Student access of lecture recordings for Cluster 1 (A: High Access) and Cluster 2 (B: Low Access) as heat maps on a calendar. The darker the blue, the higher the frequency of access. Note that the scale varies from 0–35 on A and 0–12 on B. The semester commenced on March 3, with the first lecture on March 4, and the mid-semester exam on the Saturday of week 8 (May 3). The teaching period finished on the Friday of week 13 (May 30), and was followed by a one-week study break and the end-of-semester exam period, with the final exam on June 25.

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

Closer examination of the two student clusters identified from the overall analysis (Figure 1) revealed that the extent to which the two groups accessed lecture recordings differed significantly, with the high performers accessing the lecture recordings less frequently than the low performers (Table 3). However, there was no appreciable difference in the timing of lecture recording use across the semester for the high performing cluster (Figure 7A) or the low performing cluster (Figure 7B). The difference in the extent of lecture recording use was consistent with student self-reports (Table 3), with the high performing cluster indicating that they used lecture recordings in fewer meta-learning tasks (typically 0–1 task) than the low performing students (typically 1–2 tasks). Interestingly, across the four meta-learning tasks, the low performers were more likely to report that they *used* the lecture recordings, than report that they *planned to use* the lecture recordings to study for exams. This pattern was not apparent for the high performers, who neither reported using, nor planning to use, lecture recordings. Finally, the most dramatic differences between the high and low performing clusters related to assessment submission times. For all four meta-learning tasks and the assignment, the high performers submitted their work significantly earlier than the low performers (Table 3).

4 DISCUSSION

The combination of student responses in meta-learning assessment tasks together with student access to online data was used in this study to characterize allied health science students in terms of the self-regulatory strategies they employ, their organization, and their preparedness for study. However, each of these forms of information is limited in providing accurate information on students' self-regulatory processes. Some learning processes are particularly difficult to track in a traditionally run, on-campus setting; for example, information seeking, seeking social assistance, and environment structuring. Self-reporting is particularly useful for providing insight into such processes, which are not always easily tracked online in adult learners with clear preferences for certain self-regulatory strategies (for example, "Watching YouTube videos for concepts that I don't understand and need to clarify" and "Studying with friends in group discussions where we talk about concepts, and take turns explaining things to each other, in order to come to a group consensus about more difficult topics"). However, self-reported data has well-documented flaws, including variations in the interpretation of the learning context in question, and inaccuracies in self-reporting the frequency of strategies (Perry & Winne, 2006). The combined analysis of students' online access of data together with self-reporting within meta-learning assessment tasks allows a more comprehensive view of student self-regulatory behaviour. Furthermore, using this combination of sources of information should allow both a verification of the self-reported student data and a greater understanding of the learning behaviours that underlie patterns of student access to resources on the learning management system site. For academics in an on-campus, non-research environment, the combination of these tools may assist them to increase their understanding of their students' self-regulatory processes.

Analysis of the student responses to the meta-learning tasks showed that students in this study used a broad range of self-regulatory strategies, particularly those from the performance and self-reflection phases of the self-regulatory cycle (Figure 2). The breadth of learning strategies used is likely a reflection of the educational experiences of the students. Our study cohort were second-year undergraduates and post-graduates in intellectually demanding programs at an advanced stage of their education, so consequently they appear to have a broad repertoire of learning strategies. In addition, those who utilized strategies from multiple phases of the self-regulatory cycle performed better, suggesting that a diversity of strategies is beneficial. The most common strategies reported were the review and reappraisal of records, including using both lecture recordings and notes. Many students also reported that they structured their environment to be conducive to learning, although their performance was likely to be poorer the more frequently they did so. This may reflect that students with very specific environmental needs may not be adaptable to the varying learning contexts at university or are easily distracted within them. Interestingly, based on the initial questioning of students undertaken here, fewer than 10% of students reported using the forethought strategies of planning or setting goals for their learning. For example, only a single student reported setting goals as a learning strategy. However, this finding needs to be viewed with caution as the responses to the meta-learning questions, while rich, are dependent on context.

Goal setting has previously been identified as an important learning strategy, being closely linked to academic achievement, with students who set a combination of mastery and performance goals having better academic achievement than those who do not (Sandars & Cleary, 2011; Luo, Paris, Hogan, & Luo, 2011). Strategic planning is also considered an important self-regulatory process, but is cognitively demanding, as it requires learners to understand the area of expertise they wish to acquire, have insight into their existing knowledge, and possess pedagogical knowledge to make informed decisions (Bonestroo & de Jong, 2012). Given these considerations, strategies from the forethought phase are considered both more advanced than those from the performance phase and essential for effective self-regulation (Postholm, 2011; Schunk, 2008; Turner & Husman, 2008; Zimmerman, 2008). It is therefore a concern that these students did not appear to be engaging in forethought processes. However, in the first meta-learning task, students were not specifically asked to identify their goals or describe their plans for learning; rather they were asked to describe the learning strategies they had previously used. When prompted to undertake forethought processes in the second meta-learning task, by being asked their intentions in the period leading up to the mid-semester exam, students clearly demonstrated that they could strategically plan their learning, reporting multiple strategies when prompted, but again very few mentioned setting goals for their learning. Together, these findings suggest that either generally students don't engage in forethought processes unless prompted, or that their forethought processes are primarily limited to planning. Alternatively, students may set goals, but not consider goal setting to be a "strategy" for learning.

Students who reported using or intending to use lecture recordings in the meta-learning tasks accessed the recordings significantly more frequently and on more days (Table 2). In addition, there were clearly defined clusters of student learning behaviour in using lecture recordings (Figure 7), identified both within online access and meta-learning data with high prediction accuracy, with students who exhibit extensive use of lecturing recordings having poorer academic performance than those who made less use of the recordings (Table 3). This is not surprising if the meta-learning responses of the students are considered in more detail. Where students reported simply that they reviewed lectures, regardless of the form, these were classified as "reviewing records," which is part of the performance phase of the self-regulatory cycle (Zimmerman, 2000; Nota et al., 2004). "Reviewing records" was the most commonly reported category, and most reports of lecture recording fell into this category. Whereas, if students reported that they engaged with the lecture material more actively — for example, if they used the material to create summaries, concept maps, and diagrams of the material, or developed lists or questions about key concepts — then these were classified as "active reappraisal of records" (Nota et al., 2004). Few students mentioned using the lecture recordings for this purpose. If students actively engage with learning materials, when they go beyond simple review of their notes and/or the supplied resources, then they are engaging with the material in a more reflective way. In addition, students may use the material they develop to test themselves or may use the supplied practice exams. By evaluating their knowledge and understanding in these types of strategies, students engage in key processes within the self-reflection phase of the self-regulation cycle (Zimmerman, 2000). These processes are more likely to have a positive impact on student learning than the simple review of records.

More than 70% of students reported using learning strategies to organize their learning, particularly in terms of creating optimal opportunities for study, by managing their time, their behaviour, or by modifying their environment (Figure 2). The use of these strategies reflects that students are monitoring and moderating their learning, as part of the performance phase of the self-regulatory cycle. The main meta-cognitive processes within this phase are self-control and self-observation, with self-control requiring attention to and awareness of one's actions and how they affect outcomes, and self-observation occurring when students systematically monitor their performance (Zimmerman, 2000; Schunk, 2005). Furthermore, engagement in these processes has been previously shown as a significant predictor of academic achievement (Kitsantas, Winsler, & Huie, 2008). While these types of organizational learning strategies may be difficult to identify solely via tracking student access to resources on course learning management system sites, the extent of a students' organization may be reflected in their approach to intra-semester assessment tasks. Therefore the punctuality or "earliness" of the assessment submission was analyzed (Figures 5 and 6) and compared to academic achievement and reporting of organizational strategies (Table 3). There was a significant relationship between the times students submitted the meta-learning assessment tasks and assignment with the overall course grade (Table 3). However, when comparing the self-reporting of organizational strategies by students, there was no relationship with course grade, except for environmental structuring, which had a small negative relationship with academic performance. These findings suggest that tracking student intra-semester assessment submission times is predictive of academic achievement, but not an indicator of their self-reported level of organization. However, the combination of tracking of submission times and targeted questioning within regular meta-learning tasks may allow academics to intervene and promote improved planning and organizational behaviour by their students.

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

It was difficult to verify the extent to which students prepared for learning from the access patterns of learning resources in this course. The course was run traditionally, with the online resources available prior to classes being confined to lecture notes, assessment information, and introductory videos for the practical classes. Access to lecture notes was often limited to a short time (a few days) prior to the scheduled lecture time, and consequently student access to these resources was not tracked. However, student access to the introductory practical videos provides some insight into their preparedness. Over half the students accessed these resources prior to their first practical class, most immediately prior to the class, whilst 65% of all students accessed them at some time during the semester (Figure 4). Considering this data in relation to the low number of students who reported engaging in forethought processes in the first meta-learning task, this finding suggests that students tend to under-report their planning and preparedness for learning, unless specifically prompted to articulate their forethought processes. As it has been demonstrated that the most effective self-regulated learners implement strategies from all phases of the self-regulation cycle (Cohen, 2012; Kitsantas, 2002; Schunk & Swartz, 1993; Aregu, 2013; Cleary, Zimmerman, & Keating, 2006), further verification of the extent to which students engage in forethought strategies is needed.

There is good evidence to suggest that prompting students to engage in meta-learning is beneficial to them, increasing their self-regulatory skills and deepening their understanding of their learning (Hattie, 2009), hence the inclusion of meta-learning assessment tasks in our course design. The primary purpose and value of these tasks was to prompt students to undertake self-regulation, particularly to promote the use of learning strategies from the forethought and self-reflection phases. Strategies from these phases are considered more advanced, and using a combination of strategies from all three phases has been associated with higher academic achievement (Pintrich, 1995; Zimmerman, 2002; Kitsantas, 2002). In this study, the data generated from these tasks was utilized to develop an understanding of students' self-regulatory behaviour. However, although rich, the data is rather laborious to analyse and is limited by the accuracy of self-reporting. While calibration of this form of self-reporting by students and their self-regulatory behaviour has not been undertaken, other forms of self-reportage have shown widely differing levels of accuracy (Hadwin et al., 2007; Schmitz & Wiese, 2006). Comparison of the findings from the meta-learning data with the online access data from the learning management system suggests that students under-report the extent of their self-regulatory behaviour. However, student responses to the open-ended questions in the meta-learning tasks often included lengthy and detailed descriptions on the nature and types of strategies employed, and there is close alignment in the cluster analysis on the lecture recordings with the student reports. Together, this suggests that although under-reporting, students are not necessarily inaccurate in the types of strategies they describe.

Due to time constraints, analysis of the meta-learning data was confined to categorization and quantification of the self-regulatory strategies that students reported, but the lack of correlation between individual categories of strategies with performance found here are both contrary to literature and to our previous research using more detailed analysis methods (Kitsantas, Winsler, & Huie, 2008; Ogiji et al., submitted). Analysis did show, however, that student use of strategies from multiple phases

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

of the self-regulatory cycle was positively correlated with academic performance. While the meta-learning data may allow an evaluation of the categories of strategies they use, simply categorizing and quantifying appears to be insufficient to give an accurate evaluation. Indeed, we and others (Nandagopal & Ericsson, 2012; Ogiji et al., submitted) have shown that academic performance is not simply related to the use of, or particular categories of learning strategies, but rather to the quality, diversity, and timing of those learning strategies. Thus, as Winne (2010) suggests, it is likely that higher performing students have more appropriate contingencies, evaluating and adapting their behaviour to suit the task. For example, using Winne's terminology, IF a high performing student finds that reading the textbook does not produce the desired learning outcomes, they might THEN seek information from peers or academics; while a lower performing student's THEN behaviour might be to read the text book for longer (Winne, 2010). These differing adaptations may not be discernible through simple categorization or quantification of strategies. Clearly, all students are capable of self-regulating their learning, but the quality, quantity, and timing of the self-regulatory processes differs between learners (Cohen, 2012). Potentially, utilizing learning analytics to evaluate the data generated by the course learning management system may offer a fast method to gauge students' self-regulation and progress as learners, to supplement the meta-learning findings, but the way in which such data reflects the self-regulatory behaviour of students needs to be understood.

4.1 Future directions

The methodologies used to research self-regulated learning are moving away from reliance on uni-modal data sources, such as self-report questionnaires (Pintrich, 2004), think-aloud protocols (Greene, Robertson, & Costa, 2011), or computer logs (Aleven, Roll, McLaren, & Koedinger, 2010) toward multi-modal data sources (Ben-Eliyahu & Bernacki, 2015). Recent research has attempted to triangulate multiple sources, such as case studies, interviews, and observations (Lichtinger & Kaplan, 2015), or combined quantitative surveys with qualitative open-response questions (McCardle & Hadwin, 2015). However, these approaches tend to still be bound in a single paradigm (observational, self-report, or computer logs) rather than bridging diverse data sources such as the open-ended meta-learning questions and learning management site interactions presented in this study. Admittedly, this study has used narrow slices from these different methodologies as an early exploration of the degree to which such data can be triangulated. As a next step, we plan to investigate the utility of chronologically sequencing multiple computer logs with student responses to multiple meta-learning tasks. This juxtaposition will help us to determine the temporal and sequential characteristics of students' self-regulated learning behaviours and meta-cognition, both at the micro-analysis level during tasks, and across the long-term changes that occur during a semester or over semesters. Such approaches will hopefully reveal both the critical points of change in self-regulated learning, and the degree to which such changes are retained.

However, such "tracer"-like approaches to the analysis of open-ended responses and computer log data lend themselves primarily to the detailed analysis of small numbers of purposefully selected students (e.g., subsets of students with high and low academic achievement profiles). To move into the realm of large-scale analysis of such rich data sets and patterns of temporal and sequential relationships, automated methods of text analysis would be required. As an early step in this direction, basic statistical methods in natural language processing (such as word counts and the contexts in which words occur) would be used to triangulate the computer-assisted analysis of meta-learning responses with the human inductive thematic analysis of the same responses (Sherin, 2013). By approaching the same data from two analytical approaches, an enriched model of the data is more likely to be revealed. This improved model might then be used to train and test algorithms in a machine learning approach to achieve at least semi-automated analysis of large banks of meta-learning responses in conjunction with computer logs, to predict academic outcomes (e.g., Ghiasinejad & Golden, 2013). Potentially, predictive models might assist students who are struggling to self-evaluate accurately, and help students focus on metacognitive and self-regulatory behaviours appropriate to their situation.

5 CONCLUSION

In recent years, much of the research on self-regulation utilizing online tools has supplemented data from learning management systems with other data collection methods, including computer traces, think-aloud protocols, diaries of studying, direct observation, and microanalyses (Bannert, Reimann, & Sonnenberg, 2013; Ferreira, Simão, & da Silva, 2014; Hadwin et al., 2007; Perry & Winne, 2006; Schmitz & Wiese, 2006), or has focused on courses offered primarily or entirely online (Lynch & Dembo, 2004). However, in the non-research, on-campus environment when these additional tools are not available, what can the data that is available tell us? It would appear that some aspects of the currently available data may be predictive of academic performance, but the relationship between this data the self-regulation and progress of the students is complex. Potentially though, the combination of tracking student engagement with online resources, intra-semester assessment submission times and targeted questioning within regular meta-learning tasks may allow coalface academics both to increase their understanding of student self-regulatory behaviour and to create interventions to promote improvements student learning outcomes.

REFERENCES

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2010). Automated, unobtrusive, action-by-action assessment of self-regulation during learning with an intelligent tutoring system. *Educational Psychologist, 45*(4), 224–233.

Aregu, B. B. (2013). A study of self-regulated learning strategies as predictors of critical reading. *Educational Research and Reviews, 8*, 1961–1965.

Bannert, M., Reimann, P., & Sonnenberg, C. (2013). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning, 9*(2), 161–185.

Ben-Eliyahu, A., & Bernacki, M. L. (2015). Addressing complexities in self-regulated learning: A focus on contextual factors, contingencies, and dynamic relations. *Metacognition and Learning, 10*(1), 1–13.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161–186.

Bonestroo, W. J., & De Jong, T. (2012). Effects of planning on task load, knowledge, and tool preference: A comparison of two tools. *Interactive Learning Environments, 20*(2), 141–153.

Cleary, T. J., & Zimmerman, B. J. (2004). Self-regulation empowerment program: A school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools, 41*(5), 537–550.

Cleary, T. J., Zimmerman, B. J., & Keating, T. (2006). Training physical education students to self-regulate during basketball free throw practice. *Research Quarterly for Exercise and Sport, 77*(2), 251–262.

Cohen, M. T. (2012). The importance of self-regulation for college student learning. *College Student Journal, 46*(4), 892–902.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280.

Ernst, H., & Colthorpe, K. (2007). The efficacy of interactive lecturing for students with diverse science backgrounds. *Advances in Physiology Education, 31*(1), 41–44.

Ferreira, P. C., Simão, A. M. V., & Da Silva, A. L. (2014). Does training in how to regulate one's learning affect how students report self-regulated learning in diary tasks? *Metacognition and Learning*. doi:10.1007/s11409-014-9121-3

Ghiasinejad, S., & Golden, R. M. (2013). Modeling human coding of free response data. *Computers in Human Behavior, 29*(6), 2394–2403.

Greene, J. A., Robertson, J., & Costa, L. J. C. (2011). Assessing self-regulated learning using think-aloud methods. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 313–328). New York, NY: Routledge.

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*(2–3), 107–124.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses related to achievement*. Abingdon, UK: Routledge.

Jackson, N. (2004). Developing the concept of metalearning. *Innovations in Education and Teaching International, 41*(4), 391–403.

Kitsantas, A. (2002). Test preparation and performance: A self-regulatory analysis. *The Journal of Experimental Education, 70*(2), 101–113.

Kitsantas, A., Winsler, A., & Huie, F. (2008). Self-regulation and ability predictors of academic success during college: A predictive validity study. *Journal of Advanced Academics*, *20*(1), 42–68.

Lichtinger, E., & Kaplan, A. (2015). Employing a case study approach to capture motivation and self-regulation of young students with learning disabilities in authentic educational contexts. *Metacognition and Learning, 10*(1), 119–149.

Luo, W., Paris, S. G., Hogan, D., & Luo, Z. (2011). Do performance goals promote learning? A pattern analysis of Singapore students' achievement goals. Contemporary *Educational Psychology, 36*(2), 165–176.

Lynch, R., & Dembo, M. (2004). The relationship between self-regulation and online learning in a blended learning context. *The International Review of Research in Open and Distance Learning, 5*(2). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/189/271

McCardle, L., & Hadwin, A. F. (2015). Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. *Metacognition and Learning*, *10*(1), 43–75.

Nandagopal, K., & Ericsson, K. A. (2012). An expert performance approach to the study of individual differences in self-regulated learning activities in upper-level college students. *Learning and Individual Differences, 22*(5), 597–609.

Nota, L., Soresi, S., & Zimmerman, B. J. (2004). Self-regulation and academic achievement and resilience: A longitudinal study. *International Journal of Educational Research*, *41*(3), 198–215.

Ogiji, J., Colthorpe, K., Ainscough, L., Zimbardi, K., & Anderson, S. (submitted). Meta-learning assessment tasks provide insight into undergraduate students' self-regulatory learning behaviour.

Perry, N. E., & Winne, P. H. (2006). Learning from learning kits: gStudy traces of students' self-regulated engagements with computerized content. *Educational Psychology Review, 18*(3), 211–28.

Pintrich, P. R. (1995). Understanding self-regulated learning. *New Directions for Teaching and Learning, 63*, 3–12.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational psychology review, 16*(4), 385–407.

Postholm, M. B. (2011). Self-regulated learning in teaching: Students' experiences. *Teachers and Teaching, 17*(3), 365–382.

Sandars, J., & Cleary, T. J. (2011). Self-regulation theory: Applications to medical education. AMEE Guide No. 58. *Medical Teacher, 33*(11), 875–86.

Schmitz, B., & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology, 31*(1), 64–96.

Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist, 40*(2), 85–94.

Schunk, D. H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review, 20*(4), 463–467.

Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology, 18*(3), 337–354.

Sherin, B. (2013). A computational study of commonsense science: An exploration in the automated analysis of clinical interview data. *Journal of the Learning Sciences, 22*(4), 600–638.

Turner, J. E., & Husman, J. (2008). Emotional and cognitive self-regulation following academic shame. *Journal of Advanced Academics, 20*(1), 138–173.

Vilalta, R., Giraud-Carrier, C., & Brazdil, P. (2010). Meta-learning-concepts and techniques. *Data mining and knowledge discovery handbook*. New York, NY: Springer.

(2015). Know Thy Student! Combining Learning Analytics and Critical Reflections to Increase Understanding of Students' Self-Regulated Learning in an Authentic Setting. *Journal of Learning Analytics*, *2*(1), 134–155.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244.

Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences, 8*(4), 327–353.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives,* 2nd ed. (pp. 153–189). New York, NY: Lawrence Erlbaum.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*(4), 267–276.

Winters, T. (2011). Facilitating meta-learning in art and design education. *International Journal of Art & Design Education, 30*, 90–101.

Zimmerman, B. J. (ed.) (2000). *Attaining self-regulation: A social cognitive perspective*. Burlington, MA: Elsevier Academic Press.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70.

Zimmerman, B. J. (2006). Development and adaption of expertise: The role of self-regulatory processes and beliefs. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance.* Cambridge, UK: Cambridge University Press.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166–183.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses

Justin Reich, Dustin Tingley, Jetson Leder-Luis, Margaret E. Roberts, Brandon M. Stewart
Harvard University, USA
justin_reich@harvard.edu

ABSTRACT: Dealing with the vast quantities of text that students generate in Massive Open Online Courses (MOOCs) and other large-scale online learning environments is a daunting challenge. Computational tools are needed to help instructional teams uncover themes and patterns as students write in forums, assignments, and surveys. This paper introduces to the learning analytics community the Structural Topic Model, an approach to language processing that can 1) find syntactic patterns with semantic meaning in unstructured text, 2) identify variation in those patterns across covariates, and 3) uncover archetypal texts that exemplify the documents within a topical pattern. We show examples of computationally aided discovery and reading in three MOOC settings: mapping students' self-reported motivations, identifying themes in discussion forums, and uncovering patterns of feedback in course evaluations.

KEYWORDS: Massive Open Online Courses, topic modelling, text analysis, computer-assisted reading

1 OVERVIEW

Educators are constantly asking their students to write. They articulate needs and motivations in pre-course surveys, communicate and collaborate in forums, demonstrate their understanding in assignments, and offer feedback about instructional approaches in course evaluations. In classes with low student–instructor ratios, instructional teams of faculty and teaching assistants can read, process, and provide feedback on the entire corpus of text produced by students. In large-scale learning environments, such as Massive Open Online Courses (MOOCs), there is far too much for instructors to read and process in a timely fashion. Consider two sources of student text: surveys and discussion forums. In the first year of operation, hundreds of thousands of students signed up for HarvardX courses on the edX platform, and they submitted over 240,000 answers to the open-response survey question: "Please share your reasons for signing up for edX." In the inaugural MITx class, the discussion forums included over 12,000 threads and nearly 100,000 individual posts (Breslow, Pritchard, DeBoer, Stump, & Ho, 2013). These corpora represent two troves of important data. Understanding what motivates students to sign up for courses can help course developers tailor instruction to their students. Discussion forums are central sites for advancing student learning in many MOOCs, especially in the professions (Fisher, 2014; Reich et al., 2014a) and humanities (Reich et al., 2014b), but these spaces can rapidly become overwhelming to follow or analyze, especially for faculty discovering the arduous demands of teaching a MOOC (Grainger, 2013; Kolowich, 2014).

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

In this paper, we introduce advances in computer-assisted techniques for discovery and analysis of student-produced text, and we illustrate these techniques with examples from MOOC pre-course surveys, discussion forum threads, and course evaluations. We demonstrate a method of conducting text analysis known as the Structural Topic Model (STM) (Lucas et al., 2013; Roberts, Stewart, Tingley & Airoldi, 2013; Roberts et al., 2014). Topic models, such those based on Latent Direchlet Allocation (Blei, Ng, & Jordan, 2003), can uncover meaningful patterns within collections of documents. These computer algorithms can identify syntactic patterns among texts, and these syntactic patterns often prove to have useful semantic meaning. Our methods for Structural Topic Models make three new contributions to these methods. First, STM can incorporate additional covariates about the author or context of a document. Topic models can identify patterns of responses to a course evaluation question like "How might this course be improved?" and the STM can characterize how patterns of responses to that question vary by important covariates, such a student's overall course satisfaction. Second, STM methods are built into a software package that produces a set of intuitive visualizations to support analysts in finding and exploring patterns. Finally, the software package is open source and designed for use by novices in text-analysis. These tools have opened up exciting new areas of research in many fields in the social sciences (Eggers & Spirling, 2011; Jamal, Keohane, Romney, & Tingley, 2014; King, Pan, & Roberts, 2013; Stewart & Zhukov, 2009; Stockmann, 2012; Van Atteveldt, Kleinnijenhuis, & Ruigrok, 2008), and their development is timely as MOOCs and other large-scale online learning environments expand.

One of the primary goals of the STM project is to make these text analysis methods accessible to a wide variety of researchers and practitioners. Our focus in this paper, therefore, is not on the technical details of the STM procedure, which are described in Roberts, Stewart & Airoldi (2014). Rather, we demonstrate how topic models may be used in the service of computer-assisted reading, leveraging computational technologies to allow educators to investigate massive quantities of student writing systematically in a reasonable amount of time. While there is no systematic research (that we are aware of) that characterizes emerging faculty practices, in our interviews with HarvardX faculty we found that instructional teams' reviews of discussion forums, assessments, and surveys are limited to the cursory, informal, and idiosyncratic. Our hope is that STM methods provide faculty with another option: to use real-time summaries of patterns found within student-produced text to make pedagogical decisions and course corrections. These same tools can also assist administrators and educational researchers in analyzing student learning in large courses.

We begin with an introduction, suitable for novices in the field, to methods of text analysis based on machine learning in order to situate the affordances of the STM. We then illustrate our approach through analyses of three different data sources drawn from MOOCs. First, we analyze responses to the aforementioned question about student rationales for signing up for an online course. Analyzing these free response items with STM allows educators and researchers to examine how students articulate their motivations in their own words, and then evaluate quantitatively how student motivations correlate with demographic characteristics or course-taking behaviour (Kizilcec & Schneider, 2015). In

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

this first example, we provide a detailed walkthrough of the workflow for using the STM package and analyzing the resulting output.

Next, we examine a second case study: the discussion forums in a humanities course from HarvardX, ChinaX Part 1.[1] In many courses, especially in the humanities where large-scale, valid, reliable assessment methods are considered lacking (Ho et al., 2014), discussion forums are considered essential sites of student learning central to the objectives of the course (Reich et al., 2014a; Reich et al., 2014c). Recent research in evaluating text in these forums points to drawing small analytic samples for human coding (Stump, DeBoer, Whittinghill, & Breslow, 2013) or using supervised learning methods to classify threads into topics (Brinton et al., 2013). We demonstrate the application of the Structural Topic Model to help instructional teams understand the range and distribution of themes in student posts, and we show how the distribution of these topics varies by important features of each document's context, such as whether or not it was "up-voted" by a peer. As a third case study, we analyze course evaluation responses from the same ChinaX course, where students were asked to write about strengths and weaknesses of the course. We demonstrate using the STM to reveal how themes in open-ended feedback can vary by other student characteristics, such as overall student satisfaction. Since these methods extend beyond MOOCs to other large-scale learning environments, we conclude with suggestions for additional applications of computer-assisted reading in other contexts.

2 INTRODUCTION TO TEXT ANALYSIS

The analysis of text by hand is standard practice in the social sciences, where researchers read and hand-code documents and then analyze the results. The variety of benefits to computer-assisted text analysis over hand coding include the natural improvements in speed, the ability to process high volumes of text, and the consistency of treatment of all parts of the corpus (Grimmer & King, 2011; Hillard, Purpura, & Wilkerson, 2008; Lowe & Benoit, 2013). Humans often struggle with the development of complicated coding schemes (Quinn, Monroe, Colaresi, Crespin, & Radev, 2010), and there is some experimental evidence to suggest that humans judge clusters produced by automated methods to be more semantically coherent than even an taxonomy created by the documents' authors (Grimmer & King, 2011; Grimmer & Stewart, 2013). The large amount of text produced in online educational environments motivates using computation to identify patterns in student-generated text. These patterns can then be presented to educators, researchers, and even students for more in-depth analysis. In this sense, the tools we discuss can be thought of as aides for "computer-assisted reading."

Automated text analysis is a form of machine learning and comes in two flavors: unsupervised and supervised. In supervised learning, the user manually labels some subset of the data, which guides the computational analysis to derive parameters for classifying the remainder of the data. Humans code a subset of documents, and computers then predict how humans would have coded the rest of the full set

[1] https://www.edx.org/course/harvardx/harvardx-sw12x-china-920

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

of documents. In unsupervised learning, there is no user input besides the raw data, from which parameters of interest are derived. Computers find patterns in documents based on syntactic features (like the co-occurrence of certain words in the document), and humans then examine the substantive meaning of those patterns. Note that in either case, algorithms need no semantic understanding; the computer does not need to be able to understand what humans are communicating or to associate meaning with any of the words. Instead, the content and structure of corpora are sufficient to surface syntactic variations and patterns that often prove to have semantically meaningful correlates. In other words, computers can treat words as strings of arbitrary letters and find patterns in how those arbitrary strings co-occur in documents. Humans can later look at those (syntactically derived) patterns and find that they are useful and (semantically) meaningful. In our analysis, we focus on a class of unsupervised analysis called topic modelling.

2.1 Unsupervised Topic Modelling

Topic modelling is a particular unsupervised method that provides an approach for estimating general semantic themes within a corpus of documents (Blei, 2012). Crucially, we need not specify these themes in advance or manually annotate the input documents; the only analytic preparation required is inputting the raw textual data into a software package. Topic models use the patterns of word co-occurrences to infer semantic relationships. Loosely speaking, if two words frequently co-occur across many of the documents, we infer that they reference a similar concept or theme. The topics themselves are distributions over words. For example, consider an assignment where students write a paragraph about what they do in a typical day. One topic might be about learning, and give high probability to words such as “learning,” “homework,” “class,” but low probability to words such as “cooking” or “eating.” Each document exhibits a mixture over the topics, which encode the proportion of words within the document that the software estimates to have come from each topic. The semantic themes uncovered by the model provide a useful structure for summarizing large sets of documents. These methods complement human reading by organizing the unstructured corpus. Topic models have been widely applied throughout the social sciences and digital humanities (see Blei, 2012, and references therein).

Our focus here differs both in method and purpose with many existing applications of unsupervised learning in educational research (see Romero & Ventura, 2007). Previous work has focused on applications towards clustering students into different learning types using their attributes, grades, and system-use statistics. When unsupervised quantitative techniques have been applied in educational contexts to text, the focus of this work has primarily been extracting data about the generation of text, engagement statistics, rather than analyzing properties of the text itself (Anaya & Boticario, 2011; Dringus & Ellis, 2005). Some new work in this field has moved towards models that include text features along with engagement statistics; for instance, in the MOOC context, Yang, Wen, Kumar, Xing, and Rosé (2014) use a topic model analysis of discussion forums and social network features to predict student attrition among distinct student sub-communities. We build on this this line of work with accessible

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

software and methods that facilitate the substantive interpretation of topics and themes within a corpus of student writing.

2.1.1 Comparisons to Supervised Methods

One way to understand the uses, affordances, and limitations of unsupervised topic models is to compare them to supervised methods of text analysis. Whereas unsupervised methods uncover the most prevalent and overarching themes of the text, supervised methods can uncover 1) a category scheme dictated in advance or 2) a particular facet of the text defined in advance, such as whether the text has an introduction and conclusion or whether it should be held for review. The most common form of supervised learning is classification. In this setting, the researcher carefully reads a random sample of documents from the corpus and assigns each one to a category according to a pre-specified coding scheme. Supervised learning algorithms learn from this "training set" about how to classify the documents. They can then be used to categorize the larger set of unread documents in the corpus. Thus, the algorithm is taught how to classify through the examples in the training set, and it then extends the process to more documents than the analyst would be able to read alone. Pros and cons of these approaches have been thoughtfully enumerated in other fields of the social sciences (Bishop, 2006; Hastie, Tibshirani, & Friedman 2009; Murphy, 2012; Grimmer & Stewart, 2013).

One of the most well-known examples of supervised learning methods in education is automated essay scoring, which works by classifying documents along a rubric of essay quality. Then a training set of essays is scored and each essay assigned to a category by one or (ideally) multiple raters. Once these essays are scored, the algorithm classifies the remaining set of essays, perhaps flagging anomalous essays that do not appear to fit well in any category. The classifier's effectiveness can be tested by measuring the prediction accuracy within a second "testing set" of human-graded essays. Evidence from recent studies of automated essay score prediction suggests that the reliability between human graders and machine learning-based graders is similar to the reliability among human graders, at least in contexts with highly structured writing assignments that may be graded quickly (Shermis & Hammer, 2012; Duwairi, 2006). In large-scale online environments like MOOCs, it is infeasible for faculty to evaluate the individual submissions from thousands of students, and therefore faculty who wish to assign a grade to unstructured text assignments in MOOCs need to choose among self-assessment, peer-assessment, and this kind of supervised machine learning evaluation. Even in circumstances where faculty can use supervised learning methods to assign scores to individual essays, unsupervised learning methods can uncover the themes in student writing.

Supervised algorithms are useful when researchers are interested in a particular organization of documents. These algorithms require that categories be comprehensively enumerated. In the examples that follow, we show three cases where we had no *a priori* expectations for categories of student writing, and we were particularly interested in topics that we and other researchers might not have considered.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

2.2 Structural Topic Models

One distinctive affordance of STM, compared to previous approaches to topic modelling, is the ability to incorporate additional metadata (or covariates) into the model, such as information about the author or document. This allows the analyst to "structure" the corpus prior to estimation. STM is specifically designed to leverage this existing information and facilitate accurate inferences for how the observed variables relate to the latent topics.

Running the STM provides the user with the following:

1. Estimated topics, including a small set of label words most indicative of that topic and archetypal documents from each topic
2. Relationships between covariates and topics
3. The prevalence of each topic throughout the corpus along with documents most heavily focused on each topic
4. Correlation patterns between topics (i.e., which topics are most likely to occur together within a document)

A standard workflow for the STM proceeds as follows: first, educators or researchers input a corpus of documents (discussion posts, assignments, course evaluations, etc.) into the `stm` package in the open-source statistical language, R (R Core Team, 2012). At the same time, the user imports metadata about each document — such as the age of the author or whether a forum post was "up-voted" — into the software. These metadata covariates can then be used in the estimation of topic prevalence (how often a topic is discussed), topical content (the words used in discussing a topic), or both. The only other user input is the number of desired topics, which controls the granularity of the requested summary. Together the metadata and the number of topics define a probabilistic model that might have generated the data we observe. We then perform Bayesian inference and calculate the posterior distribution. In essence this finds the parameters most likely given both the model and the data observed.[2]

Once the model is fit, we can investigate relationships between the covariates and the estimated topics. For example, if we analyzed a series of open-ended responses to a question about a student's favourite aspect of a class, relationships for topic prevalence might tell us that the preferred aspects (the estimated topics) differ markedly with the overall satisfaction with the course as measured by a Likert-scale item (the observed covariate). Topical content by contrast gives us insight into the words used to describe a particular topic. Thus for example, one favourite element of the class might be the

[2] The technical details are discussed in a companion paper (Roberts et al., 2014). Briefly, the STM is a logistic-normal mixed membership topic model. Estimation proceeds using a fast semi-collapsed, variational expectation maximization algorithm where Laplace approximations are used for the non-conjugate portions of the model. As with many modern text analysis procedures, some pre-processing is done on the texts, such as removing "stop words" (e.g., "and" and "the") and "stemming" to remove the ends of conjugated verbs and plural nouns.

professor's lecture style (the estimated topic), but the words used to describe that style might differ by gender (the observed covariate).

Using covariates in the STM differs from the predictive models discussed later, such as supervised learning in that covariates might influence a topic, but the model does not force them to be connected. This helps to alleviate concerns that relationships are "baked in" to a conclusion by incorporating metadata. Rather covariates are best thought of as a way of defining subsets of the data (by age, gender, location, etc.) that *may* have similar patterns of topic use. In a separate work we discuss the details of the STM as well as provide simulation evidence showing its ability to uncover topic/covariate relationships (Roberts et al., 2014).

STM offers several improvements over simply performing LDA and then regressing the results on metadata of interest. First, the STM explicitly models topic correlation whereas LDA does not. Second, the STM allows for topic vocabulary — not just topic prevalence — to vary by covariates. Third, Roberts and colleagues (2014) have shown through simulations that LDA can miss important covariate relationships in the data. Finally, by explicitly including the covariate relationships in the model, we are able to include measurement uncertainty from the estimation of the latent topics into our regression analyses. In theory, as the number of documents grows extremely large, LDA will correctly recover the covariate relationships. In practice, there is evidence that even with extremely large document sets, STM yields better results both in terms of predictive power and qualitative interpretability (Roberts et al., 2014; Roberts, Stewart, & Airoldi, 2014.)

In what follows, we provide three case studies of the use of STM for computer-assisted reading in surveys, forums, and course evaluations. The three case studies have several features in common that make them well suited for analysis with the STM. First, they all involve large corpora of texts that cannot be read in a reasonable timeframe by educators, motivating the use of computation to organize the texts. Second, they all involve domains where useful classifications of the text are not known *a priori*, and we can use unsupervised learning models to uncover undiscovered patterns. Finally, the three corpora have documents associated with useful metadata about the author or the document, and the inclusion of these metadata in the STM can improve the estimation of the model and reveal how themes in student writing vary across substantively interesting subgroups.

3 COMPUTER-ASSISTED READING OF QUALITATIVE PRE-COURSE SURVEY RESPONSES WITH THE STRUCTURAL TOPIC MODEL

We now illustrate the STM workflow and results by analyzing data about rationales for signing up for a MOOC platform. One of the largest providers of MOOCs is edX, and when students register on the edX site (a prerequisite for registering for any individual edX course) they are given a short survey including the free-response question: "Please share with us your reasons for registering with edX." While other MOOC platforms and specific courses ask students about their motivations using a variety of fixed

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

response items (Breslow et al., 2013; Koller, Ng, Do, & Chen, 2013; Wang & Baker, 2014), this is, to our knowledge, one of the largest data sets of unstructured text where students describe their motivations for participating in online learning experiences in their own words. These data speak directly to questions about student motivation that have come to the fore as open online learning opportunities have grown dramatically (Computing Research Association, 2013). The STM allows researchers to analyze how registrants describe their motivations in their own words.

3.1 Analytic Workflow with Structural Topic Models

In this and the following examples, we follow five analytical steps in using the Structural Topic Model. First, we prepare the data by including the corpus of documents and relevant metadata/covariates into the software package. Second, we run the software package and produce a standard set of results and visualizations, including the list of topics in descending order of prevalence, the key words for each topic, and highly associated "archetypal" documents for each topic. Human analysis then guides the third step: we assign descriptive labels to each of the computer-generated topics by evaluating key words from each topic and archetypal documents within each topic. Fourth, we examine how topic distributions vary according to covariates of interest, to understand how important subgroups differ in their written responses. Finally, when appropriate, we propose a call to action based on the findings from the STM, which might be a pedagogical intervention or an experimental study.

In this first example, we estimate a Structural Topic Model with twelve topics examining student rationales for signing up for the edX MOOC platform. We examine the universe of all responses from edX users who by 4 August 2013 had registered for one of the first six HarvardX courses: Intro to Computer Science, Justice, The Ancient Greek Hero, Health in Numbers, Human Health and Global Environmental Change, and Copyright (Ho et al., 2014). This totals nearly a quarter million responses (240,208), highlighting the need for computer-assisted reading. In the model we include several covariates: indicator variables for each course, the respondent's education level, an indicator variable for whether male, and a continuous age variable. After a small amount of automatic pre-processing, estimation of this model is done with a single line of code:

```
storage<-stm(docs,vocab,K=12,
     prevalence=~course+educlevel+male+s(age),data=meta)
```

Notice the simplicity of the code syntax; one design principle of the software package is that the level of programming sophistication for conducting these analyses should be approximately equivalent to running a regression model in a typical statistical package such as R, SAS, or SPSS. In this example, we estimate 12 topics (K) and the prevalence of each topic is modelled as a function of the course a person signed up for, their education level (treated as a factor variable), gender, and age (allowed to have a non-linear effect via a spline (*s() function*). The results are stored in an object we label "storage." A complete vignette describing all features of the `stm` package (Roberts, Stewart, & Tingley, 2014), and how to use all of these features, is available at http://structuraltopicmodel.com.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

In parametric topic models such this one, it is necessary to set the number of topics. In the computer science literature it is common to set the number of topics by maximizing the predictive power of the model on a heldout sample (Wallach, Murray, Salakhutdinov, & Mimno, 2009). However, the most predictive model may not always be the most qualitatively interpretable or substantively useful model (Chang, Gerrish, Wang, Boyd-Graber, & Blei, 2009). Instead we see the choice of the number of topics as a decision to be made on substantive grounds that reflect the desired granularity of the summary of the corpus (Grimmer and Stewart 2013). When analyzing corpora covering a very broad array of material, higher numbers of topics may be more appropriate. We selected a K of 12 in this example, though we also estimated this model with other values of K for comparison, and the topics of highest interest appear in our alternate models. For those who want to use a data-driven approach, the `stm` package includes tools for calculating heldout likelihood as well as several other popular automated metrics (Mimno, Wallach, Talley, Leenders, & McCallum, 2011).

3.2 Findings from STM Analysis of Student Motivation Registrations

Using the STM, we uncover the most common themes in student self-reported motivations for registering for edX, and in Figure 1, we present representative output from the `stm` R package.

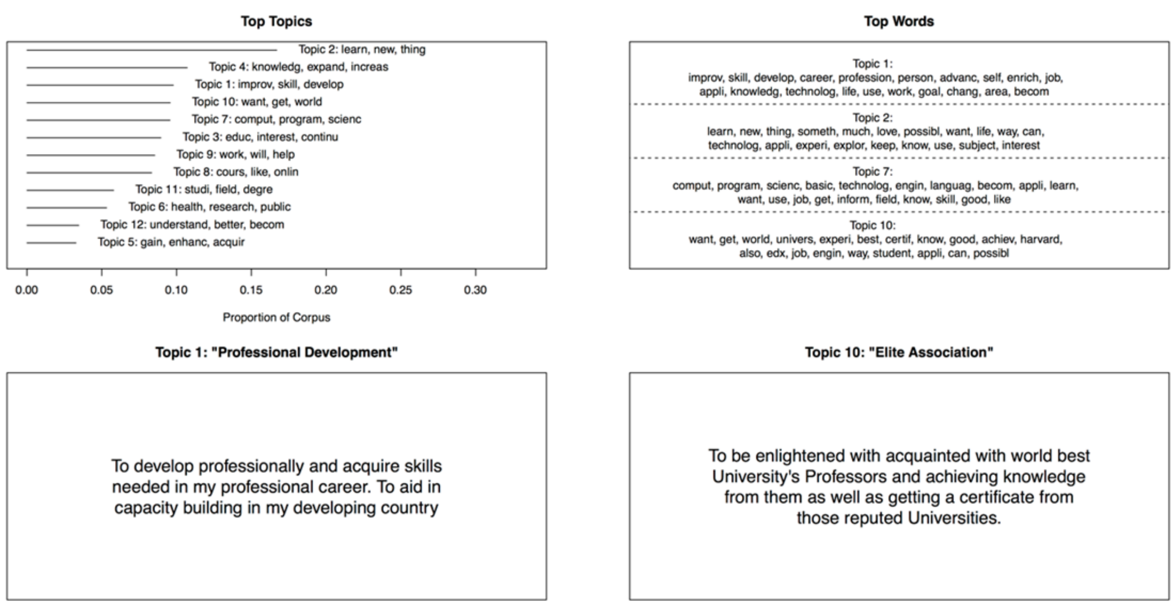


Figure 1. Representative output from a 12-topic Structural Topic Model analysis conducted in the R `stm` package from a corpus of 240,208 student responses describing their motivations for registering for edX. The top left panel shows the proportion of the corpus associated with each of the twelve topics, and three key words for each topic. The top left panel gives the twenty most probable words for four selected topics: Topic 1: "Professional Development"; Topic 2: "Lifelong Learning"; Topic 7: "Computer Science and Programming"; and Topic 10: "Elite Association." The bottom panels show highly associated texts from Topic 1 and Topic 10.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

We use two sources of data to parse the semantic meaning of these topics: 1) the word tokens mostly highly associated with each topic and 2) exemplar texts. In the top left of Figure 1, we display the 20 word stems mostly highly associated with several sample topics in their order of prevalence within the topic. In the top right panel, we show the distribution of all 12 topics across all documents. On the x-axis, we show the proportion of topic prevalence across all documents. For example, Topic 2 is the most prevalent in the corpus; it concerns learning new things and lifelong learning. Topic 12, which includes developing specific skills for job environments, is least prevalent. The system automatically produces the topic number and the three word tokens that most distinctly represent the topic. In the bottom left and right of Figure 1 we show for Topic 1 and Topic 10 the documents with the highest estimated proportion of topic-related words and constructs. Using these word frequency tables and exemplar texts, we then attach semantically meaningful descriptive labels for several topics: Topic 1: “Professional Development”; Topic 9: “Lifelong Learning”; Topic 7: “Computer Science and Programming”; and Topic 10: “Elite Association.” The appendix provides a complete depiction of all topics (in Figure 11).

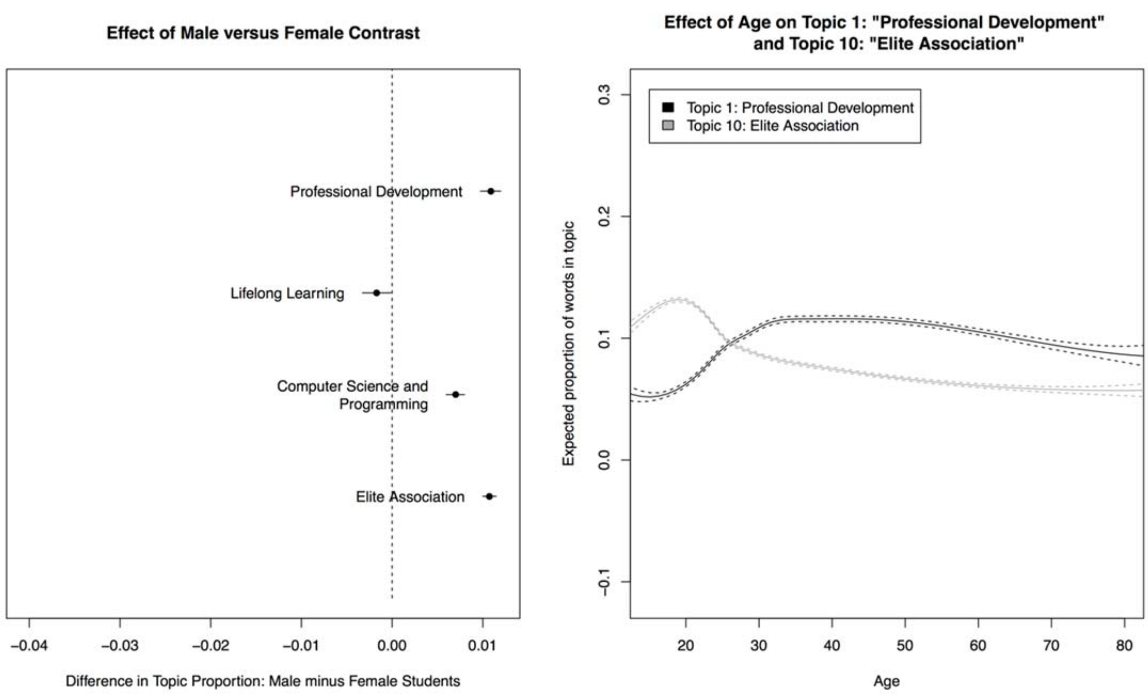


Figure 2. Results from STM analysis of a corpus of 240,000 documents describing student motivations for registering for edX. The left panel sorts four sample topics (Topic 1: “Professional Development”; Topic 2: “Computer Science and Programming”; Topic 9: “Lifelong Learning”; and Topic 10: “Elite Association”) by their respective use by males relative to females. Positive numbers indicate that males were more likely to write about the topic. The right panel shows the effect of age on usage of two sample topics. For example, younger MOOC registrants appear more motivated by elite association than career development, whereas older MOOC registrants are more likely to write about career development.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

The topic model reveals both predictable and surprising student motivations. Topic 1, for instance, uncovers students who describe their motivations as instrumental and professional in nature; they register for MOOCs to advance their careers. Given the practical nature of several of the early HarvardX courses, like Health in Numbers (biostatistics and epidemiology) and Computer Science, this topic is to be expected. Topic 10 shows the importance of associating with a leading university. This is one of the most commonly expressed reasons for wanting to sign up for a MOOC, and echoes the edX marketing language that the platform offers “the best courses, from the best professors, from the best universities.” This suggests that this element of elite branding is front-of-mind for many participants when signing up for edX. While in retrospect the importance of this topic makes sense, this dimension of elite affinity does not always appear in other surveys of student motivations in the MOOC literature. The topic model here uncovers an important dimension of student motivation that is presently under-researched. These findings informed the design of the 2013–2014 HarvardX and MITx pre-course surveys, which now ask students about the importance of career advancement and elite affiliation in their registration decisions.

After examining the topics themselves, we go on to examine how these motivations differ across substantively interesting sub-populations, which we can do with both dichotomous and continuous measures. In Figure 2, we show how the prevalence of topics differs by gender and age. In the left panel, we plot the difference in the expected proportion of words within the topic for men minus the expected proportion of words within the topic for women. When calculating effect, the values of all other variables are set at their sample median values.[3] The lines give 95% confidence intervals on the difference including measurement uncertainty. Positive numbers indicate that males were more likely to write about the topic. For instance, Topic 10 describes the desire to acquire computer science knowledge, and it was the topic most heavily correlated with the respondent being male. We would predict that documents produced by men would have more words and constructs related to computer science than documents produced by women; on average, the proportion of words in a document from this topic will be .006 more for men. While small, given our large sample size, this is statistically significant. Computer science MOOCs disproportionately enrol men — the student body of the fall 2012 HarvardX Introduction to Computer Science course was 79% men — and this unsurprising result gives us greater confidence in the performance of the model. The left-hand plots can be produced with two lines of code, the first to calculate the necessary quantities, and the second to produce a plot:

```
prep <-estimateEffect(c(1,2,7,10) ~ course+educlevel+male+s(age),
storage, meta=meta)

plot.estimateEffect(prep, "male", model="z", method="difference",
```

[3] We note that the STM model supports a broad array of specifications, including interactions between variables and allowing for non-linear effects through the use of splines. Here we do not include interactions but we do allow the effect of age to be non-linear.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

```
cov.value1=1,cov.value2=0,
xlab="Difference in Topic Proportion: Male-Female",
main="Effect of Male versus Female Contrast", verbose.labels=F,
topics=c(1,2,7,10), labeltype="custom",
custom.labels=c("Professional Development","Lifelong Learning"
,"Computer Science and Programming","Elite Association"))
```

Next, we show how to analyze the influence of a continuous covariate: the age of a student. To analyze a continuous covariate, we plot the predicted proportion of a document that comprises a topic as a function of age. To do this requires a single line of code:

```
plot.estimateEffect(prep, "age", model="z",
    main="Effect of Age on Topics 1:
    Professional Development \n and 10: Elite Association",
    method="continuous", topic=c(1,10),
    xlab="Age",ylab="Expected proportion of words in topic")
```

In the right-hand panel of Figure 2, we show the effect of age on two topics in the corpus, Topic 1: “Professional Development” and Topic 10: “Elite Association.” On the X-axis we show age, and on the Y-axis, we show the expected proportion of words in respondents’ text that come from a particular topic. The dashed lines provide 95% confidence intervals. The importance of allowing a non-linear relationship with age is quite apparent. Among this cross-sectional cohort, younger MOOC registrants appear more motivated by elite association than career development, whereas older MOOC registrants are more likely to write about career development.[4] These findings could inform recruitment efforts by universities and MOOC providers, and they suggest possibilities for segmenting marketing, where students of different ages receive recruitment materials highlighting different themes. These findings could also inform the design of more personalized learning environments; for example, by providing older students with more examples from industrial or commercial contexts.

To review, we analyzed nearly a quarter-million statements about why an individual was signing up to take a MOOC, and we used the STM to identify a set of syntactically related topics that represented substantively interesting response patterns. Evaluating all quarter-million documents would be infeasible without computer assistance, and even hand-coding methods could require coding a random sample of thousands of documents to uncover more rare topics. We then show how these distributions vary by age and gender; the STM is unique in its utilization of covariate information in this way and as such holds promise for use across educational data.

There are two caveats to using this method. First, the method organizes data for human analysis, uncovering patterns across large amounts of text, but ultimately the utility of these clusters depends

[4] We retained and used in estimation cases where individuals list low or very high ages. Results in these regions should probably be taken cautiously. Our ability to use a spline function means they are not influential outliers.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

upon thoughtful and necessarily subjective assignment of meaning to these clusters. Second, unsupervised topic modelling can uncover clusters of interest, but it cannot assign documents to a pre-defined taxonomy or assess how the distribution of documents fits into an ideal distribution. Supervised methods are more appropriate for these purposes.

4 ADDITIONAL EXAMPLES OF STRUCTURAL TOPIC MODELLING IN MASSIVE OPEN ONLINE COURSES

In this section, we examine the application of the STM to student discussion forums posts and course evaluations. Both examples come from a HarvardX course, ChinaX,[5] conducted on the edX platform in the fall of 2013. When finished, ChinaX will span nearly 15 months of course content offered in ten parts, with separate final grades and certificates for each part. The course was taught by Professors Peter Bol and William Kirby, and Part I explored the political and intellectual foundations of China. Part I launched on 31 October 2013 and finished on 23 December 2013 with 33,479 students registering for the course and over 2,000 students earning a certificate of completion. As with many online courses in the humanities, contributions to the forums are required to earn a certificate. Students are also asked to complete course evaluations and to provide open-ended feedback that can help the instructional team to make improvements in the course. We demonstrate how the STM helped the ChinaX instructional team see the broad themes in discussion forums and survey feedback and begin to use that feedback to iteratively improve subsequent sections of the course.

In each example, we continue to provide graphs of various key quantities from the STM model. To review: *topic words* give information about the top words used in each topic, listed in order of their weight in that topic, to understand the general language of a topic. *Highly associated texts* are specific examples of documents characteristic of a particular topic. In combination with the top words in each topic, they allow researchers to understand specific instances of topic usage, supplementing the overall analysis with example documents. *Covariate relationships* describe the relationship between topic usage and covariates values. *Topic prevalence* gives the relative usage of the topic across the corpus. This allows the user to find the overall themes of the corpus with relative weight of the different themes in the discussion. All analysis is done using the free, open-source R package `stm` that features a rich set of functions requiring very minimal programming knowledge of the user, equivalent to an understand of basic data entry and statistical testing in SPSS, SAS, or Stata.

4.1 Discussion Forums

Discussion forums in MOOCS provide an opportunity for students to develop and demonstrate their understanding, to ask questions, and to interact with each other and with course teaching staff. Literature on the use of forum-based learning in online education indicates that online forums can be

[5] https://www.edx.org/course/harvardx/harvardx-sw12x-china-920

important for their ability to facilitate debate, networking, and interaction with instructional staff (Mak, Williams, & Mackness, 2010). Furthermore, online discussion forums can increase individual participation and in the best instances promote a collaborative learning environment that allows for high-level critical thinking (Thomas, 2002). The best online discussion forums are facilitated by instructional staff, but in MOOCs and other large-scale learning environments, students can generate text at a pace that far exceeds an instructional team's capacity to read.

One solution to these problems has been the use of voting to promote comments of importance to the attention of students and the instructor. By design, these comments promote a biased representation of the course, where popular comments become more visible. These comments can present a skewed version of the course to the instructors, often times drawing their attention to the writings of first posters or most enthusiastic contributors rather than the writings of more typical students. In these circumstances, the use of computer-assisted reading can enable the teaching staff of a MOOC or other online-discussion-intensive course to review the forums and gain an overall understanding of the discussion topics and specific, representative points (Dringus & Ellis, 2005). The STM provides exactly this functionality. The inclusion of covariates within a topic model allows data such as up-votes to factor into a model. The STM model allows the instructor to observe general trends in discussion topics, view particular posts that capture the essence of the topics, and understand the correlation between topic usage and votes within the forum. This lets the instructor observe what kinds of conversations are provoked by the instructional content of a lesson or unit.

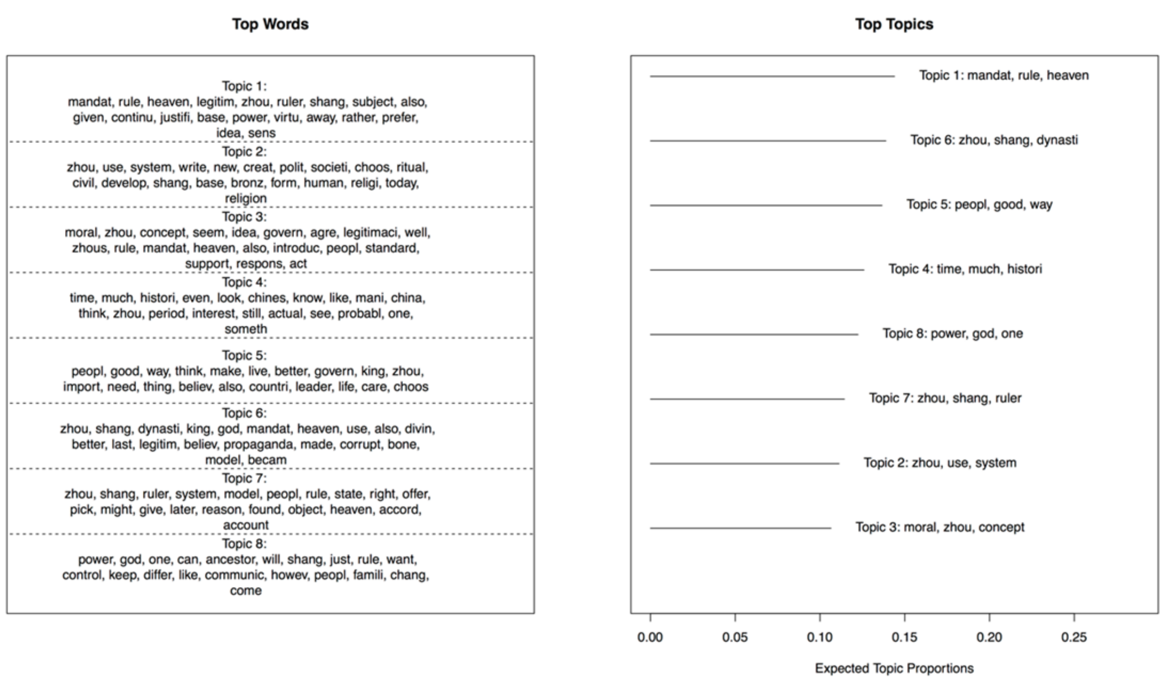


Figure 3. Results from an 8-topic STM analysis of a corpus of discussion forum posts from ChinaX examining the Zhou and Shang dynasties in China. The left panel shows key words associated with the eight topics, and the right panel shows the distribution of topics.

In the following analysis, we explore text data from a pair of discussion threads, each with more than 1,600 posts, from Part I of ChinaX. The course explores multiple dimensions of ancient Chinese history and culture, but the reader need not have a deep background in Chinese history to begin to understand our analyses. The STM provides semantically coherent topics with example posts related to each topic, which allows a non-specialist to begin understand the discussion themes. In modelling these two corpora of texts, we also include as a covariate whether or not each post was up-voted. Figure 3 is the general result of the STM modelling forum posts where students were asked to compare the Zhou and Shang dynasties for governing style: "Although both Shang and Zhou are at the beginning of China's history, later dynasties would look to Zhou as the model of civilization, rather than Shang. What did Zhou offer that Shang did not? What did Shang have that later people might have objected to?" As in the previous example, in Figure 3, the top left panel presents the top words in each topic, while the top right shows their respective prevalence throughout the corpus.

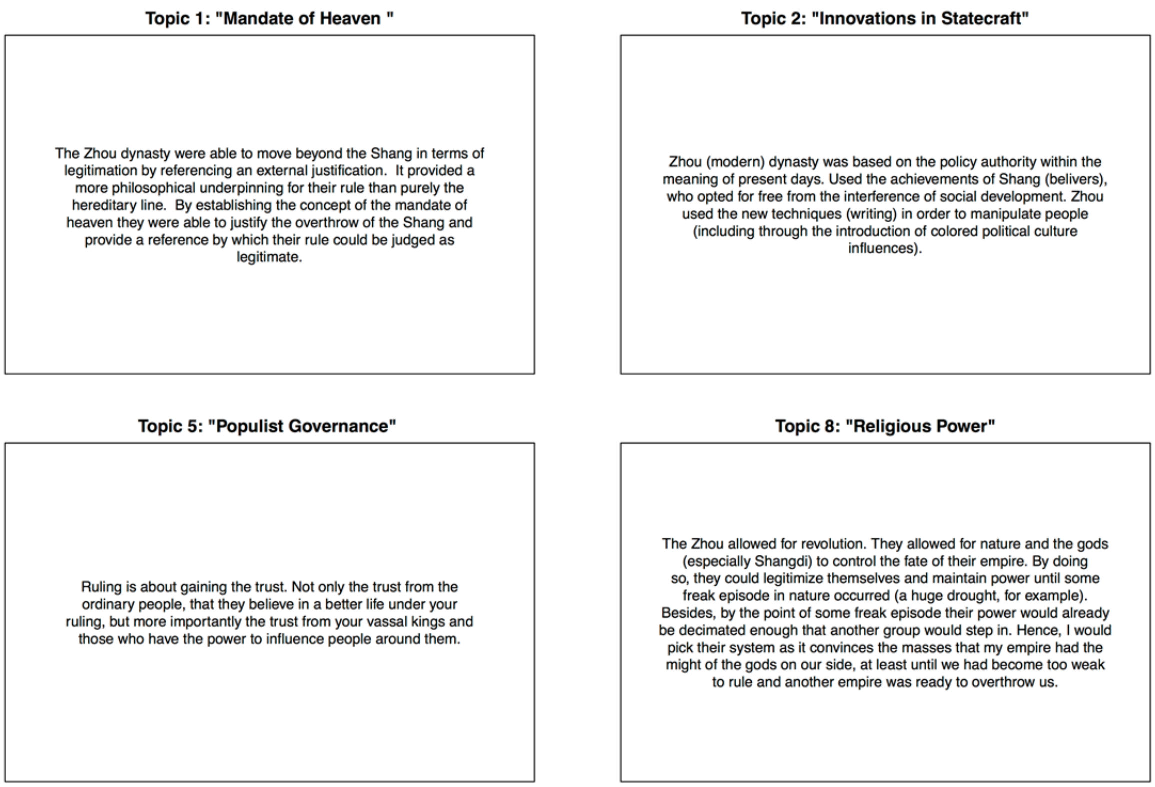


Figure 4. Example forum posts on Shang and Zhou for Topic 1: "Mandate of Heaven"; Topic 2: "Innovations in Statecraft"; Topic 5: "Populist Governance"; and Topic 8: "Religious Power."

Appropriately, the most prevalent topics in the corpus have to do with issues of religious and popular legitimacy. One of the main ideological innovations of the Zhou dynasty was the notion of the "Mandate of Heaven": that rulers maintain their position by the grace of the gods, and dishonorable rulers can have their mandate revoked (as "occurred" when the Zhou violently overthrew the Shang). Virtuous

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

rulers maintain the support of their people and earn the favour of the gods. These ideas are critical to the art and writing of the Zhou period, and Topics 1, 6, 5, and 8 are clearly related to these key issues.

When comparing topics, the `stm` package can produce visual depictions of the difference between topical themes, as shown in Figure 5, to see overlap as well as subtle differences in similar topics. Here we have compared Topics 5 and 8, to show some of the differences in language between discussions of religious power ("Religious Power") from those of populist issues ("Populist Governance").

Figure 5. Contrast in words between Topic 8: "Populist Governance" and Topic 5: "Religious Power." Words further to the outside are more heavily associated with a single topic. Larger words have a greater weight within that topic. The dashed line represents the split between words that have greater weight on Topic 8 versus Topic 5.

While these intellectual advances are important to understanding the period, the content in this unit of ChinaX also explored how these new ideologies required innovations in statecraft to spread. During this period, rulers found new uses for writing, bronze vessels, religious rituals, and public speeches to earn and maintain legitimacy in the eyes of the people. For instance, many religions functions that had been relatively apolitical in the Shang became more politicized in the Zhou. Only one topic, and one of the least prevalent topics, coheres with these ideas about advances in statecraft: Topic 2, which included words like "system," "write," "new," and "bronze."

The STM model shows that across the hundreds of posts and comments in the thread, students wrote more about the ideas that the Zhou developed to assert their legitimacy rather than the emerging statecraft methods they used to spread these ideas. This provides the instructional team with a qualitative reflection on their teaching that would be difficult to obtain without reading a substantial sample of the discussion forum. If both the ideas and the statecraft are critical to their interpretation of this period of history, then these analyses suggest that in a revision of the course content, the instructional team might consider revising the unit to place more emphasis on these ideas that students wrote less about. In this MOOC context, the STM gives faculty rapid feedback on how students are reacting and responding to instructional materials.

The forum analyses can be enhanced by including covariates, such as whether a post was up-voted by a peer in the forums. Figure 6 summarizes a set of posts in which students were asked to compare principal thinkers across Chinese history from Confucius through the era of the "One Hundred Schools." The top left graph gives the top words in several topics that we discuss below, while the top right graph gives us the respective topic weights in the corpus. In Figure 7, we show specific examples characteristic of four topics discussed below. As with the previous example, we can use these topics and word lists to review how students interpreted course materials. For instance, we are pleased to find that students discuss Han Fei and associate him with the law, as Han Fei's development of Legalist theory is an important intellectual advancement from the era. We also see that many students write responses that include comparisons of multiple thinkers. Topic 2, for instance, includes comparisons of Laozi and Mencius. Interestingly, the most prevalent topic, Topic 4: "Vague Language," includes no word tokens referencing any specific philosophers or philosophies. These distributions tell us what students wrote about; next we turn to an examination of which topics were most likely to be part of a post that received an up-vote, which provides one measure of how other students responded to these posts.

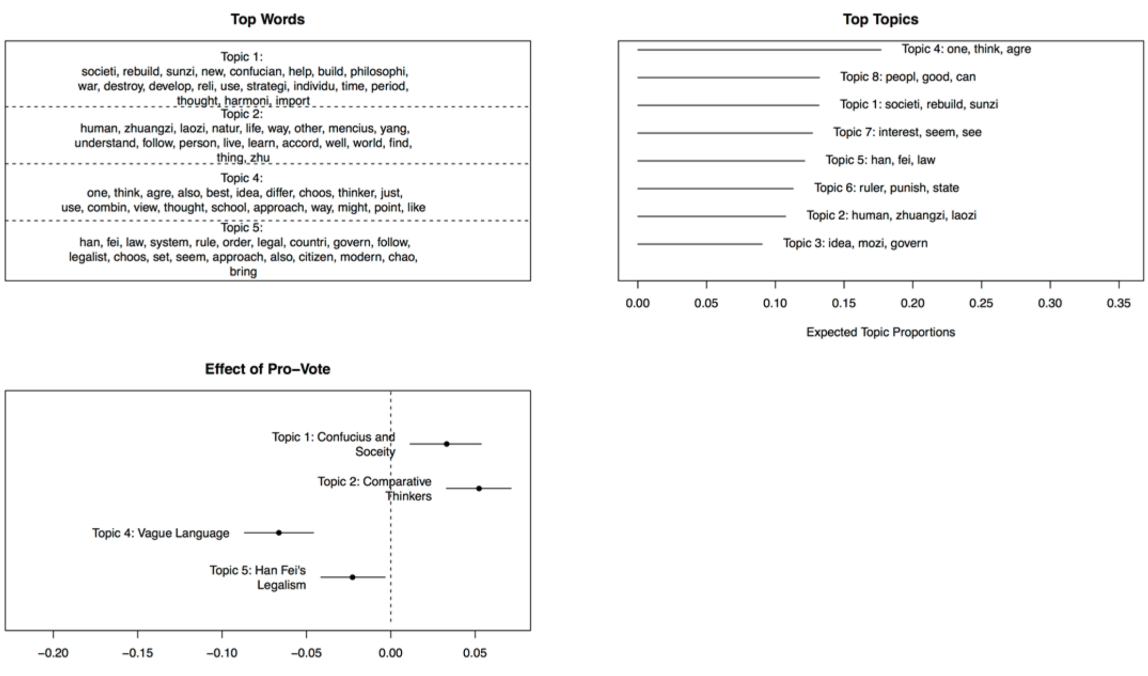


Figure 6. Output from the STM analysis of a discussion forum with 1,715 posts concerning ancient Chinese philosophers. Words associated with Topic 1: "Confucius and Society"; Topic 2: "Comparative Thinkers"; Topic 4 "Vague Language"; and Topic 5: "Han Fei's Legalism" appear in the top left panel. The top right panel shows the frequency distribution of topics the across the corpus. The bottom left panel shows the effect of topic usage on receiving at least one up-vote.

Topic 2: "Comparative Thinkers"

If that is the scenarios, I would refer to allow nature to take its course and assume the state of the Dao De Jing, which is the thought and rule written about by Lao Zi. Nature is the original state of being and human had been altering them as we deem that we are the master of this world, we forgot that Dao follows nature and we human cannot go against the flow of nature. In fact, we should learn to flow with nature and live a natural life, in an effortless and simplistic form. It is in this desireless form that one could truely explore one's potential and live life to the fullest. As Lao Zi has pointed out in his text, human floows the rule of the land, and the land follows the rule of the cosmo, and the cosmo follows the rule of the Dao and the Dao follows the rule of nature (...One should fulfil the stomach and reduce one's desire, as desire in its many forms are the roots of all evils and mishaps, returning to the basic life would be the way of returning to the Dao of nature.

Yang Zhu most closely resembles a path I am following now. After living a life that always emphasized others' needs first or at least others' opinions of what I should do with my life first, I am now discovering what my own natural inclinations are in terms of personal joy. If I explore and am aware of all the ways I experience my highest excitement and happiness in life, and we all did that, we may be able to contribute more to society than we could ever have imagined before. However, I do appreciate Han Fei's legal protections that make lifestyle and country safe from attack and exploitation.

Topic 4: "Vague Language"

In reflecting on the different schools I was somewhat surprised to find myself most drawn to Zhuangzi. However in practice relativism and "dropping out" would not be entirely practical. I agree with the main point of this thread – that no single school offers an ideal approach, rather picking aspects of all of the schools and applying them to different aspects of government is the best approach.

I cannot choose one theory or the other simply because I do not agree completely with one or another. I would use a combination of opinions and theories and apply different combinations in different countries and situations. However, I found these ideas fascinating and extremely interesting and I am very satisfied that I have participated at this course.

Topic 5: "Han Fei's Legalism"

I also admire Han Fei because he saw that a legal structure would be the best protection for all the citizens of his country. It's too bad he was asked to commit suicide by a political rival who happened to be the jealous chief minister to the king. Han Fei obeyed by drinking poison while he was imprisoned.

I'd counter with a bit of warning drawn from the history of the Qin Dynasty and Han Fei himself. This dynasty, which was built following teachings of Han Fei and other Legalists, died with its first and only Emperor. It could not achieve a transition to a new ruler. On a personal level: Li Si, the Qin prime minister and fellow Legalist, was involved in orchestrating the violent death of Han Fei (and died similarly some 25 years later).

Figure 7. Example forum posts from ChinaX comparing ancient Chinese Philosophers from four topics: Topic 1: "Confucius and Society"; Topic 2: "Comparative Thinkers"; Topic 4: "Vague Language"; and Topic 5: "Han Fei's Legalism."

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

The bottom left panel of Figure 6 shows us the relationship between topic usage and receiving at least one up-vote in the forums. Notice that Topic 2 ("Comparative Thinkers") is most associated with having at least one up-vote, and in contrast, Topic 4 ("Vague Language") was most negatively associated with having an up-vote; students did not up-vote comments that reflected indecisiveness or a lack of specific evidence. Interestingly Topic 5 ("Han Fei and Legalism") was also less likely to be up-voted. In a deeper analysis of these posts, it would be worth exploring whether these posts were less likely to be up-voted because they were poorly written or had some problem in argumentation, or if Han Fei's Legalism simply proved unpopular with a modern audience.

This analysis motivates several kinds of experiments in sharing these findings directly with students as a way of offering feedback on participation in the discussion forums with concrete examples. For instance, instructors could show students the evidence that vague responses received few up-votes while responses with specific ideas and evidence received more, and then show exemplars of each. This kind of feedback on the specific characteristics of high-quality responses in a forum would hopefully help students write better responses. Ideally, faculty would test these ideas by showing the feedback to a random subsample of students to determine if this kind of feedback helped some improve their writing. We could also show students the evidence of their own leanings and biases, for instance by demonstrating the community's low regard for Han Fei. We believe this kind of computer-assisted reading and real-time display of textual data could play an important role in providing students feedback on discussion forums too large for rapid human analysis.

In just a few graphs, the STM offers a thematic overview of the student responses from two entire forums worth of data. We have the ability to see the topics by frequency and word choice and to dive into those topics by looking at archetypal posts. This allows us to monitor student understanding of topics to prevent gaps in knowledge or misunderstanding while also understanding the discussion in terms of broad themes. These results can easily be incorporated into later coursework or when retrospectively evaluating the success of a course. At present, one of the most challenging aspects of incorporating forums in MOOCs is that they rapidly become far too extensive for any student or faculty member to follow, and the STM offers a toolset for finding patterns amid these wide-ranging conversations.

4.2 Class Feedback

Research suggests that faculty who reflectively incorporate feedback from student evaluations improve their teaching, as measured by subsequent evaluations (Winchester & Winchester, 2014). In small-scale teaching environments, it is possible to read and analyze an entire set of student evaluations that include qualitative feedback. But as class sizes grow, especially in large-scale online learning environments, reading thousands of open-ended student responses becomes logistically infeasible. We present a strategy that allows faculty to ask for rich qualitative feedback from many thousands of students and then use computer-assisted reading methods to find patterns of feedback and to uncover typical suggestions or concerns.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

At the close of the first eight-week mini-course of ChinaX, participants were invited to complete a course evaluation, which combined both open- and fixed-response answers. We have responses from 1,057 students for which we have complete covariate information. This represents approximately 2.8% of all registrants of the course and 42.3% of all who explored over half of the units of the course. Students who completed the survey were overwhelmingly those who persisted throughout the entire course; 79.9% of survey respondents earned a certificate in the course compared to 5.4% of all registrants. Our findings here describe how a subset of successful students evaluated the course.

Students were asked to articulate their feedback on the course in two open response questions:

- What were your favourite aspects of this mini-course so far?
- What could the ChinaX team do to improve your learning experience?

Students were asked additional fixed responses questions such as "Overall, how satisfied were you with this mini-course?" with response anchors ranging from "Very Dissatisfied" to "Very Satisfied." In analyzing and triangulating responses from these questions, we can provide a nuanced picture of how substantively interesting subgroups evaluate the course.

Figure 8 shows the results from a seven-topic STM of student responses to their favourite aspects of the course for several selected topics.[6] The most prevalent topic (Topic 3) connects to the relationship with the course faculty and guest lectures and specifically references "office hours." The ChinaX instructional team and faculty created video "office hours" every other week during the course, which were filmed as the course progressed, in contrast to the rest of the content, which was prepared well in advance. In these office hours, the course faculty provided a response to the discussion forums, highlighted ideas from the previous week, and showcased material to come. It was a time-intensive addition to the course, and the ChinaX teaching staff was gratified to see that students responded positively to the efforts. These findings help persuade the instructional team to continue investing effort into producing the office hours videos in subsequent weeks of the course. More broadly, the prevalence of the topic suggests that rapport with faculty is important even in these distance courses (Murphy & Rodriguez-Manzanares, 2012). The second most prevalent topic was about the course content, Topic 2, and includes specific reference to the short videos, most no longer than five minutes. This student feedback lends some evidence to the assertion that shorter videos may be a particularly appropriate content medium for the MOOC context (Guo, Kim, & Rubin, 2014).[7]

---

[6] In this STM, we include as topic prevalence covariates our satisfaction measure, levels of familiarity with the subject matter, reasons for taking the course, age, and gender.

[7] Other topics dealt with more substantive topics like Chinese culture and history, and more general comments about enjoying the class.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

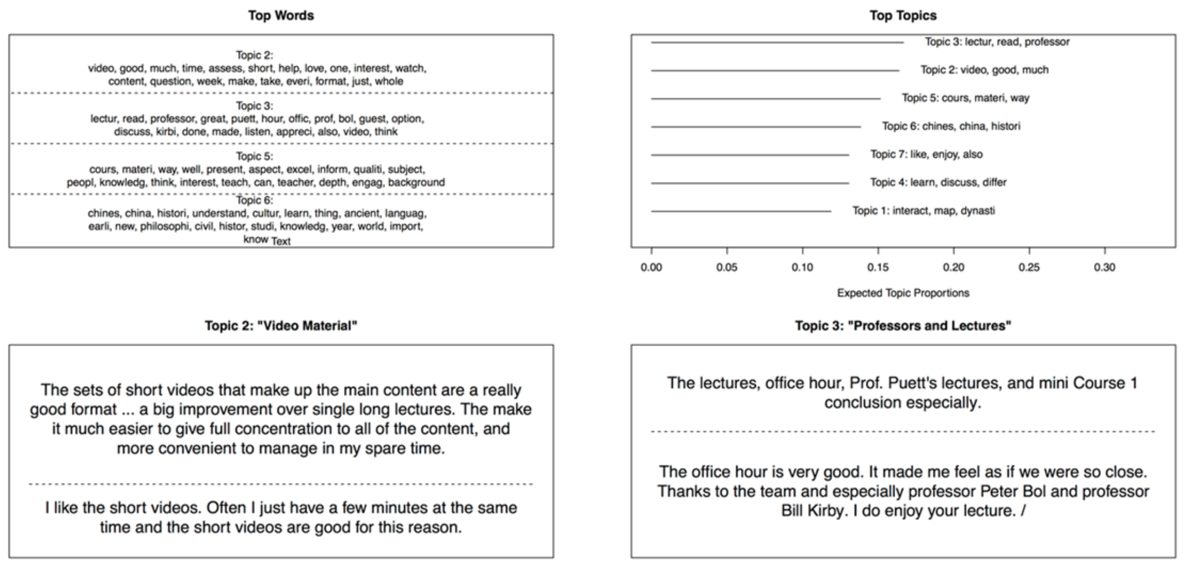


Figure 8. Output from 7 topic STM analysis of 1,057 responses to ChinaX course evaluation. The top left panel includes top words from four topics, Topic 2: "Video Material"; Topic 3: "Professors and Lectures"; Topic 5: "Course Materials"; and Topic 6: "Chinese History in Context." The top right panel includes the proportion of all seven topics across the corpus. The bottom panels have highly associated texts from Topics 2 and 3.

Figures 9 and 10 show the results of a ten-topic STM on the responses regarding what elements of the course could be improved. Here we focus our analysis on the influence of different levels of student satisfaction. Nearly all students who take the survey were either "Very" or "Extremely" satisfied with the course, so we are examining a narrow range of happy, successful students. Nonetheless, the differences in topic prevalence by student satisfaction are revealing. We see that Topic 4 ("Assessments") was associated with less satisfied students, and from the listing of the top words associated with this we recognize that these students took issue with the assessment and question format of the course. Our text samples indicate that these students felt that the questions were ambiguous or needlessly tricky. By contrast, more satisfied students raised issues with the discussion forum platform used by edX, Topic 2. These students complained about technical issues that prevented them from easily engaging with other students; they wanted to be able to participate even more fully!

Topic 7 includes references to Chinese characters, and it was the second most prevalent topic. While the course was ongoing, the ChinaX course team received several emails asking that more Chinese and English characters be displayed in the videos. The emails indicated that students were using the course to learn English or Chinese. These emails were sporadic and idiosyncratic, so the team debated whether to devote additional resources to adding Chinese characters to the videos. The prevalence of this topic in the course evaluation provided additional evidence that this was a widespread interest, and the course team decided to invest more resources in displaying more Chinese and English text in the videos

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

themselves. The computer-assisted reading of these course evaluations helped faculty to confirm the importance of this issue for learners and to improve this dimension of the course iteratively as it progressed.

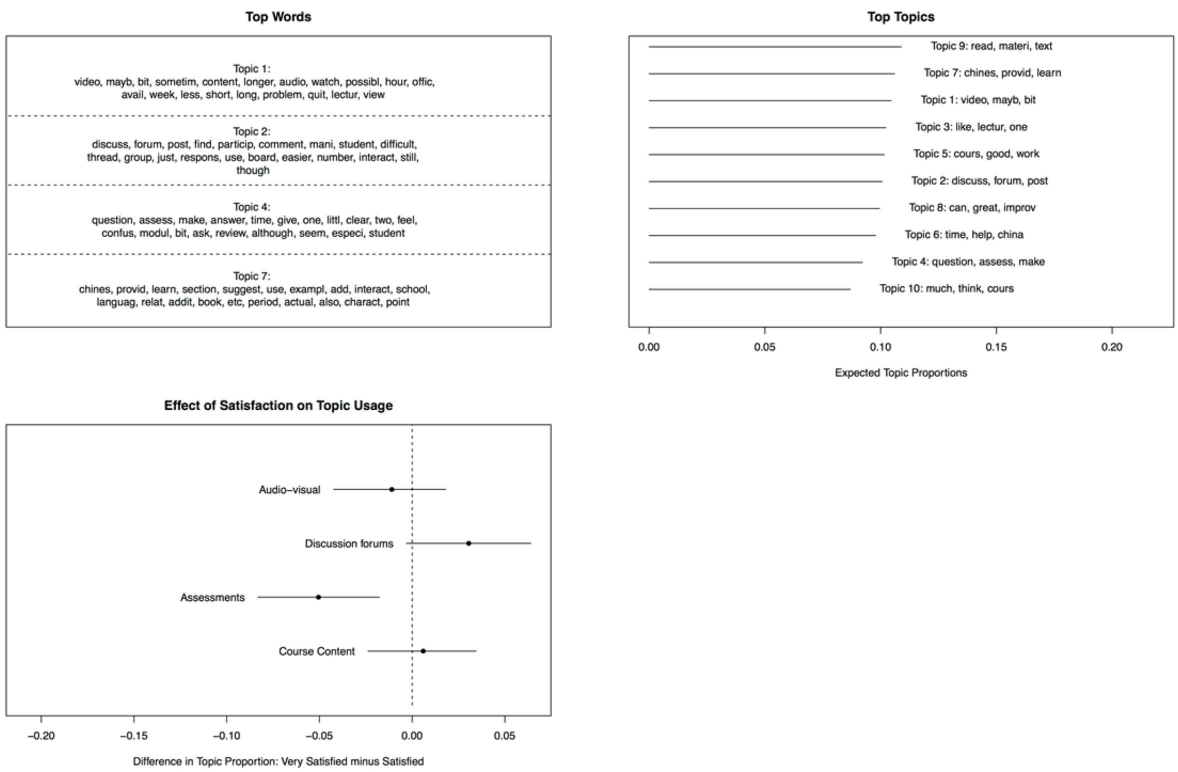


Figure 9. Output from 10-topic STM model examining 1,057 course evaluations concerning what could be improved in ChinaX. The top left panel includes top words from four sample topics: Topic 1: "Audio-visual"; Topic 2: "Discussion forums"; Topic 4: "Assessments"; and Topic 7: "Course Content." The top right panel shows the proportion of each topic across the corpus. The bottom left panel shows the effect of topic usage on being "Very Satisfied" versus "Extremely Satisfied."

In each of these examples, the STM successfully modelled responses to course evaluations, providing useful information about student learning and experience. While Likert-type items can gauge general levels of student satisfaction, effort, or learning, the rich data of open-ended responses give many more possibilities for characterizing the underlying reasons why students are satisfied or unsatisfied. In circumstances where the data are unwieldy to read, we have shown the ability of the STM to generalize the results of these responses while also incorporating quantitative factors such as student satisfaction. In this way, the STM enables instructors to use course evaluation in a meaningful way, even with short periods between iterations of a course, while lessening the time costs of reading volumes.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

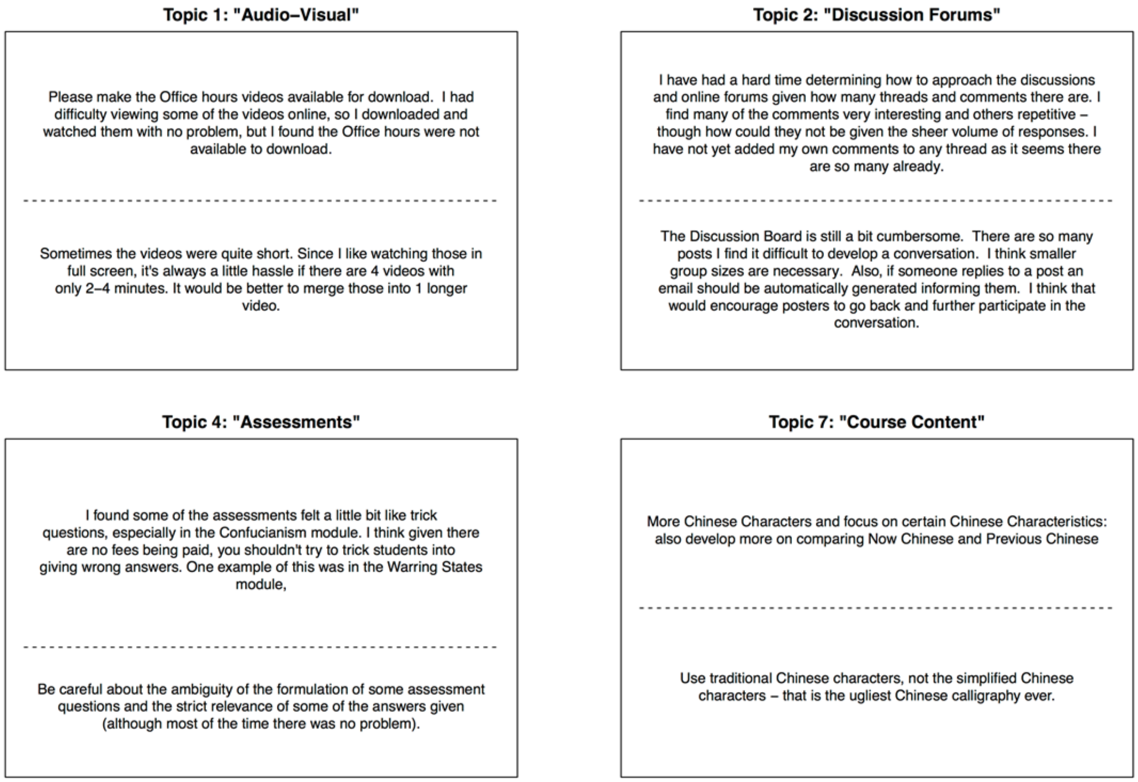


Figure 10. Students discuss what they found could be improved about the class, with examples from topics 1, 2, 4, and 7.

5 CONCLUSION

As MOOCs and other online learning environments expand in scale, the same data growth that proves overwhelming to faculty and instructional teams increases the reliability and utility of the STM. The STM becomes ever more useful in exactly the place where the human ability to process the entirety of student contributions in a timely fashion breaks down. These computer-assisted reading approaches have promising applications for helping faculty make sense of the vast conversations happening in MOOCs and large-scale learning environments. By way of conclusion, we offer three possible extensions of this work into domains beyond those discussed in this paper.

While the examples in this paper come from discussion forums, pre-course surveys, and course evaluations, there are also applications with student assessment. Much of the early research and development in MOOCs has focused on scalable mechanisms of assessing and assigning grades to individual student work. There have been important advances, but methods of peer grading and machine grading have proven controversial, technically challenging, and logistically difficult to implement (Piech et al., 2013; Rees, 2013). The supervised machine analysis of shorter pieces of student writing has proven particularly intractable (Brew & Leacock, 2013; Reich et al., 2014a). It may, therefore,

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

prove useful to complement these efforts at individual assessment of student learning with STM and other topic modelling approaches that attempt to assess student learning collectively. STM holds the promise of inviting students to submit their written work, knowing that each of their individual contributions will add to a model of student thinking that represents an entire learning community.

The examples in this paper also exclusively come from MOOCs built within learning management systems, where the focus of learning is on lecture videos and computationally graded assessments (sometimes called xMOOCs). STM technologies and methods also hold promise for connectivist learning environments (Downes, 2008), which emphasize the aggregation of student-produced text and media from sites across the open Web (sometimes called cMOOCs). STM approaches to analyzing these aggregated corpora offer connectivist educators a new set of tools to make aggregate meaning of the production of a network of learners. As we look towards a future of learning environments with larger networks of students, the most important technologies will not be those that facilitate dissemination of content from faculty, but those that allow educators to better understand the range and quality of contributions from students.

Finally, we have focused entirely on online environments, but there are promising applications for these tools in residential settings as well. In many large lecture courses, faculty use exit tickets or “mud cards” to have students articulate concepts that they are struggling with, and the STM could help characterize the topics and distributions of those challenges. Course evaluations are another promising domain for future research. Nearly every university uses some form of course evaluation to provide feedback to and evaluate instructors, but for reasons of expediency the analyses of these course evaluations is mostly limited to the quantitative elements. STM models open new possibilities for helping faculty, administrators, and instructional staff better understand not just how satisfied and engaged students were in a course, but how they qualitatively describe the strengths and weaknesses of particular courses.

Throughout higher education and across the disciplines, reasoning from evidence through writing is one of the central ways that students develop and demonstrate their understanding. Structural Topic Models and other unsupervised machine-learning methods are an important set of tools, complementary to peer grading and supervised machine learning techniques, to help instructors and educational researchers better understand students’ written contributions to learning communities and learning experiences.

REFERENCES

Anaya, A. R., & Boticario, J. G. (2011). Application of machine learning techniques to analyse student interactions and improve the collaboration process. *Expert Systems with Applications*, *38*(2), 1171–1181. doi:10.1016/j.eswa.2010.05.010

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. doi:10.1145/2133806.2133826

Blei, D. M., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. Retrieved from http://machinelearning.wustl.edu/mlpapers/paper_files/BleiNJ03.pdf

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York, NY: Springer.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., & Ho, A. D. (2013). Studying learning in the worldwide classroom: Research into EdX's first MOOC. *Research & Practice in Assessment, 8(1),* 13–25. http://www.rpajournal.com/dev/wp-content/uploads/2013/05/SF2.pdf

Brew, C., & Leacock, C. (2013). Automated short answer scoring. In M. Shermis & J. Burnstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 136–153). New York, NY: Routledge.

Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z., & Wong, F. M. F. (2013). Learning about social learning in MOOCs: From statistical analysis to generative model. *arXiv preprint arXiv:1312.2159*.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Proceedings of Advances in Neural Information Processing Systems 22 NIPS (2009)*.http://papers.nips.cc/paper/3700-reading-tea-leaves-how-humans-interpret-topic-models Computing Research Association. (2013). *New technology-based models for postsecondary learning: Conceptual frameworks and research agenda.* Retrieved from http://cra.org/uploads/documents/resources/rissues/Postseconday_Learning_NSF-CRA_report.pdf

Downes, S. (2008). Places to go: Connectivism & connective knowledge. *Innovate: Journal of Online Education*, *5*(1). http://bsili.3csn.org/files/2010/06/Places_to_Go-__Connectivism__Connective_Knowledge.pdf

Dringus, L. P., & Ellis, T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education*, *45*(1), 141–160. doi:10.1016/j.compedu.2004.05.003

Duwairi, Rehab M. (2006). A framework for the computerized assessment of university student essays. *Computers in Human Behavior, 22*(3), 381–388.

Eggers, A., & Spirling, A. (2011). *Partisan convergence in executive-legislative interactions modeling debates in the House of Commons, 1832–1915.* Retrieved from Harvard Institute for Quantitative Social Science website: 202.154.59.182/mfile/files/Jurnal/Jurnal%202011/Partisan%20Convergence%20in%20Executive-Legislative%20Interactions%20Modeling%20Debates%20in%20the%20House%20of%20Commons,%201832%961915.pdf

Fisher, W. W. (2014). *HLS1x: CopyrightX course report.* (HarvardX Working Paper Series 5.) Retrieved from SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2382332

Grainger, B. (2013). *Massive open online course (MOOC) report.* Retrieved from University of London website: http://www.londoninternational.ac.uk/sites/default/files/documents/mooc_report-2013.pdf

Grimmer, J., & King, G. (2011). General purpose computer-assisted clustering and conceptualization.

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

Proceedings of the National Academy of Sciences, *108*(7), 2643–2650. Retrieved from http://www.pnas.org/content/108/7/2643.abstract

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. Retrieved from http://stanford.edu/~jgrimmer/tad2.pdf

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the First ACM Conference on Learning@ Scale conference*, 41–50. Retrieved from http://pgbovine.net/publications/edX-MOOC-video-production-and-engagement_LAS-2014.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd Ed.). New York, NY: Springer.

Hillard, D., Purpura, S., & Wilkerson, J. (2008). Computer-assisted topic classification for mixed-methods social science research. *Journal of Information Technology & Politics*, *4*(4), 31–46. Retrieved from http://faculty.washington.edu/jwilker/TopicClassification.pdf

Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). *HarvardX and MITx: The first year of open online courses* (HarvardX and MITx Working Paper No. 1). Retrieved from SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2381263

Jamal, A., Keohane, R., Romney, D., & Tingley, D. (2014). Anti-Americanism or anti-interventionism: Evidence from the Arabic Twitter universe. *Perspectives on Politics*, *13*(1), 55–73. Retrieved from http://scholar.harvard.edu/dtingley/publications/american-eyes-arabic-tweeters

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*, 1–18. Retrieved from http://gking.harvard.edu/files/censored.pdf

Kizilcec, R. F. & Schneider, E. (2015). Motivation as a lens to understand online learners: Towards a data-driven design with the OLEI scale. *Transactions on Computer-Human Interactions, 22*(2). doi:10.1145/2699735

Koller, D., Ng, A., Do, C., & Chen, Z. (2013, May/June). Retention and intention in massive open online courses: In depth. *Educause Review*. Retrieved from http://net.educause.edu/ir/library/pdf/ERM1337.pdf

Kolowich, S. (2014). The professors who make the MOOCs. *The Chronicle of Higher Education*, *59*(28), A20–A23. Retrieved from http://chronicle.com/article/The-Professors-Behind-the-MOOC/137905/#id=overview

Lowe, W., & Benoit, K. (2013). Validating estimates of latent traits from textual data using human judgment as a benchmark. *Political Analysis*, *21*(3), 298–313. Retrieved from http://pan.oxfordjournals.org/content/early/2013/06/17/pan.mpt002

Lucas, C., Nielsen, R., Roberts, M., Stewart, B., Storer, A., & Tingley, D. (2013). Computer assisted text analysis for comparative politics. Retrieved from http://scholar.harvard.edu/files/dtingley/files/comparativepoliticstext.pdf

Mak, S. F. J., Williams, R., & Mackness, J. (2010). Blogs and forums as communication and learning tools in a MOOC. *Proceedings of the 7th International Conference on Networked Learning 2010*, 275–285. Retrieved from http://www.lancaster.ac.uk/fss/organisations/netlc/past/nlc2010/abstracts/PDFs/Mak.pdf

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011, July). Optimizing semantic coherence in topic models. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 262–272). Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=2145462

Murphy, E., & Rodriguez-Manzanares, M. A. (2012). Rapport in distance education. *The International Review of Research in Open and Distance Learning*, *13*(1). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1057/2076

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective.* Cambridge, MA: MIT Press.

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuned models of peer assessment in MOOCs. arXiv preprint arXiv:1307.2579

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, *54*(1), 209–228. Retrieved from https://www.law.berkeley.edu/files/TopicModel.pdf

Team, R. Core (2012). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: http://cran.case.edu/web/packages/dplR/vignettes/timeseries-dplR.pdf

Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., Chuang, I., & Ho, A. D. (2014a). *PH207x: Health in Numbers and PH278x: Human Health and Global Environmental Change 2012–2013 course report.* (HarvardX Working Paper No. 2.) Retrieved from SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2382242

Reich, J., Emanuel, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., & Ho, A. D. (2014b). *HeroesX: The Ancient Greek Hero: Spring 2013 course report.* (HarvardX Working Paper No. 3.) Retrieved from SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2382246

Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., Chuang, I., & Ho, A. D. (2014c). *JusticeX: Spring 2013 course report.* (HarvardX Working Paper No. 4.) Retrieved from SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2382248

Rees, J. (2013, March 5). Peer grading can't work. *Inside Higher Ed.* Retrieved from http://www.insidehighered.com/views/2013/03/05/essays-flaws-peer-grading-moocs#ixzz2MiKxNP7b

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2014). Structural topic models. *Working Paper*. Retrieved from http://scholar.harvard.edu/files/bstewart/files/stm.pdf

Roberts, M. E., Stewart, B. M., & Tingley, D. (2014). stm: R package for structural topic models. *Working Paper*. Retrieved from http://scholar.harvard.edu/files/bstewart/files/stmvignette.pdf

Roberts, M. E., Stewart, B. M., Tingley, D., & Airoldi, E. M. (2013). The structural topic model and applied social science. *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation.* Retrieved from http://scholar.harvard.edu/files/bstewart/files/stmnips2013.pdf

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B, & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*(4), 1064–1082. doi: 10.1111/ajps.12103

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*(1), 135–146. Retrieved from http://www.sciencedirect.com/science/article/pii/S0957417406001266

Shermis, M. D., & Hamner, B. (2012). Contrasting state-of-the-art automated scoring of essays: Analysis. Paper presented at *National Council on Measurement in Education Annual Meeting, Vancouver, B.C.* http://www.scoreright.org/NCME_2012_Paper3_29_12.pdf

Stewart, B. M., & Zhukov, Y. M. (2009). Use of force and civil–military relations in Russia: An automated content analysis. *Small Wars & Insurgencies*, *20*(2), 319–343. Retrieved from http://scholar.harvard.edu/files/bstewart/files/2009_stewartzhukov_swi.pdf

Stockmann, D. (2012). *Media commercialization and authoritarian rule in China*. Cambridge, UK: Cambridge University Press.

Stump, G. S., DeBoer, J., Whittinghill, J., & Breslow, L. (2013). *Development of a framework to classify MOOC discussion forum posts: Methodology and challenges.* Retrieved from https://tll.mit.edu/sites/default/files/library/Coding_a_MOOC_Discussion_Forum.pdf

Thomas, M. (2002). Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning*, *18*(3), 351–366. doi/10.1046/j.0266-4909.2002.03800.x

Van Atteveldt, W., Kleinnijenhuis, J., & Ruigrok, N. (2008). Parsing, semantic networks, and political authority using syntactic analysis to extract semantic relations from Dutch newspaper articles. *Political Analysis*, *16*(4), 428–446. Retrieved from http://pan.oxfordjournals.org/content/16/4/428.short

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009, June). Evaluation methods for topic models. *Proceedings of the 26th Annual International Conference on Machine Learning* (pp. 1105–1112). New York, NY: ACM Press. http://dl.acm.org/citation.cfm?id=1553515

Wang, Y. & Baker, R. S. J. d. (2014). MOOC learner motivation and course completion rates. *Working Paper.* Retrieved from http://www.moocresearch.com/wp-content/uploads/2014/06/MRI-Report-WangBaker-June-2014.pdf

Winchester, T. M., & Winchester, M. K. (2014). A longitudinal investigation of the impact of faculty reflective practices on students' evaluations of teaching. *British Journal of Educational Technology*, *45*(1), 112–124. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/bjet.12019/abstract

Yang, D., Wen, M., Kumar, A., Xing, E., Rosé, C. P. (2014). Towards an integration of text and graph clustering methods as a lens for studying social interaction in MOOCs. *Working paper.* Retrieved from http://www.moocresearch.com/wp-content/uploads/2014/06/C9184_ROSE_MOOC-Research-InitiativeRose9184X.pdf

(2015). Computer-Assisted Reading and Discovery for Student-Generated Text in Massive Open Online Courses. *Journal of Learning Analytics*, *2*(1), 156–184.

APPENDIX

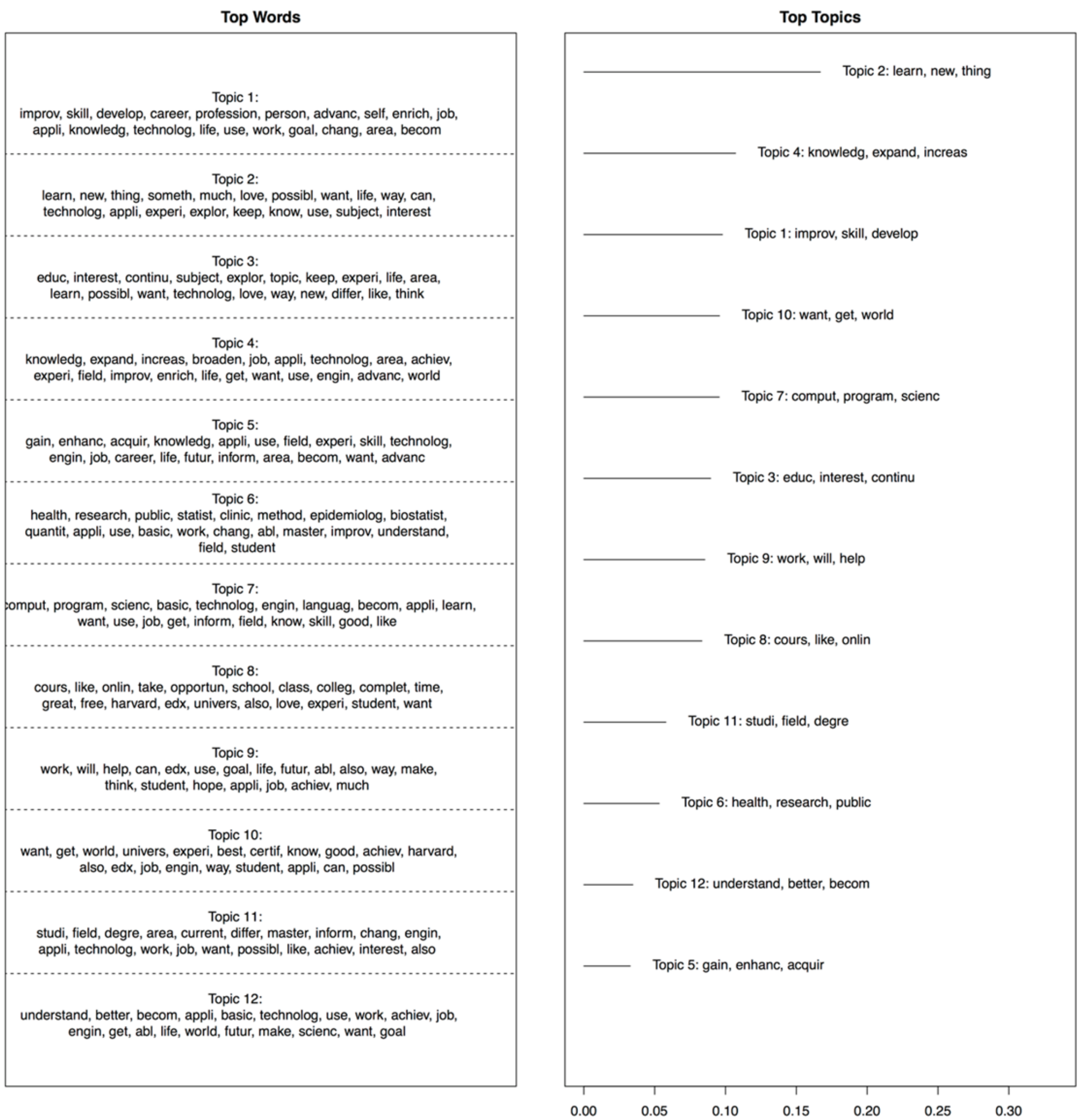


Figure 11. Educational goals. The left column lists words associated with each topic and the right column gives the distribution of topics across the corpus.

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

Discourse-Centric Learning Analytics: Mapping the Terrain

Simon Knight
Knowledge Media Institute, Open University, United Kingdom
sjgknight@gmail.com

Karen Littleton
Centre for Research in Education and Educational Technology, Open University, United Kingdom

ABSTRACT: There is an increasing interest in developing learning analytic techniques for the analysis, and support of, high-quality learning discourse. This paper maps the terrain of discourse-centric learning analytics (DCLA), outlining the distinctive contribution of DCLA and outlining a definition for the field moving forwards. It is our claim that DCLA provides the opportunity to explore the ways in which discourse of various forms both resources and evidences learning; the ways in which small and large groups, and individuals, make and share meaning together through their language use; and the particular types of language — from discipline specific, to argumentative and socio-emotional — associated with positive learning outcomes. DCLA is thus not merely a computational aid to help detect or evidence "good" and "bad" performance (the focus of many kinds of analytics), but a tool to help investigate questions of interest to researchers, practitioners, and ultimately learners. The paper ends with three core issues for DCLA researchers — the challenge of context in relation to DCLA; the various systems required for DCLA to be effective; and the means through which DCLA might be delivered for maximum impact at the micro (e.g., learner), meso (e.g., school), and macro (e.g., government) levels.

KEYWORDS: discourse, discourse-centric learning analytics, social learning analytics, data mining, computer supported collaborative learning, collaborative learning, learning analytics

1 INTRODUCTION

"Learning Analytics is an emerging research field and design discipline that occupies the 'middle space' between the learning sciences/educational research and the use of computational techniques to capture and analyze data (Suthers & Verbert, 2013)" (Knight, Buckingham Shum, & Littleton, 2014, p. 1). One interest for learning analytics is in its potential for the analysis of learning processes, and discourse data (Buckingham Shum & Ferguson, 2012). In 2014 we saw the fourth Learning Analytics and Knowledge (LAK) conference and the second Discourse-Centric Learning Analytics (DCLA) workshop (Buckingham Shum, de Laat, De Liddo, Ferguson, & Whitelock, 2014) as a part of that conference. It now

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

seems a good time to reflect on this developing field. In this paper, we discuss this sub-area of learning analytics — discourse-centric learning analytics — which we refer to specifically in order to foreground our interest in the middle space between learning and analytics (Suthers, Lund, Rosé, Teplovs, & Law, 2013; Suthers & Verbert, 2013); an interest not only in computational-analytic techniques for discourse (the "DC" of DCLA) but in the explicit learning implications of those techniques, which should be grounded in our understanding of educationally salient discourse. This paper maps the terrain of the work falling under DCLA, describing the relationship between theory and methods, the various targets of analysis, and the kinds of learning claims at which those analyses might be targeted. We do not intend to put forward any particular stance with regard to the relationship between discourse and learning, but rather by mapping the terrain, we highlight necessary considerations in any well-motivated approach to DCLA as a distinctive research area.

If learning analytics hopes to build on and contribute to learning theory, researchers should consider the ways in which learning analytics — in this case discourse-centric learning analytics — relates to, and is different from, other research on related topics. What is it that makes something DCLA, rather than natural-language processing, text-based machine learning, or some other form of learning analytics? The interest in developing DCLA techniques is in part driven by recognition of its potential bi-directional contribution; as we discuss further below, DCLA might take two forms: bottom up (inductive, exploratory), or top down (deductive, confirmatory). In either case, we must consider how the general approach relates to a distinctive sub-discipline of DCLA, and its relationship to existing theory. While in this paper we are specifically interested in DCLA, we anticipate that similar issues arise for other sub-domains of learning analytics. We consider these challenges in the context of DCLA, concluding with three key questions relating to *context, systems,* and *feedback mechanisms*. The paper thus provides a mapping of the DCLA terrain and its distinctive contribution, providing a focal point for DCLA's key challenges. A key component of our claim is that DCLA provides the opportunity to explore the ways in which discourse of various forms supports and evidences learning; the ways in which small and large groups, and individuals make and share meaning together through their language use; and the particular types of language — from discipline specific, to argumentative and socio-emotional — associated with positive learning outcomes. DCLA is thus not merely a computational aid to help detect or evidence "good" and "bad" performance (the focus of many kinds of analytics), but a tool to help investigate questions regarding the nature of discourse as learning, which are of interest to researchers, practitioners, and ultimately learners.

1.1 Preliminary Definitions for DCLA

Given its potential to support and investigate discourse in learning contexts, DCLA is thus of considerable interest. Whether in the context of Massive Open Online Courses (MOOCs), traditional university tuition, or more informal writing on the web, the automated or semi-automated analysis of linguistic data is increasingly prominent (see for example the classified sources in Table 1). Yet, it might well be noted that such analysis is not particularly new. Analysis of discourse and its communicative or illocutionary force is well established, and the ways in which new technologies mediate or alter such

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

properties are at least as old as Socrates' complaint regarding the diminished value of *written* argument (Plato, 1999). Given this context, it is important to consider the ways in which *DCLA* might differ in focus from related fields, what its potential is, what perspectives are brought to bear on the consideration of DCLA as opposed to other domains, and what are the defining features of DCLA.

The trivial response, of course, is that DCLA is just the application of learning sciences, and analytic approaches, to student discourse data. However, insofar as this leaves unspecified definitional questions regarding "discourse," "learning analytics," and indeed "centric," this is clearly inadequate. To give an example, imagine a case in which a learner's motivation is assessed during the writing of a formal assessment (via non-linguistic proxies), and this motivation is fed back to him or her in some meaningful way. While certainly we might say this is learning analytics, and indeed we could accept that there is "discourse" data involved (the written assessment), it seems less clear to us that we would wish to describe such analytics as Discourse-*Centric* Learning Analytics. This paper describes our developing perspective on this issue, and marks out the ground within which we think DCLA lies. In doing so, we describe a set of illustrative examples drawn from our own work at the Open University. These are not intended to be exhaustive, nor indeed exemplary, but simply exemplifications of the types of space within which we conceive DCLA, and the pressure points for such work.

A key focus of some earlier work (see for example, Scheuer, Loll, Pinkwart, and McLaren's 2010 review of the state of the art in computer-supported argumentation), has been the potential of automated analysis for discourse data — a key pivot point for DCLA. That review, and an earlier review (Clark, Sampson, Weinberger, & Erkens, 2007) of argumentation in Computer-Supported Collaborative Learning (CSCL) environments, largely focus on existing argumentation frameworks — that is, they take the "top-down" theory-driven approach. However, it is important to note the potential of DCLA approaches for contributing back to theory, in — well conceptualized — bottom-up, data-driven approaches, as we discuss in the section "Top-down and Bottom-up Approaches" below.

The first DCLA workshop (Buckingham Shum et al., 2013) proposed a mission statement for DCLA, to *"Devise and validate analytics that look beyond surface measures in order to quantify linguistic proxies for deeper learning."* Yet, we hold the view that, in a very real sense, linguistic activity is not simply an *indicator* (or proxy) for deeper learning, it is often the site of that learning. That is, our learning is not just *displayed* through discourse; the discourse forms a fundamental site of that learning. Indeed, Table 1 below gives an indication of some focal points for DCLA, which apply across the varieties of data that might properly be considered "discourse" (text and spoken-chat data, longer written monologues, CSCL interactions and so on). In each case, one can imagine — as is indicated in the columns — clear examples of educational foci, and benefits to such analysis. We highlight though, that as one moves down the rows, there is a shift in focus from individuals, to two types of small-group processes (Stahl, 2010); in the first two — subject knowledge and rhetorical capabilities and tendencies — the emphasis is on individual capabilities and knowledge; in the third, it is on the ways in which people co-operate and share knowledge (the collective level); while in the fourth, discourse is taken to be not only indicative of learning, but constitutive of that learning in a collaborate, co-constructive sense.

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

Table 1. Discourse Functions, Focus, Content, and Example DCLA techniques*

Function (from Knight, 2013)	**Focus**	**Content**	**Example DCLA techniques**
Supporting individuals' subject learning	Advancement of subject knowledge.	Propositional, semantic content of what is said. Appropriate reference to entities and relationships from the curriculum.	Techniques for recognizing named entities, esp. ontology-driven techniques for domain specific entity-relationship extraction (e.g., history; biology) (Rosé et al., 2008; Tablan, Roberts, Cunningham, & Bontcheva, 2013).
Supporting psychological development — the development of social and reasoning skills	Development of argumentation skills or dispositions.	The rhetorical content of language. How language is used to position and make claims/discourse moves in discourse or writing.	Techniques for extracting rhetorical and argumentation forms (Clark et al., 2007; De Liddo, Sándor, & Buckingham Shum, 2012; Scheuer et al., 2010; Weinberger & Fischer, 2006).
Promoting group understanding or commonality	The interactional nature of collaborative discourse, networks of documents.	Who says what to whom? The actors constituting the interpersonal context.	Techniques for identifying interlocutors in dialogue, and more diffuse social networks of strong/weak ties via different channels (Haythornthwaite & de Laat, 2010; Oshima, Oshima, & Matsuzawa, 2012; Sie et al., 2012).
Enabling sharing of ideas that can be improved together	Resources that emerge through the discourse, as "improvable objects" (Wells, 1999).	The ways discourse — particularly in CSCL contexts — resources future discourse and artefact development. The temporal and contextual (including social) development of ideas.	The temporal and interactional nature of the above 3 rows, and the nature of the shifts between states — the *pragmatic* context (Bannert, Reimann, & Sonnenberg, 2013; Furberg & Ludvigsen, 2008; Lid & Suthers, 2003; Medina & Suthers, 2009; Rosen, Miagkikh, & Suthers, 2011; Suthers & Desiato, 2012; Suthers & Medina, 2010).

*Table extended from Knight and Littleton (2013).

The purpose of the table is to highlight the multiple ways in which discourse may be probed, the different properties of such analysis, and the ways in which such analysis might interact. This is particularly important if, through our data collection, task setting, and analytic methods, we reify a particular perspective on the purposes of discourse that neglects the multi-functionality of discourse.

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

This is true in both the context of summative and formative assessments of discourse quality; we should be careful to avoid giving feedback on only one aspect of discourse, of obscuring the nature of the discourse in attempts to reduce data to manageable, neat categories (enabling, for example, visualization and dashboards), and of failing to appropriately scaffold both teachers and students in their use of talk for learning. In the latter case, alongside those concerns, the issue of performativity (see, for example, Ball, 1999) is raised — wherein those aspects of a curriculum directly assessed, become those aspects most focused on in classroom contexts. Indeed (as with any assessment instrument) there is the risk that as certain forms of DCLA become widespread, teaching and assessment become driven by those facets of discourse made visible by these techniques; that we value what we can measure, rather than measuring what we value (Wells & Claxton, 2002). Moreover, while quantification offers analytic opportunity, it is not the end goal of learning analytics properly understood; closing the loop for meaningful feedback is crucial to differentiating learning analytics from other approaches. This was the focus of the second workshop (Buckingham Shum et al., 2014), asking *"Once researchers have developed and validated discourse-centric analytics, how can these be successfully deployed at scale to support learning?"* Understanding this full cycle is crucial, and it is to this issue that we turn in the section "The Learning Analytics Lifecycle." While DCLA of various sorts may be productive, we should be aware of the interpretive flexibility of the tools we develop, and the pedagogic, assessment, and epistemological context of which they form a part (Knight et al., 2014). However, if we are able to operationalize some stable and general categories and patterns in discourse use, then we can model such patterns in classifiers, recognizing that such modelling involves a balance around acceptable levels of information loss (Rosé & Tovares, 2015).

In the following section ("The Learning Analytics Lifecycle") we discuss how this process might occur, moving on to discuss the potential for "Top-down and Bottom-up Approaches" in this analytics lifecycle. We then move on to "DCLA: UNDERSTANDING WHERE DISCOURSE MEETS ANALYTICS," discussing some specific examples of Discourse Sites and analytic levels, around which our concluding definition and questions are focused.

1.2 The Learning Analytics Lifecycle

By an "analytics lifecycle" we mean the cycle through which a need or desire for an analytic technique is identified, the analytic itself is developed, and the new analytic tool is then implemented in some form. Consider the developed learning analytics lifecycle (Clow, 2012) that moves through learners, data from or about learners, processing of data into metrics, and intervention. The process of "evidence centered design" (Mislevy, Behrens, Dicerbo, & Levy, 2012) follows a similar pattern, beginning 1) with the identification of target constructs, then 2) identifying behaviours indicative of those constructs, before 3) developing tasks likely to elicit those salient indicators. It is apparent, then, that DCLA should consider the types of tasks in which learners — formally or informally — engage, and the types of data that may be captured about those contexts. Theorizing is needed to consider how to process this data into appropriate metrics — how to interpret the data. Once metrics are developed, consideration of interventions should focus on ways in which data might be meaningfully represented either for students

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

or for expert educators in support of learners (in line with Clow's point that interventions do not necessarily need to involve a return of data to students).

Consideration of other broad models of analytic cycles including that of learning analytics (Elias, 2011) — as summarized in Table 2 — indicate a broad alignment, in particular with the need to define goals (in our context, learning goals), in making decisions regarding data selection and capture. Each of the example cycles noted in the columns of Table 2 involves a set of key steps, from data selection through to use. It is important to note that, even at the stages of selecting and processing data, a range of irreducible elements involving theoretical models, target constructs, and data constraints are involved. For example, the segmentation of discourse data into chunks for processing is both a pragmatic decision regarding the selection of statistical technique, and also a theoretical one around the units at which data may be meaningfully discussed. A crucial question to ask at each DCLA stage — from task design and data collection to the steps of analysis and feedback to students — is "what is it for?" In this sense, DCLA is positioned in the "middle space."

Table 2. Comparison of Analytics frameworks and models (Elias, 2011, p. 10)

Knowledge Continuum	**Five Steps of Analytics**	**Web Analytics Objectives**	**Collective Applications Model**	**Processes of Learning Analytics**
		Define Goals	Select	Select
Data	Capture	Measure	Capture	Capture
Information	Report		Aggregate	Aggregate & Report
Knowledge	Predict		Process	Predict
Wisdom	Act	Use	Display	Use
	Refine			Refine
		Share		Share

Given, then, the emphasis of various analytic models on selecting data and defining goals, an interesting question is raised — to what extent can "bottom up" approaches play a role in developing our theory and empirical understanding of phenomena, given the need for theory in data selection?

1.3 Top-down and Bottom-up Approaches

As we note in the introduction, the interest in DCLA is in part driven by recognition of its potential bi-directional contribution; DCLA might take two forms: bottom up (inductive, exploratory), or top down (deductive, confirmatory). Either might appropriately address a learning analytic development cycle — feeding into how we understand our desired construct, the types of behaviours associated with it, or the kinds of feedback likely to support some particular facet of learning.

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

In the former case, discourse data and learning outcomes are mined for patterns indicative of particular associations between discourse use and learning. In the latter, prior educational research is applied using new analytic techniques. As Cooper (2012) notes, both types are important contributions to the field of analytics. As Gibson (2013) points out (Table 3), data-driven approaches are different; the potential of such models is to simulate sophisticated aspects of higher order learning (with commensurate ethical concerns about auto-classifying students (Johnson, 2013). In such a view, inductive ("data-driven") approaches are used to derive data, from which hypotheses can be produced (validated), while deductive approaches (the traditional scientific hypothetico-deductive model) construct hypotheses to test through data collection and analysis.

Table 3. Comparison between traditional and data-driven science methods (Gibson, 2013, p. 6)

	Traditional scientific method	**Data-driven science method**
Step 1	Ask a Question	Ask a Question
Step 2	Do Background Research	Do Background Research
Step 3	*Construct a Hypothesis*	*Simulate Theory or Search for Patterns*
Step 4	*Test Hypothesis with Experiments*	*Generate and Analyze Data*
Step 5	*Analyze Data and Draw a Conclusion*	*Conduct Validation Experiments*
Step 6	Communicate Results	Communicate Results

In both cases, the relationship between prior theorizing and empirical work, and new techniques should be established. Validation is fundamental here, and understanding the ways in which concepts are operationalized across the various manual, semi-manual, and automated methods is important (see for example, Chi, 1997). Of course, there are also cases where modelling computationally can inform theory. For example, Reimann (2009) discusses process models that analyze sequences of events — of relevance here insofar as they may be identified automatically from log data — in contrast to more traditional approaches that seek to model the variance in the occurrence of dependent variables. In this discussion, Reimann notes that

> Process mining can serve a number of purposes, among them: (a) Discovery—No a priori model exists. Based on an event log, a model is constructed; (b) Conformance—An a priori model exists. Event logs are used to determine the extent to which the enacted collaboration corresponds to the model; (c) Extension—An a priori model exists. The goal is not to test but to extend the model, for instance with performance data (e.g., durations of activities). (Reimann, 2009, p. 251)

As Reimann notes in later work,

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

> For [Educational Data Mining] EDM to contribute more directly to theory building, it needs to be applied to data that measures theoretically relevant properties and mechanisms. These are not necessarily found in the log files of software that has been designed for practical educational rather than research purposes. It is unlikely that we will find many theoretically interesting relations by looking at more of the essentially same kind of data. "Big data" and "more data" are not identical with conceptually "rich data" and "deep data" that capture not only multilayered phenomena but also a rich account of learning contexts. The latter, we believe, could be a productive methodological direction. (Reimann, Markauskaite, & Bannert, 2014, p. 11)

Crucially, the two aspects of analysis that Chi (1997, p. 311) notes in relation to quantitative and verbal analyses hold true here too: first, generation of the right questions is crucial for determining the types of analyses one conducts; and second, understanding the means (the "mechanics") through which one conducts that analysis is a crucial consideration, not to be removed from (nor integrated in) the first. Similarly, we should take lessons about the translation of research to practice, and the transfer of practice from site to site into account in our analytics, bearing in mind that "educational data for learning analytics is context specific and variables carry different meanings and can have different implications across educational institutions and area of studies" (Ifenthaler & Widanapathirana, 2014, p. 1).

2 DCLA: UNDERSTANDING WHERE DISCOURSE MEETS ANALYTICS

The preceding discussion of learning analytics lifecycles, and the role of top-down and bottom-up processes indicates that, although there is clear interest in the application of existing educational research to online contexts, particularly where analysis and learning support may be automated, more research may be needed to understand these applications. That is, the application of our existing theory to the context of novel forms of data — the types of data capturable, the contexts in which the data is created and captured, and the differences in the medium of online contexts — may need further research.

For example, we have a strong interest in "exploratory dialogue" (see, for example, Mercer & Littleton, 2007), which in small group contexts has a demonstrated relationship with improved learning outcomes. However, little research has been conducted on exploratory dialogue in online contexts (see, for example, Littleton & Whitelock, 2005), with only a few studies of such asynchronous dialogue (see, for example, Ferguson, Whitelock, & Littleton, 2010), and none that we are aware of in the context of large multi-party and multi-modal conference chat systems. Yet, it is clear that there are differences between face-to-face and online communication, and many online contexts provide different types of opportunities for communication. This is an important point regarding a "middle space" between the learning sciences and analytic techniques (Suthers & Verbert, 2013); even a meeting in the middle space can be approached from either side, and such trajectories affect the ways we action our theories. If our operationalization of prior (offline) educational research is driven by the analytic techniques available,

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

then the types of validation in which we must subsequently engage, and the ways in which we approach the on- and offline dialogue as related or not, should be considered. Similarly, our analytic techniques should not be limited by the ways in which educational dialogue has previously been operationalized; new analytic techniques afford new opportunities to theorize and reconsider what constitutes productive dialogue.

Indeed, if we consider the classic "quantification of verbal analysis" text (Chi, 1997), in that (manual) case, the point is made that all stages of analysis involve both top-down and bottom-up processing; coding is driven by top-down theory, but the application of codes is refined through bottom-up protocols (Chi, 1997, pp. 309–310). To give a concrete example in the machine learning case, Rosé and Tovares note, "more attention should be given to the problem of representing data appropriately" (Rosé & Tovares, 2015, p. 6), particularly with regard to the issue of segmentation where different levels of granularity in segmentation for the training and classification of various forms of natural language data may have an impact. This concern is a complex one, and in earlier work Arguello and Rosé (2006) propose that regarding topic segmentation, researchers first need a definition of "topic," which should be:

1. Reproducible by human annotators
2. Not reliant on domain-specific or task-dependent knowledge
3. Grounded in generally accepted principles of discourse structure (in order that shifts in topic are recognizable from surface characteristics of the dialogue)

Thus, although in that case hidden Markov models were used to detect topic shifts, the point is to reproduce a human-recognized topic model (rather than anything more ephemeral). That is, although "bottom-up" methods may create productive means of segmentation, the relationship of such methods — and their outputs — to theoretically grounded assumptions should not be forgotten.

In the next two sub-sections, we give some examples of contexts in which discourse data is produced, and is associated with learning outcomes. In the section on discourse sites, we introduce discourse contexts in which data is collected, and the implications of those learning contexts and the data collected for the analytics we conduct. This section is broadly concerned with the "top down" approach to DCLA, derived from existing learning theory and empirical work. In the section on learning analytics, we discuss a broad set of analytic techniques for discourse data, and the implications of these for our learning-discourse analysis, and feedback representations. This section is broadly concerned with the "bottom-up" approach to DCLA, derived from rapidly advancing analytic techniques.

2.1 Discourse Sites

The aim of this (and the next) section is to introduce some examples of work on the "sites" of discourse noted in Table 1, around which there has been interest in developing analytics. We use the examples to illustrate the ways in which people develop their reasoning through dialogue, build arguments in CSCL

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

environments, and make rhetorical moves in written texts. Returning to Table 1, recall rows that focus on

1. The kinds of subject knowledge people evidence
2. The kinds of argument skills or dispositions they evidence
3. The ways in which learners interact with each other and with ranges of documents
4. The ways in which meaning is built up through shared interaction on documents

While identifying subject knowledge (content, entities, concepts, etc.) is important, it is not the only crucial element. So too for interaction — whether with citations, connecting "nodes" in CSCL knowledge-building environments, or simply through response to others' arguments or moves (e.g., responding to a question). Instead, *combinations* of these important features of a text are crucial to understanding what is going on. The first two rows of the table, then, are about individual levels of knowledge, the third about collective levels, and the fourth about collaborative co-construction.

We use the following sections to highlight some exemplar (but by no means exhaustive) learning sciences literature around discourse, taking as our focal point theorizing around argumentation and reasoning. In the following paragraphs, we present three types of discourse data: 1) unstructured dialogue, 2) free-text, and 3) dialogue within structured CSCL environments. In each case, we are interested in the forms of data available and the contexts in which they are obtained. These are of particular interest for the following section, which discusses the ways in which such data has been treated analytically, and raises some potential areas of interest. The point here is to note particularly how theory informs our collection of, and analysis of, data and to foreground the levels at which DCLA might operate.

2.1.1 Unstructured dialogue

Perhaps the most obvious form of discourse data is the kind of computer-mediated communication seen in "chat" programs, and free-form dialogue of classrooms (and indeed, everyday conversation). In everyday classroom contexts, teachers model the use, and pay attention to the uptake of, key content and rhetorical terms in student dialogue (rows 1 and 2 of the table). A common complaint in school classrooms at least, is that whole-class discussion focused on questioning tends to focus too much on these levels of — individual — content-based understanding, rather than whole class commonality (Alexander, 2008). For these levels of discourse use (the latter 2 rows of the table) classroom "assessment for learning" techniques such as "mini-whiteboards" and interactive whiteboards (Hennessy & Warwick, 2010) for all students to display an answer, or more dialogic modes of discussion (Alexander, 2008), may be more important. In our own work we have a strong interest in the ways in which unstructured dialogue is used to co-construct common knowledge (Edwards & Mercer, 1987; Littleton & Mercer, 2013), contributing to the strong consensus among researchers that in a variety of contexts, high-quality dialogue is associated with learning (see the collection edited by Littleton & Howe, 2010). A core component of this research — along with that around Accountable Talk (Michaels, O'Connor, Hall, & Resnick, 2002; Resnick, 2001) — is the analysis of how ideas (including subject knowledge) are deployed in arguments, and taken up by others, to help build new knowledge together.

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

Fundamental to this process is critical but constructive engagement with others' ideas, participation from all members, respectful challenges and justifications, and the consideration of opinions before agreement is sought — in short, to make the reasoning visible in the dialogue (Mercer & Littleton, 2007).

2.1.2 Written texts

Of course, spoken dialogue through short exchanges is not the only site of learning. In particular in formal learning contexts — although by no means exclusively — much learning also occurs through the construction of, and is represented in, longer free-form written texts. Thus, one means through which the kind of co-construction noted in the 4th row occurs, is in the interaction with "the given," with written texts of various forms (see, for example, Knight & Littleton, 2015). Indeed, this too should be a core area of interest for DCLA researchers — how learners engage with texts, to construct longer pieces of writing, and how that writing is a site for, and representation of, their learning. Again, subject knowledge is key here, but so too is genre — including the kinds of "rhetorical moves" (or arguments) available, and so on. As Simsek, Buckingham Shum, Sándor, De Liddo, and Ferguson (2013, p. 3) note, analysis of texts for factual statements provides only a partial picture; in scientific research publications, authors make connections between claims of various sorts, including supporting, refuting, and so on, and this meta-discourse is fundamental to analysis of such papers.

Again, we see the importance not only of subject knowledge, but argumentation skill (rhetorical moves made), use of other's ideas (through citation and background knowledge), and genre-specific knowledge building (each discipline having its own "style" of making sometimes similar claims). Additional Natural Language Processing (NLP) approaches of interest to DCLA can be found in the emerging field of Argument Mining, which sits at the convergence of machine learning, information retrieval and argumentation theory (Palau & Moens, 2009). Again, the four rows of the table can be explicated in terms of this data — at the first two levels (or rows), the use of particular writing genres or scripts, and deployment of key concept terms, with the third and fourth rows seen in deeper engagement with genres, and deeper interactive reasoning around texts.

2.1.3 Structured dialogue and CSCL environments

As we note above, the relationship between "given," or sole-author student created texts and student dialogue are not firm. Indeed, there have been calls for "abandoning the forced dichotomy between two genres of collaboration tool" (Enyedy & Hoadley, 2006, p. 414), calling for fusion between information (document), and communication interfaces in part to facilitate the kinds of learning identified in the third and fourth rows of Table 1. The use of CSCL tools to support document-oriented, or less directed reasoning tasks is widespread; aiming to support the kinds of individual learning identified in the first two rows of Table 1. Much work has been conducted on developing environments that support, and make available for researchers, particular types of dialogue. Importantly though, while de Vries, Lund, and Baker (2002), to take one example, note important characteristics of productive epistemic dialogue, which bear striking resemblance to those described in this paper, they note that such dialogue is not "automatically" produced in structured environments. Indeed, Dillenbourg (2002) notes that some such environments risk "over structuring" and thus restricting the use of important types of dialogue. We

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

note, then, that while design may reduce computational difficulties (for example, by introducing threading to discussions), the points made in this paper with respect to the importance of context are still fundamental to understanding the dynamic features of dialogue through which learning is co-constructed. Computer environments may be seen as complementary to such dialogue, in particular where they formalize, through the user interface conceptual model, some of the systems through which exploratory and accountable dialogue are more likely to occur — the "ground rules" of each. It is thus that systems have been developed specifically to support particular types of formalized argumentation schema (Clark et al., 2007; De Liddo, Buckingham Shum, Quinto, Bachler, & Cannavacciuolo, 2011; Scheuer et al., 2010; Weinberger, Ertl, Fischer, & Mandl, 2005; Weinberger & Fischer, 2006). Again, we highlight the range of ways in which CSCL argumentation aligns with the table with some scripting aimed at individual learning gains (particularly around argumentation schema), and other methods aimed to aid in intersubjective meaning making.

2.2 Analytic Levels

We have highlighted some of the data forms and contexts of learning discourse above. The aim of this section, then, is to discuss the data-types and learning theory in light of analytic techniques. We note that even in contexts in which inductive or exploratory approaches to data analysis are used, the data itself is still imbued with theory insofar as the learning contexts from which that data is collected require theorizing; for example, genre detection requires a notion of "genre," and the collection of documents displaying such classes. Indeed, in the rather different context of transcribing audio and video recordings, Hammersley (2010) makes similar points: transcriptions cannot be said to be purely "constructed" as they bear an obvious and undisputable relationship to their object of representation, which is closely associated with their theoretical ontogeny, the prior knowledge of the transcriber, and the purposes for which the transcription is being created. Similarly though, they are not "given" direct insights into the world; shared understanding is constructed in the process of transcription itself and relies heavily on a (given) set of cultural and theoretical values including the best ways to represent given classes of behaviour (for example, whether to include notes on intonation). In this section, then, we highlight the ways in which theory should penetrate analytics for DCLA, and the complexity arising from the intersection between our data type (dialogue, text, CSCL) and analytic focus (subject knowledge, argumentation skill, interaction, and broader knowledge-building culture).

2.2.1 The individual level

If we consider again Table 1, recall that the first two rows regard individual level learning gains — in terms of content- or subject-based learning and personal psychological development. Analysis of such learning through discourse focuses on the ways in which individuals use terms, whether in individual or collective contexts; what is key is not the type of discourse data, but rather the analytic focus. The most basic kinds of analysis take simple "bag of word" or cue phrase approaches — in which classes of dialogue are assigned based on the presence of key words or phrases — to detection of concepts. Such approaches can be enriched with ontologies to match key phrases from some domain, with those present in the discourse data. More complex variants on this approach might include a broader feature

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

set — such as grammatical markers, particularly in the case of argumentation features — to analyze the presence of particular types of discourse. As such, analysis focuses not on the interactional features of discourse, which such techniques are not methodologically equipped to analyze, but on the individual's use of terms and styles of discourse, and their relationship to individual learning (of concepts, and perhaps of rhetorical moves of some kind, and so on).

This kind of approach can be applied across dialogue, written text, and CSCL environment discourse data, although the former more naturally lends itself to these techniques. Written essays also offer opportunity for a variety of other semantic-related analyses, such as latent semantic analysis (LSA), which goes beyond cue-phrase analysis to look at underlying (latent) semantic content in the text. Such analyses can also be conducted on appropriately segmented dialogue data; for example, Epistemic Network Analysis (Shaffer et al., 2009) conducts a principal component analysis to explore connections between themes throughout a dialogue. A major advantage of CSCL environments for such analysis is the ability to build in "tagging" systems to enable learners to "tag" their entries with semantic labels, which can then be analyzed. In each of these cases, feedback can focus on the types of *concepts* and *facts* included in the discourse, and their incidence. Similarly, analysis for argumentation, rhetorical moves, and tendency (or disposition) to use such devices can also be analyzed through these methods. Again, feedback can focus on incidence of particular markers, but it can also — when combined with content analysis — focus on the ways concepts and facts are connected using argument-markers (for example, "it is argued that Broca's area is related to language because evidence from studies (ref1, ref2, etc.)…," where Broca's and the references noted are conceptual or factual claims, and the marker "because" links them). Again, CSCL knowledge-building/argument-mapping environments are particularly well equipped for this kind of analysis, in which the "connections" learners make between claims (often represented as "nodes") are often pre-labelled, and thus in machine-readable format.

These notions then cover the ways in which individuals use language to share meaning, and express their knowledge. However, we are also interested in how learners interact with each other using dialogue, how they move beyond *expressing* to "taking up." There are two levels of interest here. The first (and third row of the table) regards the ways in which learners interact with each other and ranges of documents, and how they pass information to each other, which is then used in the collective context. The second regards how meaning is built up through shared interaction, how knowledge is co-constructed. Again, the first two rows of the table are about individual levels of knowledge, the third about collective levels, and the fourth about collaborative co-construction.

2.2.2 The collective small group level

At the collective level of analysis, the focus of attention shifts from expressions by individuals, to the ways in which individuals — and the things that they say — interact. Thus, for example, Rosé and colleagues have explored transactivity (sometimes called intertextuality) in small groups (Sionti, Ai, Rosé, & Resnick, 2011; Stahl & Rosé, 2011), dyads (Gweon, Jun, Lee, Finger, & Rosé, 2011), whole-class discussions (Ai, Sionti, Wang, & Rosé, 2010), and its use for summarizing group discussions (Joshi & Rosé, 2007). Similarly, the ways in which written texts can be used to resource further written texts — in

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

or out of relatively structured environments — is clearly important (Enyedy & Hoadley, 2006; Knight & Littleton, 2015; Lid & Suthers, 2003; Suthers, Vatrapu, Medina, & Dwyer, 2007). Again, the structured nature of CSCL environments supports analysis to explore the ways in which ideas are "taken up" by others, and knowledge is pooled — collectively — to create (perhaps individual) answers. For example, "contingency graphs" in CSCL environments (Medina & Suthers, 2009) provide a visual insight into the ways in which one "idea" (a node), is "contingent upon" earlier ones. In each of these cases, analysis can focus on how terms are stated by an individual or given text, and then subsequently repeated by other individuals. More sophisticated analysis — particularly in the case of longer written texts — might also focus on the latent-semantic content of a student-produced text in comparison to those from which it is supposed to draw (see for example, Hastings, Hughes, Magliano, Goldman, & Lawless, 2012). Indeed, experimental work can also be modelled on Azmitia and Montomery (1993), who artificially manipulated which pieces of information participants in a group had, such that information *needed* to be passed between members in order to complete a task.

However, as Stahl (2013) recently noted, in such cases transactivity is being explored at the analytic level of individuals who have (differing) partial information regarding a problem that requires pooled information to solve it, rather than at the group level, which has come to be of great interest to the CSCL community. This focus on individuals in collaborative contexts, in contrast to collaborative units, is common to much group-work research; for example, in exploring transactivity, Azmitia and Montgomery's (1993) interest was in dialogue used to build upon a partner's explicit reasoning statements, rather than on the language used to co-construct knowledge. This, then, brings us to our last interest — that of the co-constructive level (the 4th row) of group-cognition.

2.2.3 The co-constructive small-group level

A fourth level of analysis is at the group-co-constructive level. Indeed, notions of transactivity are highly relevant here; for example, Sionti et al. (2011) summarizes notions of transactivity as sharing a requirement for interlocutors' reasoning to be explicitly displayed, and usually for contributions between speakers to be connected such that a speaker's utterances resource future utterances. In addition to this interesting construct, it is also worth mentioning other work — relating more closely to active participation and the reasoned consideration of others' arguments — in which the notion of "heteroglossia" is explored. Heteroglossia (Bakhtin, 1986) is related to the multi-vocality of perspective, the characteristic of a text as displaying, and being open to, multiple views — a significant element of dialogic education (Wegerif, 2011). Building on Martin and White's (2005) description of dialogic expansion (in which alternative positions are available), and dialogic contraction (in which the scope of permitted perspectives is restricted), heteroglossia has recently been operationalized in the computational linguistics context as:

> the extent to which a speaker shows openness to the existence of other perspectives apart from the one that is reflected in the propositional content of the assertion being made... Within our Heteroglossia analysis, assertions framed in such a way as to acknowledge that others may or may not agree, are identified as heteroglossic. We describe it as identifying wording choices that do or don't treat other perspectives than

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

> what is expressed in the propositional content of the assertion as open for consideration within the continuing discussion. (Rosé & Tovares, 2015, pp. 10–11)

In the context of a science classroom task, this has been operationalized at the coding level for a four-part scheme that takes co-constructive dialogue as an "openness to the other" in the way phrases are used. By building in further features at the small-group level, such as looking at dialogue-acts for interaction properties within the group (Erkens & Janssen, 2008; Král & Cerisara, 2012; Stolcke et al., 2000), we can draw further conclusions about how small-groups work together in co-constructive ways to create new meaning. In such analytic approaches, the focus is on the ways in which terms are "taken up" by others, and the collective accretion of knowledge over time, as indicated in the fourth row of the table. To analyze such learning, a temporal analysis of features — understanding how terms are used, by whom, and how they relate to each other — is crucial (Mercer, 2008). While temporality is, of course, relevant to other areas of learning (including rows 1–3), it is particularly salient to this fourth, co-constructive level.

2.3 Section Summary

In learning contexts, we are interested not only in the deployment of particular terms, but of their *effective deployment* in contexts that indicate an understanding of the inferential properties of the term to other concepts. Analytics that explore key words in abstracted ways may obscure the misuse of terms, or — moreover — their simple copying from texts that students use for contextualizing purposes, such as task instructions. The pragmatic level of description is therefore important: syntactic or semantic levels of description can be blind to understanding what is being done in interaction. For example, the "same" question (in syntactic and semantic terms) might be asked at the beginning and end of a lesson, while serving different (pragmatic) functions: in the first instance, to gauge baseline understanding and provide a reference point for the second posing of the question, which is to see how the question may now be reinterpreted. Describing what is happening is not the same as understanding what is being done.

We are mindful that, regarding the complexity of discourse data, one solution — for example through digital games (Gee, 2008; Shaffer, 2008) — is to "design in" for the types of language we want to teach, and analyze; providing opportunities for their display and capture in ways easy to process by machines. Certainly this point is important, and DCLA should consider the ways it might be informed by, and inform, such structured environments. The potential to build complex, discursive learning contexts is important and the full breadth of ways in which discourse is constitutive of learning should not be glossed by limiting analytic techniques.

3 CONCLUSIONS: WHAT IS DCLA?

There is a growing interest in, and availability of, data and techniques to analyze that data. Discourse is an important feature of learning, and fundamentally associated with the context of social learning

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

analytics, dispositional analytics, and other emerging analytic-sub-fields. We should strive to understand discourse in this wider context. A further driver here is the fact that interest in online text summarization tools, and support for writing both in and out of formal educational contexts will continue to grow — again, DCLA may play an important role here, perhaps complemented by other forms of analytic tools (such as social network analytics).

3.1 A Definition for DCLA

This paper has highlighted the ways in which we see DCLA as distinctive; in doing so, we have tried to contextualize it with respect to the wider research. We began the paper by noting two early focal points for DCLA:

1. Devise and validate analytics that look beyond surface measures in order to quantify linguistic proxies for deeper learning
2. Once researchers have developed and validated discourse-centric analytics, how can these be successfully deployed at scale to support learning?

As we have highlighted throughout this paper, we must be cautious here — linguistic activity is not simply an *indicator* for deeper learning, it is the very site of that learning; when we talk to each other, write texts, engage with knowledge-building environments, we are not simply representing what is known — providing *indicators* for that knowledge — we are actively engaging in a co-constructive activity, which constitutes learning. With this understanding of high-quality learning discourse, we propose a definition of DCLA:

> DCLA focuses on analytics to support high-quality discourse for learning contexts; it consists of the analysis of discourse data, the creation of effective feedback to learners and educators, and the validation and theorizing of our analytic techniques.

This definition, of course, leaves open the particular theorized account of discourse, and pedagogic approach associated with particular types of discourse. Throughout the paper, we have tried to highlight some core pivot points for DCLA, seeking to elaborate our understanding of learning analytics as centred around discourse data. In particular, we have noted the ways in which various sites of discourse might be seen as related to, or constitutive of, learning at a variety of analytic levels — from the individual to the small group. The definition given does not preclude the use of analytic devices for analysis of learning *indicators* in discourse, of interest particularly to evidence learning at an individual level, but rather it leaves open the space to explore the full range of levels at which learning and discourse are related. It also foregrounds the need for analytics *centred* on discourse to consider validation in transferring analysis from existing learning theory to analytic techniques, and vice-versa, as well as the need to consider the wider learning and feedback systems within which analytic techniques are deployed — particularly as targeted at *learning* outcomes. We close the paper with three key issues that researchers in DCLA will need to consider further: 1) the role of context, 2) DCLA systems, and 3) effective impactful feedback.

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

3.1.1 What is the relationship between discourse in different contexts (particularly on- and offline), and what are the implications for detection of particular types of discourse?

As we note above, the differences between on- and offline instantiations of particular discourse-genres are complex, and this area will be of much importance. In order to apply "offline" research in online contexts, we must be sensitive to the differences and the possible need to "validate" such a shift. Similarly, lessons from online contexts may not translate directly to offline ones — not least because the types of data we can gather are rather different, and thus our insights may be different as well. For example, creating written texts using technologies — including collaboration tools such as etherpads or google docs — provides different affordances, and research data, to "pen-and-paper" exercises. Consider also the wealth of research on computer-mediated communication (CMC) that recognizes and explores differences in CMC and "offline" talk. DCLA researchers should consider the role of existing theory, and the place of analytics in that context. This issue thus concerns:

1. The practical-context sensitivity of DCLA, and indeed all discourse analysis, regarding pedagogic, assessment, cultural, and individual differences aspects
2. The theoretical-context sensitivity of DCLA and how alignments and mismatches between on- and offline contexts might be addressed
3. The role of inductive and deductive techniques in advancing analytics and theory
4. The need for validating learning analytics techniques

3.1.2 What systems need to be in place for the successful deployment of DCLA?

DCLA rely on particular tools for data capture and analysis. However, their use may also make commitments to particular pedagogic, assessment, and epistemological assumptions (Knight et al., 2014). While tools have particular affordances (for example, twitter replies, forum threading, knowledge-mapping connections), these tools can be used in many learning contexts, and their data analyzed for many purposes. This highlights that the issue of "systems" relates not only to the analytic device, but also to the importance of meaningful activities in learning contexts for learning analytics. We agree that more attention should be paid to such claims, and that design strategies have strong potential for the development of meaningful, and enjoyable, learning activities centred on language use (Gee, 2008; Shaffer, 2008). Indeed, much work has been conducted on developing environments that support, and make available for researchers, particular types of discourse. DCLA can play a role in this. In considering the necessary systems for the successful deployment of DCLA we should thus consider the following:

1. The technical factors involved in DCLA, the types of data we can (and want to) collect, and how they are processed
2. The pedagogic and assessment context within which data is collected — collection of data should occur in purposeful, pedagogically valuable contexts
3. An understanding of the contexts within which certain discourse types are likely to occur, such that we do not deploy the "wrong" analytic device in contexts where we would not expect its target class to occur

3.1.3 What is the most effective way to feedback results to learners, educators, and policy makers for maximum impact?

As discussed above, the learning analytics cycle is not solely about the creation of numerical insights into the target-data discourse in DCLA. Rather, we are interested in how we can develop tools to support education. As such, understanding how best to feedback *discourse* supporting data — to whom, in what form, and from what data — is crucial. While computationally complex analytics might hold potential, simpler techniques, might also be impactful. For example, analytics that indicate which topics are being talked about, by whom, giving information such as turn taking — without evaluating "good" or "bad" performance — might provide pedagogic support, while being very computationally simple. In addition, learners, educators, and policy makers each have different needs; analytics may provide representational tools to resource discussion around learning impact at all levels. A part of analytics is creating representational tools to understand data we already have; this question relates to that data, and new data emerging from learning analytic techniques. Whether computationally simple or complex tools are deployed, interpretive flexibility plays a role. This issue thus concerns the following:

1. The ways in which the same data can be collected, analyzed, and fed back to relevant stakeholders at the macro, meso, and micro levels (see Figure 1 below)
2. Research around visualization, at the various levels, to support "in the moment" and "after the fact" discourse
3. In addition to supporting learning *through* DCLA, research should also be conducted on how to support learning *about* DCLA — how learning situations can make best use of DCLA and in particular how at each level (macro-meso-micro) stakeholders can communicate with others around DCLA sensemaking (Knight, Buckingham Shum, & Littleton, 2013)

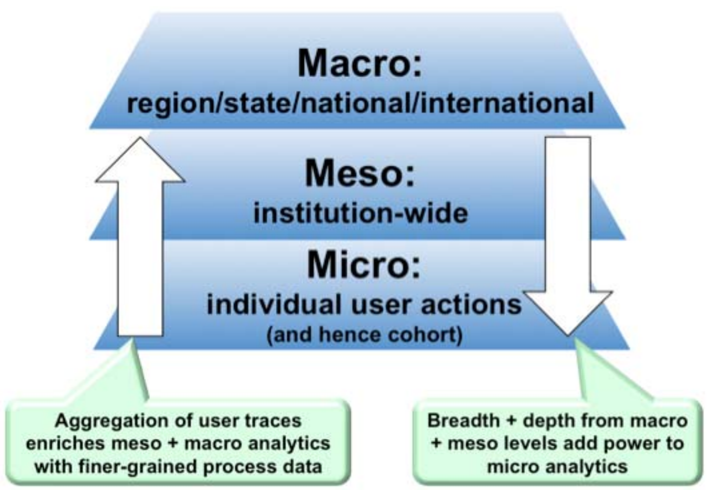


Figure 1. The macro-meso-micro levels of analysis for learning analytics (Buckingham Shum, 2012, p. 3)

We can envisage scenarios, therefore, in which leaders at the institutional level and above see aggregated DCLA (for example, as proxies for critical thinking, constructive knowledge building, or successful deployment of online forums), and are able to compare datasets from many contexts. The extent to which comparisons can be made with integrity within and between contexts is, of course, an issue afflicting all types of analytics (educational or otherwise), such is the ease with which digital data may be fused. The particular roles that discourse plays in learning, which we have sought to illuminate in this paper, alert us to the risk of decontextualizing and comparing the different types of DCLA without due care. That being said, there is an intriguing array of opportunities within the middle space of learning sciences and computational techniques around discourse data; by staking the distinctive contribution of DCLA, we highlight this potential and the core challenges for DCLA researchers moving forwards.

ACKNOWLEDGEMENTS

We are grateful to Simon Buckingham Shum for feedback on earlier drafts, and to the anonymous reviewers whose comments helped to improve this version. Our thanks to David Gibson and Simon Buckingham Shum for kind permissions to include Table 3 and Figure 1, respectively. Particular thanks to Tanya Elias for releasing Table 2 under a Creative Commons BY-NC-ND license.

REFERENCES

Ai, H., Sionti, M., Wang, Y.-C., & Rosé, C. P. (2010). Finding transactive contributions in whole group classroom discussions. *Proceedings of the 9th International Conference of the Learning Sciences, Volume 1* (pp. 976–983). Chicago, IL: International Society of the Learning Sciences. Retrieved from http://dl.acm.org/citation.cfm?id=1854360.1854485

Alexander, R. (2008). *Towards dialogic teaching: Rethinking classroom talk* (4th ed.). Cambridge, MA/London, UK: Dialogos.

Arguello, J., & Rosé, C. P. (2006). Topic segmentation of dialogue. *Proceedings of the HLT-NAACL 2006 Workshop on Analyzing Conversations in Text and Speech* (pp. 42–49). Retrieved from http://dl.acm.org/citation.cfm?id=1564542

Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development*, *2*(3), 202–221.

Bakhtin, M. M. (1986). *Speech genres and other late essays* (Vol. 8). Austin, TX: University of Texas Press.

Ball, S. J. (1999). Labour, learning and the economy: A “policy sociology” perspective. *Cambridge Journal of Education*, *29*(2), 195–206. doi:10.1080/0305764990290203

Bannert, M., Reimann, P., & Sonnenberg, C. (2013). Process mining techniques for analysing patterns and strategies in students’ self-regulated learning. *Metacognition and Learning, 9*(2), 161–185. doi:10.1007/s11409-013-9107-6

Buckingham Shum, S. (2012). Learning Analytics Policy Brief. UNESCO. Retrieved from http://iite.unesco.org/pics/publications/en/files/3214711.pdf

Buckingham Shum, S., de Laat, M., De Liddo, A., Ferguson, R., Kirschner, P. A., Ravencroft, A. … Whitelock, D. (2013). 1st International Workshop on Discourse-Centric Learning Analytics. *Proceedings of the 3rd International ACM Conference on Learning Analytics and Knowledge*.

Leuven, Belgium. New York, NY: ACM Press. Retrieved from http://www.solaresearch.org/events/lak/lak13/dcla13/

Buckingham Shum, S., de Laat, M., De Liddo, A., Ferguson, R., & Whitelock, D. (2014). 2nd International Workshop on Discourse-Centric Learning Analytics. *Proceedings of the 4th International ACM Conference on Learning Analytics and Knowledge*. Indianapolis, IN. New York, NY: ACM Press. Retrieved from http://dcla14.wordpress.com/

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society*, *15*(3), 3–26.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, *6*(3), 271–315. doi:10.1207/s15327809jls0603_1

Clark, D. B., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review*, *19*(3), 343–374. doi:10.1007/s10648-007-9050-7

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 134–138). Retrieved from http://dl.acm.org/citation.cfm?id=2330636

Cooper, A. (2012). *A framework of characteristics for analytics* (briefing paper Vol 1, No. 7). JISC CETIS. Retrieved from http://publications.cetis.ac.uk/wp-content/uploads/2012/12/A-Framework-of-Characteristics-for-Analytics-Vol1-No7.pdf

De Liddo, A., Buckingham Shum, S., Quinto, I., Bachler, M., & Cannavacciuolo, L. (2011). Discourse-centric learning analytics. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 23–33). New York, NY: ACM Press. doi:10.1145/2090116.2090120

De Liddo, A., Sándor, Á., & Buckingham Shum, S. (2012). Contested collective intelligence: Rationale, technologies, and a human–machine annotation study. *Computer Supported Cooperative Work (CSCW)*, *21*(4–5), 417–448. doi:10.1007/s10606-011-9155-x

de Vries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *Journal of the Learning Sciences*, *11*(1), 63–103. doi:10.1207/S15327809JLS1101_3

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. *Three Worlds of CSCL: Can We Support CSCL?*, 61–91. Retrieved from https://telearn.archives-ouvertes.fr/hal-00190230

Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. London, UK: Routledge.

Elias, T. (2011). Learning analytics: Definitions, processes and potential. Retrieved from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

Enyedy, N., & Hoadley, C. M. (2006). From dialogue to monologue and back: Middle spaces in computer-mediated learning. *International Journal of Computer-Supported Collaborative Learning*, *1*(4), 413–439. doi:10.1007/s11412-006-9000-2

Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *International Journal of Computer-Supported Collaborative Learning*, *3*(4), 447–470. doi:10.1007/s11412-008-9052-6

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

Ferguson, R., Whitelock, D., & Littleton, K. (2010). Improvable objects and attached dialogue: New literacy practices employed by learners to build knowledge together in asynchronous settings. *Digital Culture & Education*, *2*(1), 103–123.

Furberg, A., & Ludvigsen, S. (2008). Students' meaning-making of socio-scientific issues in computer mediated settings: Exploring learning through interaction trajectories. *International Journal of Science Education*, *30*(13), 1775–1799. doi:10.1080/09500690701543617

Gee, J. P. (2008). Learning and games. In K. Salen (Ed.), *The ecology of games: Connecting youth, games, and learning* (Vol. 3, pp. 21–40). Cambridge MA: MIT Press. Retrieved from http://mitpress2.mit.edu/books/chapters/0262294249chap2.pdf

Gibson, D. (2013). Psychometric considerations for digital assessments. *10th World Conference on Computers in Education*. Torun, Poland. Retrieved from http://edu.rsei.umk.pl/wcce2013/publications/v1/V1.27_008-Gibson-fullR-FPR.pdf

Gweon, G., Jun, S., Lee, J., Finger, S., & Rosé, C. P. (2011). A framework for assessment of student project groups on-line and off-line. In *Analyzing Interactions in CSCL* (pp. 293–317). Cham, Switzerland: Springer International Publishing AG. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-7710-6_14

Hammersley, M. (2010). Reproducing or constructing? Some questions about transcription in social research. *Qualitative Research*, *10*(5), 553–569. doi:10.1177/1468794110375230

Hastings, P., Hughes, S., Magliano, J. P., Goldman, S. R., & Lawless, K. (2012). Assessing the use of multiple sources in student essays. *Behavior Research Methods*, *44*(3), 622–633. doi:10.3758/s13428-012-0214-0

Haythornthwaite, C., & de Laat, M. (2010). Social networks and learning networks: Using social network perspectives to understand social learning. *7th International Conference on Networked Learning*. Retrieved from http://celstec.org.uk/system/files/file/conference_proceedings/NLC2010_Proceedings/abstracts/PDFs/Haythornwaite.pdf

Hennessy, S., & Warwick, P. (2010). Research into teaching with whole-class interactive technologies. *Technology, Pedagogy and Education*, *19*(2), 127–131. doi:10.1080/1475939X.2010.491211

Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning*, 1–20. doi:10.1007/s10758-014-9226-4

Johnson, J. A. (2013). *Ethics of data mining and predictive analytics in higher education*. Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2156058

Joshi, M., & Rosé, C. P. (2007). Using transactivity in conversation for summarization of educational dialogue. *SLaTE workshop on speech and language technology in education*. Retrieved from http://www.isca-speech.org/archive_open/archive_papers/slate_2007/sle7_053.pdf

Knight, S. (2013). Creating a supportive environment for classroom dialogue. In S. Hennessy, P. Warwick, L. Brown, D. Rawlins, & C. Neale (Eds.), *Developing interactive teaching and learning using the IWB*. Maidenhead, UK: Open University Press.

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Collaborative sensemaking in learning analytics. *CSCW and education workshop: Viewing education as a site of work practice, 16th ACM*

Conference on Computer Support Cooperative Work and Social Computing (CSCW 2013), 23 February 2013, San Antonio, TX.

Knight, S., Buckingham Shum, S., & Littleton, K. (2014). Epistemology, assessment, pedagogy: Where learning meets analytics in the middle space. *Journal of Learning Analytics, 1*(2), 23–47.

Knight, S., & Littleton, K. (2013). Discourse, computation and context: Sociocultural DCLA revisited.*1st International Workshop on Discourse-Centric Learning Analytics*. Leuven, Belgium. Retrieved from http://oro.open.ac.uk/36640/

Knight, S., & Littleton, K. (2015). Thinking, interthinking, and technological tools. In R. Wegerif, J. C. Kaufman, & L. Li (Eds.), *The Routledge international handbook of research on teaching thinking*. Abingdon, UK: Routledge.

Král, P., & Cerisara, C. (2012). Dialogue act recognition approaches. *Computing and Informatics*, *29*(2), 227–250.

Lid, V., & Suthers, D. D. (2003). Supporting online learning with an artifact-centered cross-threaded discussion tool. *Proceedings of the International Conference on Computers in Education* (pp. 319–314). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.70.8914&rep=rep1&type=pdf

Littleton, K., & Howe, C. (2010). *Educational dialogues: Understanding and promoting productive interaction*. Abingdon, UK: Routledge.

Littleton, K., & Mercer, N. (2013). *Interthinking: Putting talk to work*. Abingdon, UK: Routledge.

Littleton, K., & Whitelock, D. (2005). The negotiation and co-construction of meaning and understanding within a postgraduate online learning community. *Learning, Media and Technology*, *30*(2), 147–164.

Martin, J. R., & White, P. R. (2005). *The language of evaluation*. Basingstoke, UK/New York, NY: Palgrave Macmillan.

Medina, R., & Suthers, D. D. (2009). Using a contingency graph to discover representational practices in an online collaborative environment. *Research and Practice in Technology Enhanced Learning*, *4*(3), 281–305.

Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences*, *17*(1), 33–59. doi:10.1080/10508400701793182

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. Abingdon, UK: Routledge.

Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. (2002). Accountable talk: Classroom conversation that works. Pittsburgh, PA: University of Pittsburgh.

Mislevy, R., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centred design, psychometrics, and educational data mining. *Journal of Educational Data Mining*, *4*(1), 11–48.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge building discourse explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, *60*(5), 903–921. doi:10.1007/s11423-012-9265-2

Palau, R. M., & Moens, M.-F. (2009). Argumentation mining: The detection, classification and structure of arguments in text. *Proceedings of the 12th international conference on artificial intelligence*

(2015). Discourse-Centric Learning Analytics: Mapping the Terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

and law (pp. 98–107). New York, NY: ACM Press. Retrieved from http://dl.acm.org/citation.cfm?id=1568246

Plato. (1999). *Phaedrus*. (B. Jowett, Trans.). Retrieved from http://www.gutenberg.org/ebooks/1636

Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 239–257. doi:10.1007/s11412-009-9070-z

Reimann, P., Markauskaite, L., & Bannert, M. (2014). e-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology, 45*(3), 528–540. doi:10.1111/bjet.12146

Resnick, L. B. (2001). Making America smarter: The real goal of school reform. In A. L. Costa (Ed.), *Developing minds: A resource book for teaching thinking* (3rd ed.) (pp. 3–6). Alexandria, VA: Association for Supervision and Curriculum Development.

Rosé, C. P., & Tovares, A. (2015). What sociolinguistics and machine learning have to say to one another about interaction analysis. In L. Resnick, C. Asterhan, & S. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.

Rosé, C. P., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *3*(3), 237–271.

Rosen, D., Miagkikh, V., & Suthers, D. D. (2011). Social and semantic network analysis of chat logs. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 134–139). New York, NY: ACM Press. doi:10.1145/2090116.2090137

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, *5*(1), 43–102.

Shaffer, D. W. (2008). *How computer games help children learn*. Basingstoke, UK/New York, NY: Palgrave Macmillan.

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., … Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media, 1*(2), 33–53. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/ijlm.2009.0013

Sie, R. L. L., Ullmann, T. D., Rajagopal, K., Cela, K., Bitter-Rijpkema, M., & Sloep, P. B. (2012). Social network analysis for technology-enhanced learning: Review and future directions. *International Journal of Technology Enhanced Learning*, *4*(3), 172–190. doi:10.1504/IJTEL.2012.051582

Simsek, D., Buckingham Shum, S., Sándor, A., De Liddo, A., & Ferguson, R. (2013). XIP Dashboard: Visual analytics from automated rhetorical parsing of scientific metadiscourse. *1st International Workshop on Discourse-Centric Learning Analytics*, 8 April 2013, Leuven, Belgium. Retrieved from http://libeprints.open.ac.uk/37391/1/LAK13-DCLA-Simsek.pdf

Sionti, M., Ai, H., Rosé, C. P., & Resnick, L. (2011). A framework for analyzing development of argumentation through classroom discussions. *Educational Technologies for Teaching*

Argumentation Skills, Bentham Science. Retrieved from http://www.cs.cmu.edu/~huaai/pub/Sionti.pdf

Stahl, G. (2010). Group cognition as a foundation for the new science of learning. In M. S. Khine & I. M. Saleh (Eds.), *New science of learning: Cognition, computers and collaboration in education.* London: Springer.

Stahl, G. (2013). Transactive discourse in CSCL. *International Journal of Computer-Supported Collaborative Learning*, *8*(2), 145–147. doi:10.1007/s11412-013-9171-6

Stahl, G., & Rosé, C. P. (2011). Group cognition in online teams. *Theories of Team Cognition: Cross-Disciplinary Perspectives.* New York, NY: Routledge/Taylor & Francis. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.190.3215&rep=rep1&type=pdf

Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D. … Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, *26*(3), 339–373.

Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. *45th Hawaii International Conference on System Science (HICSS)* (pp. 3368–3377). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6149232

Suthers, D. D., Lund, K., Rosé, C. P., Teplovs, C., & Law, N. (Eds.). (2013). *Productive Multivocality in the Analysis of Group Interactions* (Vol. 15). New York, NY: Springer. Retrieved from http://www.springer.com/education+%26+language/learning+%26+instruction/book/978-1-4614-8959-7

Suthers, D. D., & Medina, R. (2010). The temporal development of representational practices: Implications for theory and analysis of situated learning. *43rd Hawaii International Conference on System Sciences (HICSS)* (pp. 1–10). doi:10.1109/HICSS.2010.397

Suthers, D. D., Vatrapu, R., Medina, R., & Dwyer, N. (2007). Information sharing and interaction in collaborative convergence. In T. Hirashima, U. Hoppe, & S. S.-C. Young (Eds.), *Supporting Learning Flow through Integrative Technologies* (Vol. 162, pp. 167–174). IOS Press.

Suthers, D. D., & Verbert, K. (2013). Learning analytics as a “middle space.” *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 1–4). New York, NY: ACM Press. doi:10.1145/2460296.2460298

Tablan, V., Roberts, I., Cunningham, H., & Bontcheva, K. (2013). GATECloud.net: A platform for large-scale, open-source text processing on the cloud. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *371*(1983), 20120071. doi:10.1098/rsta.2012.0071

Wegerif, R. (2011). Towards a dialogic theory of how children learn to think. *Thinking Skills and Creativity*, *6*(3), 179–190. doi:10.1016/j.tsc.2011.08.002

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer supported collaborative learning. *Instructional Science*, *33*(1), 1–30. doi:10.1007/s11251-004-2322-4

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*(1), 71–95. doi:10.1016/j.compedu.2005.04.003

Wells, G. (1999). *Dialogic inquiry: Towards a sociocultural practice and theory of education*. Cambridge, UK: Cambridge University Press.

Wells, G., & Claxton, G. (2002). *Learning for life in the 21st century: Sociocultural perspectives on the future of education*. Hoboken, NJ: Wiley-Blackwell.