

Editorial

Charalampos Karagiannidis and Sabine Graf

Welcome to the Bulletin of the IEEE Technical Committee on Learning Technology, Volume 18, Number 1 issue.

This issue includes articles related to the theme of *Technology-Enhanced Science, Technology, Engineering and Math Education*. Lasica et al highlight different types of laboratories, trends, and areas of research to take into account in the use or adoption of virtual laboratories for the STEM education, and discuss examples of studies or tools developed for remote laboratories in STEM education. El-Demerdash et al present the design and evaluation of digital “c-book units”, i.e. digital books produced within a socio-technological environment allowing meshing narratives with interconnected, interactive and dynamic digital artefacts (widgets), to promote Creative Mathematical Thinking. Liu et al present a study which explored the scalability of conversation-based assessment to measure constructs related to scientific reasoning, where virtual students interact with students to demonstrate their abilities to use evidence to support a prediction.

The issue also includes a review by A. Wagner of the book “Minds Online: Teaching Effectively with Technology” published by Harvard University Press.

We sincerely hope that the issue will help in keeping you abreast of the current research and developments in Learning Technology. We also would like to take the opportunity to

invite you to contribute your own work (e.g. work in progress, project reports, dissertation abstracts, case studies, event announcements) in this Bulletin, if you are involved in research and/or implementation of any aspect of advanced learning technology. For more details, please refer to the author guidelines at <http://www.ieeetclt.org/content/authors-guidelines>.

Special theme of the next issue: **Adaptive and Intelligent Learning Systems**

Deadline for submission of articles: **July 22, 2016**

Articles that are not in the area of the special theme are most welcome as well and will be published in the regular article section.

Charalampos Karagiannidis (karagian@uth.gr) is an Associate Professor at the Department of Special Education of the University of Thessaly, Greece.

Sabine Graf (sabineg@athabasca.ca) is an Associate Professor at the School of Computing and Information Systems of the Athabasca University, Canada.

Research Challenges in future laboratory-based STEM Education

I. E. Lasica, K. Katzis, M. Meletiou-Mavrotheris, and C. Dimopoulos

Abstract— Science, Technology, Engineering and Mathematics (STEM) Education is strongly supported by laboratory experiments, which constitute a significant educational tool to promote better understanding among students. Virtual and Remote Labs have become an interesting field of study during the past two decades. Several studies have been devoted to the different types of labs arising with the evolution of technology and their efficacy within all educational levels. In this conceptual paper, some current trends are identified after a brief overview of the state-of-the-art concerning laboratory experiments within secondary / higher education and lifelong learning. Moreover, challenges are discussed and potential future areas of research are identified. The overarching aim of these research areas is to assess the usefulness, effectiveness, acceptability, adaptability, social impact and cost of the described types of laboratories to fulfill the high standards required for a career in a STEM related profession.

Index Terms—Augmented Reality, Distance Learning, Remote Labs, Virtual Labs.

I. INTRODUCTION

Science, Technology, Engineering and Mathematics (STEM) Education is an area receiving increased attention during the past decades not only within the educational but also the commercial and the entertainment field. Experiments constitute a critical part of the engineering education since they attract students' interest, promote better understanding of the taught theories, and provide opportunities for the acquisition of practical knowledge.

The evolution of technology offers new opportunities in lab-based STEM education, overcoming the drawbacks of using traditional hands-on labs such as high costs, limited availability, expensive maintenance, etc. [1] [2]. Modern approaches have been initiated including remoteness (remote labs), virtuality (virtual labs) and recently immersion (augmented reality labs) [3] [4]. The new types of labs provide important benefits, including reduced costs, enhanced

availability and accessibility, large-scale observability, and increased safety [5]. A significant number of studies underline the potential of laboratory-based learning for the enhancement of the educational process within traditional classrooms or within an e-learning and/or a b-learning context [3] [6] [7].

In spite of the numerous benefits and the popularity of laboratory experiments within STEM education, there are still some drawbacks that need to be addressed concerning remote labs [8]. Some of the limitations mentioned in different studies are the following: (a) virtual environments are imitations of reality and thus suffer from a weak realistic representation of real equipment [9]; (b) virtual and augmented reality labs enhance students' isolation [3]; (c) since data are obtained from theoretical calculations based on mathematical models, students often fail to take into account instrument or other errors inherent in practical experimentation [9].

The main research question of this work, steamed from the NeReLa Project's objectives (a TEMPUS project carried out with partners from Serbia and EU - Slovenia, Spain, Portugal, Cyprus and Greece), including teachers' inspiration (Higher and Secondary Vocational Education) in Serbia to take advantage of remote labs as a means of enhancing engineering teaching and making it more effective and attractive. In this conceptual paper, we identify current trends in laboratory-based STEM education by drawing upon existing research articles and bibliometric analyses in the field of virtual and remote labs in education [2] [3] [6] published in international high-rated journals in the field of education, computer sciences, and information technology. The analysis was exploratory and based on 30 papers which were reviewed to draw the conclusions while results of recent extensive analyses in the field were referred as main foundations to be extended. The keywords used in our search for relevant literature included terms like Virtual Laboratory, Remote Laboratory, Augmented Reality and STEM. Further selection criteria for articles reviewed included the following: (a) published during the last ten years (2006-2016) so as to reflect recent trends, (b) explicitly addressing the STEM field; and (c) reporting on technologies related to laboratory education.

After a brief overview of the state-of-the-art concerning laboratory experiments within STEM-related education and distance learning, we describe the different lab types identified within studies so far and their main characteristics. Moreover,

Manuscript received May 9, 2016.

I. E. Lasica is with the School of Arts & Education Sciences, European University of Cyprus, 6, Diogenis Str., Engomi, P.O. Box: 22006, 1516 Nicosia-Cyprus (e-mail: e_ilona@outlook.com).

K. Katzis and C. Dimopoulos are with the School of Sciences, European University of Cyprus, 6, Diogenis Str., Engomi, P.O. Box: 22006, 1516 Nicosia-Cyprus (corresponding author to provide phone: +35722713296; fax: +35722713013; e-mail: K.Katzis@euc.ac.cy).

M. Meletiou-Mavrotheris is with the School of Arts & Education Sciences, European University of Cyprus, 6, Diogenis Str., Engomi, P.O. Box: 22006, 1516 Nicosia-Cyprus (e-mail: M.Mavrotheris@euc.ac.cy).

some current trends are distinguished and some additional challenges are pointed out as identified in the bibliography. These challenges could become areas of further scientific research interest, referring to the usefulness, effectiveness, acceptability, adaptability, social impact and cost of the described types of laboratories. More specifically, Section 2 introduces the reader to various types of laboratory-based education, Section 3 discusses laboratory-based STEM education and presents a number of examples of remote laboratories, and Section 4 describes the research challenges and research directions that need addressing. Finally, conclusions are drawn.

II. TYPES OF LABORATORY-BASED EDUCATION

As already mentioned, experiments constitute a critical part of STEM since they enhance students' understanding of the theories taught. However, an important question arising from the literature review is the variety of types of labs one can identify when implementing an experiment. This section describes the specific characteristics of each type of lab, as

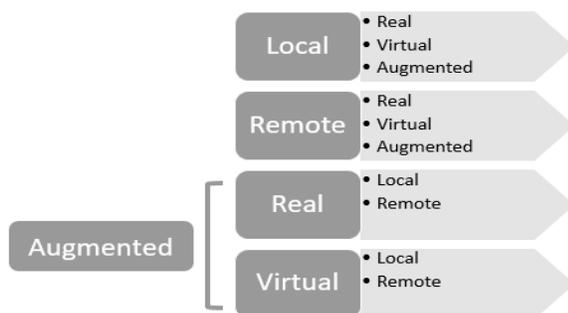


Fig. 1. Lab categories concerning resources access and physical nature.

recognized by us and categorized by Villalba et al. [4] according to: (a) the way resources are accessed (remote or local) and (b) the physical nature of the lab (simulated or real). Taking into consideration a blended category mentioned within recent studies as “augmented”, which combines virtual with real equipment, we have added it as an extra category [5] [10].

According to the way resources are accessed, labs could be distinguished between *local* and *remote*. More specifically, traditional hands-on labs, existing in a specific physical location (within a school, a university or anywhere experiments can take place), could be considered as local labs where students implement an experiment in real time (synchronous) mode. Usually only a single student or a small group of students participate in the implementation of the experiments. Local Labs, as presented in Figure 1, could include Real Local Labs (RLLs), Virtual Local Labs (VLLs) and Augmented Reality Local Labs (ARLLs), referring to the physical nature of the lab [11]. Remote labs differ in such a way that they allow students to perform experiments in real-time (synchronously) or at a time of their own choice (asynchronously) through the Internet [12]. Remote Labs, as presented in Figure 1, could include Real Remote Labs

(RRLs), Virtual Remote Labs (VRLs) and Augmented Reality Remote Labs (ARLLs) referring to the physical nature of the lab [11]. Figure 1 illustrates the combinations based on the categories mentioned.

III. LABORATORY-BASED STEM EDUCATION

There is a general agreement on the benefits of labs to teach science and engineering [2] [13] [14]. Local and Remote labs, as presented in Figure 1, are not exclusive alternatives, but they are valuable educational resources which could be combined in one integral and complementary curriculum concerning STEM education [15] [16].

A number of studies cast doubt on the effectiveness of VRLs within the educational process, underlining that RLLs are necessary to acquire haptic skills and instrumentation awareness, skills difficult to achieve through VRLs and ARLLs [16] [17]. These studies emphasize the need for employing a combination of the different types of labs during different educational phases, e.g. VRLs during preparation, RLLs in live lectures in order to implement real experiments and finally, RRLs to support repetitive experimentation [16].

The most common consensus within studies concerning laboratory-based STEM education, is that virtual labs constitute a useful tool for students to implement experiments without space or time limitations, as opposed to local laboratories which are limited to a specific place and time constraints [3] [14]. Some studies also mention that bringing remote experiments into classrooms (secondary, higher, secondary-vocational education etc.) is an efficient way to achieve better learning outcomes [18], and to attract students to STEM related fields of study and careers [19].

As part of our research activities at EUC, we have been involved in a Tempus project called Network of Remote Laboratories (NeReLa) along with our partners from Serbia, Slovenia, Spain, Portugal, etc. [20]. NeReLa aimed at inspiring university teachers (Higher Education) as well as school teachers of secondary vocational education in Serbia, to take advantage of remote labs as a means of enhancing engineering teaching and making it more effective and interesting [21]. Some remote and virtual experiments were developed and carried out within the laboratories of partner universities, thus enhancing the NeReLa library of remote experiments for further use. Examples of remote laboratories demonstrated by partner universities are described below:

(a) **University of Maribor**: Two solutions for RRLs were proposed: (i) an in-house developed remote control laboratory and (ii) an external remote laboratory, based on CEyeClon network [22]. Both were integrated with a Moodle Learning Management System for the booking process. The university's RRLs are included into three networks: (1) EDIPE [23], (2) E-PRACTIC [24] and (3) SustEner [25].

(b) **University of Deusto**: A remote laboratory for multidisciplinary experimentation was presented, based on an open-source Remote Laboratory Management System. WebLab-Deusto was designed to enhance the development of RRLs supporting all necessary features to deploy a remote lab, such as authentication, authorization, scheduling, load

balancing, user tracking, sharing and administrative tools [18] [20]. Three main experiments were performed within WebLab-Deusto, VISIR LXI (with real electronic circuits), WebLab-Bot (for programming a microcontroller based on a popular mobile robot) and WebLab-Box (for performing Field Programmable Gate Array (FPGA) based embedded systems).

(c) **Faculty of Engineering of University of Porto (FEUP):** The first remote laboratory of sensitive type (Meteorological Station) was provided on a regular basis in 1998. Many projects' funding have promoted the sharing of remote labs and technical expertise within the university's repository since then. In 2010 a wide goal was set: the creation of a consortium in Portugal of online experimentation and a relative platform, to initialize a database for retrieving all national initiatives in remote and virtual labs, following the International Association of Online Engineering (IAOE) ontology [18]. Such a repository would be able to promote collaborative activities and foster the sharing of resources at national level and generally, in Portuguese speaking countries. Within this goal, FEUP has been sharing relative resources with national and international institutions.

The abovementioned case studies mostly referred to technologies concerning RRLs, whereas some of the experiments introduced Augmented Reality similarly to other studies [5]. NeReLa is modelled based on the available infrastructure for remote experimentation at partner universities and, by allowing people from other institutions to contribute towards developing new remote engineering experiments, is expected to enrich LiReX library, a "living organism" collection of remote experiments within European institutions.

IV. RESEARCH CHALLENGES AND RESEARCH DIRECTIONS

There is an obvious trend in education towards the extensive use of different types of labs, generally in the field of STEM and especially in the field of engineering [1]. Laboratory-based education has been spreading across all educational levels (primary schools, higher education, vocational learning, universities etc.) from self-education to formal curricula [3] [8]. Moreover, there are studies that focus on virtual learning environments concerning special categories of students (e.g. students with autism) [26]. The use of virtual environments can enhance students to acquire more specific technical skills, which can be considered as essential 21st century skills for lifelong learning.

We believe that there are many research areas worth investigating related to laboratory-based STEM education [18] [20]. Heradio et al. [3] specify four main areas for intensive research which could be developed in the near future: (a) efficient combination of virtual and remote labs, (b) collaborative learning, (c) VRLs assessment, and (d) VRLs sharing. Additionally to these specific areas, some studies have identified the need to build platforms with the capability of adaptation to different users' profiles [15] [27], which should take into consideration the specific needs of a user and his/her background (such as educational level (e.g. secondary

education), knowledge field (e.g. engineering), role (e.g. student, trainee) and suggest relative experiments.

Moreover, the assessment of the educational effectiveness of the different types of laboratories presented in this paper is an interesting topic for further research, in order to define whether the use of innovative ways to implement experiments can lead to higher performing students [13] [18]. Establishing RRLs within current formal curricula would bring an innovative pedagogical approach to STEM education [16] and maybe introduce the basis for a fully online Engineering degree. This could lead to new Instructional Design approaches within Distance and Blended Learning, based on the usage of RRLs, VRLs and ARRLs. Additionally, collaboration within such approaches should be enhanced, through Learning Management Systems (LMS) like Moodle and social networks, as many studies mention isolation as an important drawback of RRLs [28].

Another topic worth mentioning for further research, taking into consideration the objectives of projects such as NeReLa, could be the increase in the number of students choosing to follow STEM related studies. Attractiveness of STEM has already improved since relevant games and equipment are used not only for educational but also for entertainment purposes, providing a strong indication that industry approves this new type of teaching and learning methodology. However, an interesting study could focus on the impact of the different types of labs on students' professional choices.

Concerning VRLs sharing and reuse - a hot topic for further study [3] - methodologies of forming specific metadata could be proposed for organizing such repositories and defining the way and the context in which they could be used to facilitate search, retrieval and reuse processes [28]. Finally, measuring the cost of modern labs (including setup, running and maintenance) would be beneficial for academic institutions, since it would enable them to define if the running costs of these labs could decrease and as a result lower tuition fees could be achieved to enhance access to STEM education.

We believe that research directions with respect to STEM laboratories should be answering critical and practical questions at the same time in order to assist academics and academic institutions to choose the best possible way of implementing a laboratory-based pedagogical approach, so as to enhance its learning impact. Such questions include the following: *Which laboratory is most suitable for a STEM course / degree? What is the expected impact on students' skills? Does this approach fulfill the learning outcomes set at the beginning of the course? Is it too complicated and expensive to implement? Does the student require expensive equipment to be able to participate? What is the industry's opinion regarding this matter? Would employers prefer someone who has conducted most of his/her experiments remotely or would they choose a candidate with conventional studies? Can the laboratories be frequently updated or is this difficult to do due to the cost and complexity of the system to make changes to the current syllabus?* These are some of the questions someone might have to attend to before selecting between the different types of labs currently available.

V. CONCLUSIONS

To conclude, in this conceptual paper, we drawn upon existing bibliometric analyses in the field of virtual and remote labs in education and other relevant literature, in order to specify: (a) the different lab types identified within conducted to-date, and (b) some additional challenges to these mentioned within the existing bibliography that could become areas of further scientific interest, referring to useful research areas related to the usefulness, effectiveness, acceptability, adaptability, social impact and cost of the described types of laboratories. There is an obvious trend in education towards the extensive use of RRLs, VRLs and ARRLs, generally in the field of STEM and especially in the field of engineering. Finally, these topics could be main areas of focus for interesting research studies to be conducted in the future.

ACKNOWLEDGMENT

Authors would like to thank the European University Cyprus and the NeReLa Project for the financial support concerning this study.

REFERENCES

- [1] J. Ma, J. Nickerson, "Hands on, simulated and remote laboratories: A comparative literature review", *ACM Computing Surveys*, vol. 38, no.3, 2006.
- [2] L. Gomes, S. Bogosyan, "Current trends in remote laboratories", *IEEE Transactions on Industrial Electronics*, vol. 56, no. 12, pp. 4744-4756, 2009.
- [3] R. Heradio, L. de la Torre, D. Galan, F. J. Cabrerizo, E. Herrera-Viedma, S. Dormido, "Virtual and remote labs in education: A bibliometric analysis", *Computers & Education*, pp. 9814-38, 2016.
- [4] C. M. Villalba, A. U. Moraleda, S. D. Bencomo, "Distance Education On Virtual-Lab Implementation For Science Teachers", *RIED Revista Iberoamericana de Educación a Distancia*, 11(2), 2008.
- [5] S. Odeh, S. A. Shanab, M. Anabtawi, R. Hodrob, "A Remote Engineering Lab Based on Augmented Reality for Teaching Electronics", *International Journal of Online Engineering*, vol. 9, 2013.
- [6] C. Gravier, J. Fayolle, B. Bayard, M. Ates, J. Lardon, "State of the art about remote laboratories paradigms - foundations of ongoing mutations", *International Journal of Online Engineering*, vol. 4, no. 1, pp. 19-25, 2008.
- [7] R. Krneta, M. T. Restivo, A. Rojko, D. Urbano, "Evaluation of remote experiments by different target groups: NeReLa project case study", in *2016 13th International Conference on Remote Engineering and Virtual Instrumentation (REV)*, pp. 326-331, IEEE, 2016.
- [8] M. Zappatore, A. Longo, M. Bochicchio, "The bibliographic reference collection GRC2014 for the online laboratory research community", in: *Conference on Remote Engineering and Virtual Instrumentation*, Bangkok, Thailand, pp. 24-31, 2015.
- [9] S. Odeh, "A Web-Based Remote Lab Platform with Reusability for Electronic Experiments in Engineering Education", *iJOE*, vol. 10, no. 4, pp. 40-45, 2014.
- [10] H., Vargas, G., Farias, J., Sanchez, S., Dormido, F. Esquembre, "Using Augmented Reality in Remote Laboratories", *International Journal of Computers Communications & Control*, vol. 8, no. 4, pp. 622-634, 2013.
- [11] I. Lasica, K. Katzis, M. Meletiou-Mavrotheris, C. Dimopoulos, "STEM Education: Current and future trends in laboratory-based education", in *3rd International Conference on Electrical, Electronic and Computing Engineering IcETRAN*, to be presented June 2016.
- [12] N. Aliane, "LABNET: A Remote Control Engineering Laboratory," *Universidad Europea de Madrid, DACA, Madrid, Spain*, 17 November, 2006.
- [13] T. de Jong, M. C. Linn, Z. C. Zacharia, "Physical and virtual laboratories in science and engineering education", *Science*, vol. 340, pp. 305-308, 2013.
- [14] T. Karakasidis, "Virtual and remote labs in higher education distance learning of physical and engineering sciences", in *Global Engineering Education Conference (EDUCON)*, pp. 798-807, IEEE, 2013.
- [15] A. Cardoso, M. Vieira, P. Gil, "A Remote and Virtual Lab with Experiments for Secondary Education, Engineering and Lifelong Learning Courses", *iJOE*, vol. 8, no. S2, pp. 49-54, 2012.
- [16] M. Abdulwahed, Z. K. Nagy, "The trilab, a novel ict based triple access mode laboratory education model", *Computers & Education*, vol. 56, pp. 262-274, 2011.
- [17] J. Y. Ma, J. S. Choi, "The Virtuality and Reality of Augmented Reality", *Journal of Multimedia*, vol. 2, no.1, pp.32-37, 2007.
- [18] R. Radhamani, H. Sasidharakurup, G. Sujatha, B. Nair, K. Achuthan, S. Diwakar, "Virtual labs improve student's performance in a classroom", in *E-Learning, E-Education, and Online Training*, pp. 138-146), Springer International Publishing, 2014.
- [19] R. Krneta, A. Rojko, O. Dziabenko, T. Restivo, "NeReLa project: Building network of remote labs using EU best practice, in *Tecnologias Aplicadas a la Enseñanza de la Electronica (Technologies Applied to Electronics Teaching)(TAE)*", pp. 1-8, IEEE, 2014.
- [20] Network of Remote Laboratories, NeReLa, available at <http://www.nerela.kg.ac.rs>, 2016
- [21] R. Krneta, "Introduction to the NeReLa project", in *Remote Engineering and Virtual Instrumentation (REV), 2014 11th International Conference on*, pp. 386-388, IEEE, 2014.
- [22] CEyeClon network, RRLs, available at <http://ceyeclon.com/>, 2016.
- [23] Power WebLab DC components Library, available at <http://www.pemcweblab.com/>, 2016.
- [24] E-pragmatic Project, available at <http://www.e-pragmatic.eu/>, 2016
- [25] Sustainable Energy, available at <http://www.sustener.eu/>, 2016
- [26] C. Volioti, T. Tsiatsos, S. Mavropoulou, C. Karagiannidis, "VLSS--Virtual Learning and Social Stories for Children with Autism", in *2014 IEEE 14th International Conference on Advanced Learning Technologies (ICALT)*, pp. 606-610, IEEE, 2014.
- [27] S. Graf, C. Ives, "A flexible mechanism for providing adaptivity based on learning styles in learning management systems", in *Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on*, pp. 30-34, IEEE, 2010.
- [28] C. A. Jara, F. A. Candelas, F. Torres, S. Dormido, F. Esquembre, O. Reinoso, "Real-time collaboration of virtual laboratories through the internet", *Computers & Education*, vol. 52, no. 1, pp. 126 - 140, 2009.
- [29] P. Zervas, A. Kalamatianos, E. Tsourlidaki, S. Sotiriou, D. G. Sampson, "A methodology for organizing virtual and remote laboratories", in *Digital systems for open access to formal and informal learning*, pp. 235-255, Springer International Publishing, 2014.

Design and Evaluation of Digital Resources to Enhance Creative Mathematical Thinking in a Biomathematics Context

Mohamed El-Demerdash, Christian Mercat, Jana Trgalova, Nataly Essonnier and Pedro Lealdino Filho

Abstract—In this article, we present the design and the *a priori* evaluation of digital resources, carried out within the framework of the MC Squared project (<http://mc2-project.eu>). These resources are the so-called “c-book units” (“c” for creative), which are digital books produced within a socio-technological environment allowing meshing narratives with interconnected, interactive and dynamic digital artefacts (widgets), to promote Creative Mathematical Thinking (CMT). We have developed a methodology to support creative design and the *a priori* evaluation of the CMT potential of the c-book units’ affordances. We illustrate here the design and evaluation of a c-book unit within a biomathematics context aimed at supporting learning and mathematical reasoning, using a wide range of digital widgets. We discuss the design choices resulting in the resource affordances to promote mathematical creativity in terms of personalized non-linear path, constructivist approach, and meta-cognition based activities, among others. We present experts’ *a priori* evaluation of the resource CMT potential.

Index Terms—Creativity, mathematical creativity, mathematics education, technology-enhanced learning.

I. INTRODUCTION

PROMOTING Creative Mathematical Thinking (CMT) to enable innovation is a central goal of the European Union [1]. CMT is a highly valued asset in industry and addresses current and future economic challenges. It is seen as an individual and collective construction of mathematical meanings, norms and uses in novel and useful ways [2], [3], which can be of relevance to a larger (academic, learning, professional, or other) community.

New exploratory and expressive digital media created by new designers are providing users with access to and potential for engagement with CMT in unprecedented ways taking into account the co-evolution of teaching and technology [4] or the constructionist approach [5].

The MC Squared project (<http://mc2-project.eu/>) aims at

Manuscript received May 10, 2016. This work was supported in part by the European Union Seventh Framework Programme (FP7/2007-2013) under the grant agreement n° 610467 - project “MC Squared” ICT STREP 100712. <http://mc2-project.eu> This publication reflects only the authors’ views and the Union is not liable for any use that may be made of the information contained therein.

The authors are with S2HEP EA4148 Université Claude Bernard – Lyon 1, France. M. El-Demerdash is furthermore with Menoufia University, Egypt (corresponding author e-mail: m_eldemerdash70@yahoo.com).

designing and developing an intelligent computational environment, called C-book technology, to support stakeholders from creative industries involved in the production of media content for educational purposes to engage in collective forms of creative design of appropriate digital media. The C-book technology provides an authoring dynamic environment extending e-book technologies to include diverse dynamic widgets, an authoring data analytics engine and a tool supporting asynchronous collaborative design of educational resources, called “c-books”. The specificity of this tool lies in the tight integration and possible cross-communication between different types of widgets coming from existing standalone software using an open Application Protocol Interface (API)¹, allowing for intelligent adaptive feedback and instance storage (the user comes back to the activity in the actual learning state s/he left it). The project focuses on studying processes of collaborative design of digital media intended to enhance CMT. The article then has twofold objectives. First, it highlights, through the development of “Math for Biology” c-book unit, the design choices to foster CMT in its users. Second, it enlightens, through experts’ *a priori* evaluation, the CMT affordances of this c-book unit.

II. CREATIVE MATHEMATICAL THINKING

Based on a literature review and prior studies led by researchers involved in the project related to studying creativity and mathematical creativity [6-11] a definition that reflects our vision of CMT has been adopted which defines it as a directed intellectual activity generating novel mathematical ideas or responses over the known or familiar ones in a non-routine mathematical situation, relatively to a pedagogical context. Drawing on Guilford’ model [12], the generation of new ideas shows the abilities of fluency, flexibility, originality/novelty, and elaboration that are defined as follows:

- **Fluency** means the individual’s ability to pose or come up with many mathematical ideas or configurations related to a mathematical problem or situation in a short time.
- **Flexibility** refers to the individual’s ability to vary the approach or suggest a variety of different methods toward a mathematical problem or situation.

¹ Software candidate for html-5 integration shall follow the Application Protocol Interface specification described on <http://mc2-project.eu>

- **Originality** means the individual's ability to try novel or unique approaches toward a mathematical problem or situation.

- **Elaboration** is the individual's ability to redefine a single mathematical problem or situation to create others, by changing one or more aspects by substituting, combining, adapting, altering, expanding, eliminating, rearranging, or reversing and then speculating on how this single change would have a ripple effect on other aspects of the problem or the situation at hand.

III. THE DESIGN OF "MATH FOR BIOLOGY" C-BOOK UNIT

"Math for Biology" c-book unit² proposes a sequence of activities aiming at supporting learning and mathematical reasoning, using a wide range of widgets useful in biology such as EpsilonWriter (Dynamic Algebra System - DAS), Cinderella and GeoGebra (Dynamic Geometry System - DGS). Cross-widget communication³ allows the user to send messages such as geometric or algebraic information between EpsilonWriter and Cinderella. Some open-ended activities are designed to train a high degree of CMT such as reflecting on models, modelling spirals and code tweaking in a half-baked logo microworld. Such a tight integration among these software factories would not be possible in existing design frameworks: the open API allowed to insert instances of these software on the same page so that the user doesn't have to switch among these software, and their common state is saved in a bundled live document, but more importantly, the widgets communicate so that the work in one widget influences the other ones, especially in intelligent feedback aimed at unblocking students learning. The analysis of the usage of the c-book is as well eased by automatic logging of activities, tailored to specific requests from teachers in a much more detailed and integrated way than in usual Learning Management Systems (LMS), because usage is traced to the level of individual widgets.

The c-book unit consists of four sections interrelated through a narration about a children visiting a Fauna and Flora park. Section 1 introduces biomathematics, its meaning and its study fields and ends up with giving users the opportunity to pick and choose among different activities on biomathematics proposed in sections 2 - 4 according to their knowledge and interest. Section 2, demonstrates the need of mathematical approximation as a tool to model biological traits, and enables to reflect on magnitudes. In section 3, golden ratio is a pretext search, for doing mathematics by looking at its occurrences in nature, paintings, architecture and Fibonacci numbers. The c-book unit ends with section 4, where users are exposed to different kinds of spirals in nature. Users have the opportunity to upload their pictures of spirals (Fig. 1) and try out some constructs that might fit them, using different digital widgets either based on algorithms or algebra. Finally, they can build their own spiral-based artwork reminding of Celtic art.

² <http://6.mc2dme.appspot.com:80/dwo/dwo.jsp?profile=78&language=en&courseViewNr=5723809264959488>

³ Cross widget communication is a way to allow to send messages between widgets

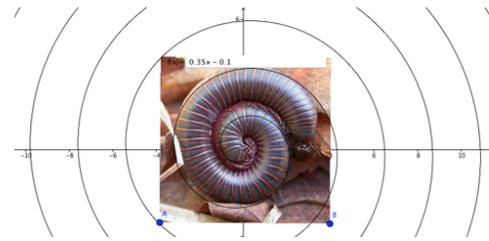


Fig. 1. A DGS widget enabling students to upload their own spiral pictures and try out some formulae that fit them.

IV. DESIGN CHOICES AND RATIONALE

The designers in the project share a common interest of producing digital resources promoting CMT. The French Community of Interest (CoI) [12] designing "Math for biology" c-book unit is a heterogeneous group gathering mathematics teachers, teacher educators, computer scientists and researchers in mathematics education.

A. CMT affordances

The designers focused on supporting the development of the CMT cognitive components (fluency, flexibility, originality and elaboration), social and affective aspects on the user's side through a relevant use of specific widgets. The C-book technology boasts affordances that can promote CMT. To take a technical metaphor, a door has a specific use; a knob is an affordance that allows for an acceptable use of the door, allowing to open and close it. Similarly, the activities in our c-book unit engage the users in compelling problems and the widgets that we setup allow for creative ways to think about them, investigate and solve them. The c-book unit opens up as well an open space for reflection and problem posing. The description of three instances of CMT affordances follow in Section E.

B. Storytelling Approach and Personalized Non-linear Path

Throughout all the activities of this c-book unit, the designers use a storytelling approach supported by interactive widgets, videos and pictures to motivate and engage users and guarantee unity of the c-book activities. After introducing the meaning of the term biomathematics and its fields of study in section 1, users are free to choose their path of learning. Therefore, the structure of the c-book unit allows users to autonomously navigate according to their interest and knowledge.

C. Constructivist Learning Approach

The c-book unit activities in sections 2 - 4 are developed based on the constructivist learning theory practices enabling students to create new experiences and link them to their prior cognitive structure, supported with learning opportunities for conjecturing, exploration, and mathematics communication.

D. Metacognition Activities

All c-book sections terminate with a meta-cognitive activity that has been designed to encourage students to reflect about their learning and enable them to further understand, analyze and control their own cognitive processes. These activities

have also been designed to develop students' written mathematical communication skills through the use of EpsilonChat, a widget for communicating mathematics in which users can easily write mathematical contents.

E. Technology Added-value

Three instances of designers' use of C-book technology are presented to highlight the technological added-value provided by such a system:

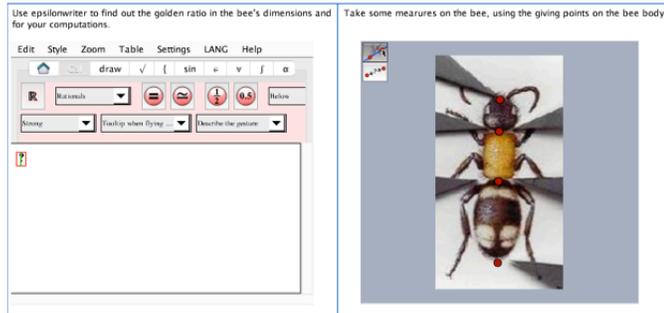


Fig. 2. Cross-widget communication to allow sending messages between Cinderella and EpsilonWriter.

The cross widget communication between Cinderella and EpsilonWriter, two widget factories⁴, enables to check whether the golden ratio models the dimensions of a bee (Fig. 2). The idea behind the cross-widget communication is to send the measurements from Cinderella widget to EpsilonWriter where users can write text and mathematical content, as well as doing calculations, in particular comparing the ratio of values. The digital golden ratio compass tool is a DGS widget enabling users to discover proportions seemingly like golden ratios in other contexts such as in insects, and in plants, artwork, etc. Users look for a picture of an object, upload it and check with the digital compass whether its dimensions satisfy the golden ratio. They can share their thoughts and findings via EpsilonChat and EpsilonWriter. The main added value here is meshing the three interconnected widgets in the same educational environment so students work on a single document which stores all their activities in their present state, a feature not presently available in other educational environments.

The C-book technology allows using external materials that couldn't be provided by the environment but are reachable by external URL links. For instance, a 3D-GeoGebra widget that enables users to change the dimensions of a cone and a cylinder in order to get the same (See: <https://tube.geogebra.org/m/2552113>).

V. THE EVALUATION OF "MATH FOR BIOLOGY" C-BOOK UNIT CMT AFFORDANCES

In this section, we present the methodology for evaluating a priori CMT affordances in pedagogical resources and present the results of the evaluation of "Math for Biology" c-book unit.

A. The Evaluation Question

We evaluate the degree to which the four cognitive components of CMT (fluency, flexibility, originality and elaboration), and social and affective aspects have been integrated and promoted through the design of the c-book unit, as perceived by the evaluators.

B. Evaluation Methodology

The evaluation of the CMT affordances is based on an evaluation tool called "CMT Affordances Grid" (See Appendix A), which has been elaborated and refined during a 3-cycle process of MC Squared project. The grid is divided into three sections. The 13 first items evaluate the c-book unit affordances towards the development of CMT in users/students. These items address the c-book unit affordances such as nature of the activities or variety of representations of mathematical concepts at stake and ask the evaluators to what extent these affordances are likely to enhance the user's cognitive processes (fluency, flexibility, originality, elaboration). The second and third sections are dealing with the social and affective aspects of the c-book unit that are likely to enhance CMT in c-book unit users. As for the first aspect, the responders were asked to evaluate the item in relation to each one of the four cognitive components of CMT in a scale from 1 (weak affordance) up to 4 (strong affordance). There was additionally an extra option called N/A in case the affordance was not applicable for the specific item. The evaluation of the CMT affordances of this c-book unit was done by three members of the French CoI, not involved in its design. It was organized in three steps. First, the evaluators had to use the c-book unit and be acquainted with the affordances. Second, a teleconference was organized by the main designer of the c-book unit to address evaluators needs for understanding and clarification. Third, the evaluators evaluated the c-book unit affordances based on the grid using an online form⁵ prepared for this purpose.

C. Evaluation Results

The chart, shown in Fig. 3, represents the evaluation of the cognitive components of CMT from the evaluators' point of view. For each of the 13 questionnaire items, the height of a color represents the proportion of the associated CMT component, while the thickness represents the mean between the four aspects for each item (the mean of means, the thicker the better). This representation is debatable since creativity is known not to foster consensus, therefore we don't expect inter-rater evaluators agreement and using the mean is blurring this effect. Indeed, the Cohen's Kappa value of 0.3 means low agreement between the evaluators: because of their different expertise, some evaluators might foresee specific creative usage where some others won't. From the evaluators' point of view, two items, Item 12: The c-book unit includes non-standard problems calling for mathematical solutions and item 13: The c-book unit includes half-baked constructs that call for intervention of the teacher were assigned the N/A value (no affordance).

⁴ Widget factory is a software system (often developed independently of the c-book environment over many years) allowing to produce sophisticated parameterizable c-book-widgets easily.

⁵ See https://docs.google.com/forms/d/1xb-wDHLVG4w2bqMFfj6h0t-6w_GCLhQIXIPD7VZAGAo/viewform

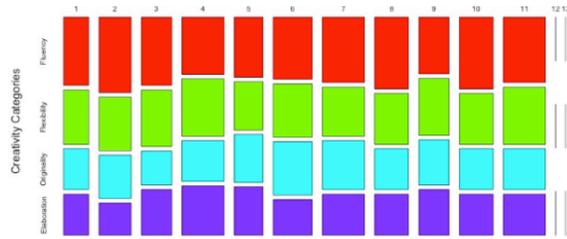


Fig. 3. Evaluation of the four cognitive components of CMT from the evaluators' point of view on the 13 items of the grid.

In Table I, we present the qualitative data for each component computing the mean from *No Affordance* (Scored 1) to *Strong Affordance* (Scored 4).

TABLE I
CMT EVALUATION SUMMARY

| Fluency | Flexibility | Originality | Elaboration | Social | Affective |
|---------|-------------|-------------|-------------|--------|-----------|
| 3.03 | 2.46 | 2.11 | 1.92 | 2.83 | 3.33 |

The highest value for this c-book in terms of cognitive aspects was fluency for which the value achieved the rank of "good affordance". It means that, in general, the c-book unit could boost the students' development of their ability to provide many responses or to come up with *many* strategies to solve a mathematical problem or challenge. Elaboration and originality are the components with lower values of weak affordance. The radar chart (Fig. 4) shows the distribution of the evaluation among the evaluated categories. This chart shows which component of CMT is most likely to be enhanced by the use of the c-book unit. In the case of "Math for Biology", it is the affective aspect, followed by fluency and social factors.

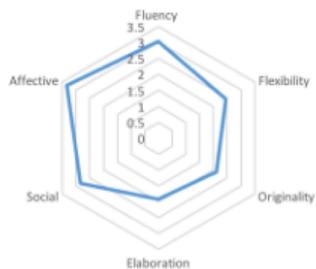


Fig. 4. Radar chart of the distribution of the evaluation among different aspects of CMT.

In line with findings reported in the literature [6-10], in Table II, collating the 13 questionnaire items, correlations among the four cognitive components of CMT are strong between some cognitive aspects.

TABLE II
CORRELATION VALUES OF CMT COMPONENTS

| | Fluency | Flexibility | Originality | Elaboration |
|-------------|---------|-------------|-------------|-------------|
| Fluency | 1.00 | 0.89 | 0.67 | 0.75 |
| Flexibility | 0.89 | 1.00 | 0.69 | 0.90 |
| Originality | 0.67 | 0.69 | 1.00 | 0.60 |
| Elaboration | 0.75 | 0.90 | 0.60 | 1.00 |

It means that, considering a significant value of $r > 0.70$ ($p = 0.05$), we may conclude that fluency, flexibility and elaboration can be fostered at the same time. In the case of originality, there is no statistical evidence that supports the hypothesis that this component can be fostered by the other

ones. Moreover, experts did not comment that much on their quantitative evaluation. Though, they raised the need for a good orchestration to achieve the goals intended by the c-book unit.

VI. CONCLUSION

This article presents a methodology for CMT supported design and evaluation of a pedagogical product. It is illustrated on the "Math for Biology" c-book unit, an interdisciplinary digital resource within the framework of the MC Squared project. The c-book unit was well evaluated in terms of fluency and affective aspects. The c-book unit calls students to make relations with proportionality and golden ratio, emphasizing the role of mathematics and how it can be perceived in nature.

We limited ourselves to an *a priori* evaluation by experts of the resource CMT affordances. Tests with students will be conducted to validate the evaluation results.

APPENDIX

Appendix A: [CMT Affordances Grid.](#)

REFERENCES

- [1] Education Council, "Recommendation of the European Parliament and the Council of 18 December 2006 on key competencies for lifelong learning," Brussels: *Official Journal of the European Union*, vol. 30, no. 12, 2006.
- [2] R. K. Sawyer, "Creative Teaching: Collaborative discussion as disciplined improvisation," *Educational Researcher*, vol. 33, no. 3, pp.12-20, Mar. 2004.
- [3] R. J. Sternberg, *Creativity and development*. New York: Oxford University Press, 2003, pp.91–138, ch. 3.
- [4] C. Hoyles and R. Noss, *What can digital technologies take from and bring to research in mathematics education?* In Second international handbook of mathematics education, pp. 323-349. Springer Netherlands, 2003.
- [5] L. Healy and C. Kynigos, "Charting the microworld territory over time: design and construction in learning, teaching and developing mathematics," *ZDM Mathematics Education*, vol. 42, no. 1, pp. 63-67, 2010.
- [6] M. El-Demerdash, "The effectiveness of an enrichment program using dynamic geometry software in developing mathematically gifted students' geometric creativity in high schools," Ph.D. dissertation, University of Education, Schwäbisch Gmünd, Baden-Württemberg, Germany, 2010.
- [7] M. El-Demerdash and U. Kortenkamp, "The effectiveness of an enrichment program using dynamic geometry software in developing mathematically gifted students' geometric creativity," in *Proc. 9th International Conference on Technology in Mathematics Teaching - ICTMT9*, Metz, France, 2009, pp.1-5.
- [8] D. Haylock, "Recognizing mathematical creativity in schoolchildren," *ZDM Mathematics Education*, vol. 27, no. 2, pp. 68-74, 1997.
- [9] E. L. Mann, "Creativity: The essence of mathematics," *Journal for the Education of the Gifted*, vol. 30, no. 2, pp.236-260.
- [10] M. Hany and M. El-Demerdash, "The effectiveness of a proposed unit in bio-mathematics for developing secondary stage students' deep understanding skills - in Arabic," *Journal of Egyptian Association for Science Education*, vol.18, no. 5, Mar. 2015.
- [11] J. P. Guilford, "Creativity," *American Psychologist*, vol. 5, pp. 444-454, 1950.
- [12] G. Fischer. (2001) Communities of interest: learning through the interaction of multiple knowledge system, Available [online]: <http://13d.cs.colorado.edu/~gerhard/papers/iris24.pdf>

Conversation-based Assessments: An Innovative Approach to Measure Scientific Reasoning

Lei Liu, Jonathan Steinberg, Farah Qureshi, Isaac Bejar, and Fred Yan

Abstract— This study explored the scalability of an innovative assessment approach, conversation-based assessment, to measure constructs related to scientific reasoning, where virtual students interact with students to demonstrate their abilities to use evidence to support a prediction. We designed two parallel tasks using the same design pattern and compared the equivalence of these tasks. We designed an experimental study to investigate whether the two parallel tasks performed equivalently by comparing student performance on both tasks. The results of the study shed light on the challenges of scalability of conversation-based assessment and the implications for next steps.

Index Terms—Conversation-based assessment, virtual agents, scientific reasoning

I. INTRODUCTION

NEW views of learning (e.g., constructivism and activity theory) call for innovative assessment modes that track students' thinking processes more closely. One underlying assumption of these new learning theories is that knowledge is socially constructed and mediated through communicating ideas with others. Conversation-based assessment is one promising approach to track and record students' higher-order thinking processes such as scientific reasoning. Building on previous research on natural language intelligent tutoring systems (Graesser, Person, & Harter, 2001), a conversation-based approach entails computer-mediated conversations. Students may take different paths through scripted conversations based on the content of their responses. The present study focused on the scalability of a conversation-based assessment approach by posing the following specific question: how can the production of conversation-based assessments be scaled up to measure higher-order thinking across different scientific topics? This problem is important given the cost-effective and validity concerns of innovative assessments, which is an area that lacks research. We describe an approach evaluating scalability whereby a successful

implementation is used as the basis for developing parallel or isomorphic tasks, an approach that has been found useful in domains as diverse as architecture (Bejar, 2002) and networking (Kunze, Mehta, & Levy, 2015). In our study, we focused on within-domain scalability, namely, building a parallel task (Weather) to our base conversation-based task in the science domain measuring the same set of constructs (Volcano), and evaluating the psychometric equivalence of the tasks. We designed an experimental study to investigate whether these two science tasks performed equivalently and therefore could be used interchangeably. Specifically, we conducted preliminary and confirmatory analyses of the data. The preliminary analysis consisted of item analysis and exploratory factor analysis; the confirmatory analysis evaluated whether performance, response time, and internal structure were comparable across the two versions of the task.

II. CONVERSATION-BASED ASSESSMENT

A. Trialogues

Conversations have been used in assessment mostly in language proficiency exams (Wong, 2000). However, it is possible that these can also be used to measure student cognitive skills in different domains. Conversations have been frequently used to support formative assessment practices in instruction for collecting evidence of students' understanding by tracking and recording students' higher-order thinking process over time. In computer-supported communication (e.g., chat box format), displaying the conversation history allows students to access their cognitive processes when communicating with virtual agents or real partners. Given all the above opportunities, conversations can be applied in assessments to support eliciting evidence of student cognition. In our study, we applied one kind of conversation-based assessment – trialogues – in our assessment design.

A trialogue is a form of adaptive testing where the student is assessed by means of one or more simulated conversations with two virtual agents. Underneath the conversation is a tree-like structure or conversation-space diagram (Zapata-Rivera, Jackson, & Katz, 2015) that contains all the possible interactions designed to elicit student knowledge on certain topics. For example, students can be assessed on their scientific thinking skills such as using evidence to support and argue for a prediction, through conversing with virtual agents in the trialogue assessment. Based on students' response to a

Manuscript received May 25, 2016.

L. Liu is with ETS, Princeton, NJ 08541 USA (609-734-5183; e-mail: lliu001@ets.org).

I. Bejar is with ETS, Princeton, NJ 08541 USA. (e-mail: ibejar@ets.org).

F. Qureshi is with ETS, Princeton, NJ 08541 USA. (e-mail: fqureshi@ets.org).

J. Steinberg is with ETS, Princeton, NJ 08541 USA. (e-mail: jsteinberg@ets.org).

F. Yan is with ETS, Princeton, NJ 08541 USA. (e-mail: fyang@ets.org).

question that asks them to make a prediction, students are branched through the conversation space to elicit their evidence base for that prediction. Students also get a chance to revise their original prediction during their reasoning process. Some characteristic features of conversation-based item types, such as providing immediate feedback through virtual agents, can lead to greater student motivation. This is a desired goal for assessments because more accurate measurement is derived from students “trying their best”, and possibly leading to more valid student assessment. However, in practice, such benefits may not be affordable unless the tasks can be developed economically, and more importantly proven as effective measures of higher-order thinking which is often hard to measure through traditional measurement. To explore the scalability of the innovative assessment approach, we developed two parallel prototypes with the goal of evaluating their equivalence in assessing student knowledge, skills, and abilities with respect to both Earth Science knowledge and related inquiry skills.

B. Parallel triologue tasks

The development of the base triologue Volcano task was informed by a design pattern and designed based on evidence-centered design principles. The purpose of the parallel prototype development was to explore whether conversation-based assessment is a scalable and valid approach to measure higher-order thinking. Design patterns are tools that can help task designers to think through substantive aspects of an assessment argument and can help capture design rationales in a re-usable and generative form (Mislevy & Haertel, 2006). The design pattern aimed to support the construction of computer-delivered conversation-based tasks that engage students in conversational interactions with virtual agents and data collection tools. The design pattern focused on the skills of planning and carrying out data collection in the virtual field and using collected data as evidence to predict a natural event. The Volcano task asked students to design and carry out a data collection plan for the purpose of making a prediction of the likelihood of a volcanic eruption. When developing the parallel Weather task, we used the same design pattern to guide our prototype development about collecting data for making a prediction of the likelihood of a thunderstorm.

Our design pattern focuses on a subset of components of scientific reasoning, in particular using observation data to make a prediction for natural events, and includes four focal areas of knowledge, skills, and abilities (KSAs) relevant to scientific reasoning: earth science knowledge, analyzing data and identifying patterns, conducting data collection, and making predictions based on data. All items in both prototypes were mapped onto these four constructs that are aligned with the new science standards.

III. STUDY DESIGN

To evaluate the equivalence of the Weather and Volcano prototypes in measuring the same constructs, we administered both Volcano and Weather tasks to 210 students in the fall of 2014. Students were randomly assigned to one of two versions

of alternating order of the same assessment. Students with an odd-numbered student ID were in one condition and students with an even numbered student ID were in the other condition. Below are three specific research questions to explore the equivalence of the two tasks:

1. Do the two prototypes meet basic psychometric criteria?
2. How do the relationships among constructs compare across the two prototypes?
3. Do randomly equivalent samples of students perform comparably on both triologue tasks with respect to the focal constructs and the time demands of the task?

A 2x1 between-condition design was applied in the study. Students were randomly assigned to two conditions. In condition 1, students took the Volcano task first and then took the Weather task. In condition 2, the order of the tasks were counterbalanced. All participants took the background information questionnaire (BIQ) at the beginning of the study and a post-survey at the end of the study. Both instruments were administered online. The BIQ survey consisted of demographics, self-reported course grades, grade level, technology use, prior game use, prior knowledge of volcanoes and weather, and non-cognitive attributes (e.g., persistence, willingness to learn, domain relevant self-efficacies, interest in domain, etc.). The post-survey included a combination of constructed response items and multiple choice items that measured both conceptual understanding and science inquiry skills related to the constructs of the Weather and Volcano prototypes. In addition, the posttest also included questions to collect evidence of participants’ perceived motivation and engagement when interacting with triologue systems.

IV. DATA ANALYSIS

A. Basic psychometric criteria: Item and Exploratory Factor Analyses

We first conducted basic analyses of the psychometric functioning of the two versions of the task. Specifically, we conducted item analyses (Penfield, 2013) and exploratory factor analyses. Item analysis evaluates the difficulty of each item as well as its discrimination power (i.e., is the item able to sort students into those that know the answer to the item and those that do not). The results of the item analysis are relevant to the scoring assumption in a validity argument (Kane, 2006) and provides an aspect of the foundation for a validity argument regarding the scores from the task. In addition, given that the items constructed for each prototype were intended to be equivalent, and administered to randomly equivalent samples, we explored whether the item difficulties and discrimination were comparable. We followed the item analysis with an exploratory factor analysis of the inter-relationships among items to determine the structural similarity of the Volcano and Weather tasks.

B. Relationship among constructs: Confirmatory analyses

We evaluated comparability in the level of student performance exhibited on the Volcano and Weather tasks. As the data for the initial analyses consisted of observed scored responses on multiple-choice items as either right (1), wrong

(0), or in some cases partial credit (0.5 or 0-3 in the case of constructed response items), exploratory factor analyses (EFA) were conducted separately on each form in SAS 9.3 using a polychoric correlation matrix first generated from PRELIS 2.57. Within-prototype confirmatory factor analyses (CFA) were conducted in LISREL 8.80 testing for the fit of the data to a single latent dimension.

C. Performance analysis: Task scores and response times

For purposes of comparing the equivalence of Volcano and Weather prototypes, we conducted descriptive analyses to compare the mean construct scores as well as the total scores across the two prototypes. Construct sub-scores were computed by simply summing the individual item scores and then a normalized mean was computed (sum/items), also regarded as percent correct. We also compared the total time spent on each of the two prototypes.

V. RESULTS

A. Preliminary Item and Factor Analyses

The item analysis examined the difficulty and discrimination of the scored items in each task. The ranges of difficulty were 0.06-0.98 for Volcano and 0.05-0.81 for Weather. The ranges of discrimination were 0.34-0.84 for Volcano which was good and 0.08-0.88 for Weather which indicated some underperforming items. There was also a difference in presentation for one item between tasks: Weather is a single selection versus the counterpart in Volcano which is a multiple selection item, making the item inherently more difficult in Volcano, possibly due to a higher tendency to misread or misunderstand the question (Cassells & Johnstone, 1984; Haladyna & Downing, 2002).

Results from the exploratory factor analysis for each prototype indicated the presence of a relatively strong first factor but more emphasis was placed on the confirmatory factor analysis to allow items to load on four particular constructs: Earth Science Knowledge, Analyzing Data & Identifying Patterns, Conducted Data Collection, and Making Predictions Based on Data. A fifth construct, Evidence-Based Reasoning while, part of the assessment, was not included for this analysis, but is discussed with respect to the performance analysis.

B. Confirmatory Factor Analyses

The fit statistics from the within-prototype confirmatory factor analyses (CFA) shown in Table 1 were all better for Weather than Volcano.

Table 1. CFA Fit Statistics for Volcano and Weather.

| | Volcano | Weather |
|----------------|----------------------|----------------------|
| RMSEA (90% CI) | 0.085 (0.070, 0.101) | 0.074 (0.058, 0.090) |
| CFI | 0.880 | 0.949 |
| SRMR | 0.145 | 0.133 |

Note: RMSEA = Root Mean Squared Error of Approximation; CFI = Comparative Fit Index; SRMR = Standardized Root Mean Residual.

The latent inter-correlations shown below in Table 2 indicate low to moderate relationships for each prototype, suggesting

four constructs may exist, even though the magnitudes may differ

Table 2. Inter-correlations among postulated skills for Volcano and Weather.

| Factor | Volcano | | | | Weather | | | |
|--------|---------|-------|-------|-------|---------|-------|-------|-------|
| | ESK | ADP | CDC | MP | ESK | ADP | CDC | MP |
| ESK | 1.000 | -- | -- | -- | 1.000 | -- | -- | -- |
| ADP | 0.432 | 1.000 | -- | -- | 0.802 | 1.000 | -- | -- |
| CDC | 0.223 | 0.422 | 1.000 | -- | 0.545 | 0.580 | 1.000 | -- |
| MP | 0.319 | 0.115 | 0.232 | 1.000 | 0.397 | 0.249 | 0.384 | 1.000 |

Note: ESK = Earth Science Knowledge; ADP = Analyzing Data & Identifying Patterns; CDC = Conducting Data Collection; MP = Making Predictions Based on Data.

The efficacy of the task model with multiple dimensions appears to be relatively good. For Volcano, all but one item loads above 0.40 on these latent dimensions, while for Weather all items load above 0.33. Additionally, there is evidence that the items generally load saliently on individual prototype factors, but the model fit is better with four factors for each prototype. Therefore, while one can examine the prototype as a whole, its underpinnings as represented by these four constructs are worthy of examination too.

C. Performance Analysis

As shown in Figures 1 and 2 using a percent correct metric, in general and regardless of task order, students performed better in Weather (54-56%) than on Volcano (47-49%) and particularly in Earth Science Knowledge (Weather = 41-50%; Volcano = 27-30%), possibly because that students had more opportunities to learn about weather knowledge in their daily lives. Aside from that, students' performance on items related to some constructs (e.g., Conducted Data Collection and Making Predictions Based on Data) was similar across prototypes, but performance on other items was not comparable related to Analyzing Data & Identifying Patterns (Weather = 79-84%; Volcano = 68-73%) and Evidence-Based Reasoning (Weather = 41-62%; Volcano = 12-26%), again partly because those constructs have more demands on applying Earth Science Knowledge.

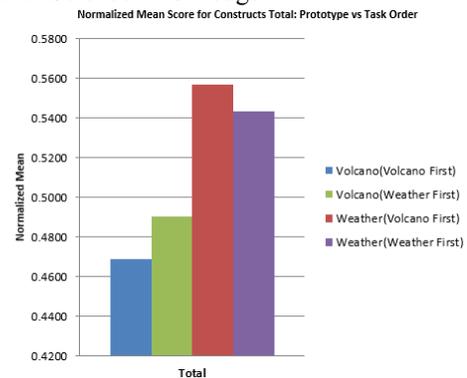


Fig. 1. Normalized mean scores for total scores across prototypes in two conditions.

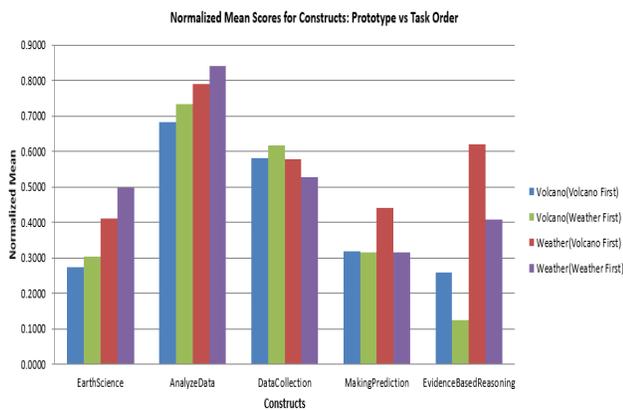


Fig. 2. Normalized mean scores for constructs across prototypes in two conditions.

In general, students spent more time (minutes) on Volcano (Mean = 30.2; SD = 13.4) than on Weather (Mean = 27.5; SD = 11.8), which can also be explained by a possible difference in familiarity with the contexts. However, the correlation in time spent between prototypes was close to zero. This necessitated the analyses to be broken out by task order. Time (minutes) spent on the second prototype was shorter (Mean = 23.6; SD = 7.1) than for the first prototype (Mean = 34.1; SD = 14.8) with a correlation of 0.32. The correlation in task times when Volcano was administered first was 0.35 and when Weather was administered first, it was 0.28.

VI. CONCLUSION

Even with all the limitations of the study (e.g., a limited data set), as shown in both item analyses and factor analyses, there are some encouraging results to show evidence of the comparability across two science conversation-based prototypes that were designed to measure similar constructs in two different science contexts. At the item level, both the item discrimination and difficulty were very much similar across the two prototypes. Performance on the aggregation of items into construct sum-scores also suggested a high degree of comparability. The analysis of the item-level performance through factor analysis suggested a high degree of cohesion among the items reflected in the strong presence of a primary dominant factor. In addition, there was evidence that students' performance on items related to some constructs (e.g., data collection and making predictions) was comparable, but performance on items related to other constructs (e.g., earth science knowledge, analyzing data, evidence-based reasoning) was not comparable. Our interpretation is that those incomparable items share highly similar surface level features (e.g., multiple choices) which may have enhanced student performance despite isomorphic item design. However, the comparable items mostly include conversation-based and performance-based items, indicating the potential advantage of

using such items to measure student understanding in different contexts (Zapata-Rivera et al., 2015).

Furthermore, the distribution of response times was reasonably comparable across the two prototypes. In both conditions, students spent less time on the second prototype, suggesting some increased familiarity with the interface. In general, students performed better on the Weather prototype than the Volcano prototype, possibly because students were more familiar with the content through their daily life experiences. Such departures from strict isomorphism are to be expected and do not necessarily preclude the possibility of using the two as if there were exchangeable, especially in a low-stakes assessment environment. The preliminary results showed promise with respect to scalability in both developing parallel prototypes in the domain of science. The possible use of similar conversation-based structures to develop assessments in different contexts indicate possible cost savings for innovative assessment development. We also realize the limitation of our research given the small sample size. For future studies with a larger sample size, the underlying issues encountered in the factor analyses with the use of polychoric correlations and other analyses could be resolved through MIRT procedures.

REFERENCES

- [1] Bejar, I.I. (2002). Generative testing: From conception to implementation. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 199-218). Mahwah, NJ: Earlbaum.
- [2] Cassels, J. T., & Johnstone, A. H. (1984). The effect of language on student performance on multiple-choice tests in chemistry. *Journal of Chemical Education*, 61, 613-615.
- [3] Graesser, A. C., Person, N. K., Harter, D. (2001) The Tutoring Research Group: Teaching tactics and dialogue in AutoTutor. *International Journal of Artificial Intelligence in Education* 12, 257-279.
- [4] Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15, 309-334.
- [5] Kane, M.T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: Praeger Publishers.
- [6] Kunze, K. L., Mehta, V., & Levy, R. (2015). *Leveraging psychometric isomorphism in assessment development*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL, April 2015.
- [7] Mislevy, R. J., & Haertel, G. D. (2006). Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice*, 25, 6-20.
- [8] Penfield, R.D. (2013). Item analysis. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology*, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology (pp. 121-138). Washington, DC: American Psychological Association.
- [9] Wong, J. (2000). Delayed next turn repair initiation in native/non-native speaker English conversation. *Applied Linguistics*, 21(2), 244-267.
- [10] Zapata-Rivera, D., Jackson, T., & Katz, I.R. (2015). Authoring conversation-based assessment scenarios. In R. A. Sottolare, A. C. Graesser & J. Hu (Eds.), *Design Recommendations for Intelligent Tutoring Systems* (pp. 169-178). Army Research Institute.

Review of *Minds Online: Teaching Effectively with Technology*

Angela S. Wagner

Abstract—This article reviews Michelle Miller’s book, *Minds Online: Teaching Effectively with Technology* with a brief introduction, review of the its contents, and practical suggestions for professionals to utilize this work. This 2014 release includes 9 chapters, 296 pages, sample syllabus, notes, and suggestions. It is published by Harvard University Press: Cambridge, MA. <http://www.hup.harvard.edu>, ISBN: 9780674660021.

Index Terms— blended learning, online learning, psychology, teaching, technology.

I. INTRODUCTION

THE dynamic changes to education in the digital age propel instructors into new teaching environments: blended and online classrooms. Despite the push to critically examine new technologies and their potential to impact learning, Miller’s book uniquely focuses on the cognitive aspects of learning. She elucidates how technologies can complement psychology through course design and strategy to produce success for students in online and blended classrooms. In *Minds Online*, Miller posits supportive ways to make online learning successful through thoughtful course design and formatting. As a Redesign Scholar for the National Center for Academic Transformation and co-creator of the First Year Learning Initiative at Northern Arizona University, Miller informs and encourages teachers in incorporating learning techniques supported by cognitive research when structuring courses utilizing technology to capitalize on learning opportunities given that technology allows us to “amplify and expand the repertoire of techniques that effective teachers use to elicit the attention, effort, and engagement that are the basis for learning” [xii].

II. BOOK CONTENTS

A. *Online Learning’s Durability, Effectiveness, & Psychology*

Chapters 1 through 3 address three majors concerns in regards to online learning: Is it here to stay? Does it work? How do computing aspects change our learning?

Chapter one explores the four factors promoting the growth of online learning: economics, student demand, measureable evidence, and new technologies. Given the growth of technology in education for these reasons, it is prudent to learn to harness its power.

Miller undertakes the major concern of performance in online learning in chapter two. Here, the quality of identical best practices in traditional classrooms versus online courses are compared, and research is presented to support similar outcomes despite modalities. She refocuses our attention to maximizing the quality of instruction [41].

The third chapter examines the ways computing changes us and dispels three myths related to negative changes of technology use by presenting the facts surrounding each idea; Miller then critically analyzes how each change truly affects teaching.

B. *Attention, Memory, & Thinking*

Chapters 4 through 6 delve more deeply into the cognitive and brain science behind quality, effective learning. Miller defines each area of focus, explains its processes, and supports its implications for online learning with helpful strategies to make learning advantageous, especially by understanding how the frameworks of cognitive psychology impact learning.

The focus of chapter four presents the mental processes of attention, a necessary ingredient to learn material. It investigates the impact of working memory, automaticity, and voluntary control at work in classroom learning, and furthermore, observation is given to situations when attention works in a differing manner (ADHD). Miller provides four strategies to harness the processes of attention to translate learning into a digital environment.

Chapter five tackles the center of teaching and learning: memory, and it thoroughly delineates memory theories, old and new, while also exploring how and when memory is affected during learning. The testing, spacing, and interleaving effects are illustrated and applied to online courses along with useful techniques to apply to course design, making activities and content more memorable.

Miller focuses on thinking in chapter 6, primarily noting the basic differences in novice and expert thinking, which is the use of effective thinking strategies. By reviewing the structural elements of problems and utilizing analogical reasoning, she proposes how novices become expert thinkers and then addresses ways to teach metacognition and critical thinking via online course activities.

Manuscript received June 20, 2016.

Angela S. Wagner is an instructor at Athens High School, Athens, AL, and a doctoral student in Boise State’s Educational Technology program, Boise, ID 83725 USA (phone: 256-233-6613; fax: 256-233-6617; e-mail: angelawagner@u.boisestate.edu).

C. Incorporation, Motivation, & Design

The final three chapters of Miller's book combine the research behind cognition and learning with application to online learning.

Chapter seven targets effective incorporation of multimedia into coursework, expounding upon ways to gain "educational mileage" [149] from materials we create. Sensory modes, multimedia theory, animations, and simulations comprise the main support behind the helpful accommodations suggested to increase learning for all students.

Given the emphasis on student motivation in contemporary teaching complaints, in chapter 8 Miller enumerates the varying components (and hindrances) to student motivation, a major component in both traditional and digital classrooms. Common roadblocks to student performance are often rooted in motivational factors, which are met in this chapter along with practical suggestions to allay these issues. Again, research meets practice as theories and trends are applied to the online format. Self-determination theory, intrinsic/extrinsic motivation, and growth mindsets are explored along with self-efficacy, anxiety, and procrastination.

Closing this work, chapter nine culminates all of the theories, research, and classroom experiences Miller presented in previous chapters, recapping it all in a clean, bulleted Q&A format. Areas crucial to all courses such as objectives, activities, assessments, peer interaction, and grades connect to guiding principles and instructor questions; additionally, a short list of mentioned resources from each driving cognitive principle is compiled for quick reference; this simplifies design elements for teachers looking to create success in online courses.

III. SUMMARY & CRITICAL REMARKS

Though *Minds Online* targets online courses in higher education, it addresses topics K-12 educators and instructional professionals would value as well. With the steady growth of virtual courses, Miller's psychology-based approach to leverage technology's benefits offers a foundation for the construction of knowledge for beginners new to course design. Concomitantly, it also while gives a convincing take on the tools and resources available to enhance current coursework for seasoned professionals. The author's suggestions for instructional design even denote points one may consider when evaluating current coursework and activities. Designers and instructors need to continually manipulate the online environment in unique, memorable ways to capture student attention and foster creative thinking

The book's conversational tone and thoroughly explicated aspects of psychology and pedagogy make it accessible to a wide audience seeking to understand the basics of teaching and learning virtually. Miller alternates between cognitive theories and their applications in a classroom, but she successfully explains such concepts without relying heavily upon technical jargon. She follows each concept with a clear classroom connection, an actionable step aligned with the way the mind works. This readability is helpful and appealing to educators who might lack the time to sift through the formidable stacks of educational psychology research

currently available. The centered exploration of the underlying cognitive influences on student learning makes it relevant to teachers of all levels and content areas incorporating technology into the classroom, traditionally or otherwise, because we cognitively learn various content in a similar fashion.

The perceived competition between traditional and distance courses has created a myopic view of online instructional design with worry surrounding emerging technologies' capability to shorten attention spans or produce lackluster results. Despite cautionary examples of the drawbacks of technology found in works such as Nicholas Carr's book *The Shallows*, which attributes emergent technologies with *rewiring* our brains, Miller warns we must avoid misinterpretation of such works and confidently redirects this sentiment by denoting that "technically speaking, computing experience does alter our brains at a neural level, but so does just about anything else that we remember" [45]. In this work, she explains how minds learn differently online and reminds readers that technology is a powerful tool when integrated effectively. Miller clearly juxtaposes misconceptions of online learning with best practices to elucidate ways one can utilize and manage resources to enhance learning. The thorough research behind each element and suggested applications leave readers with sound, duplicable strategies, ideas, and online tools to try in the classroom to improve student learning for all.

Given the variety of learners, new technologies and tools, and instructors, it is imperative not to view this book as a quick fix for any one online classroom issue but rather an opportunity to seek out the science behind learning to examine ways to strengthen online and blended instruction. Readers finish with a clear understanding of how cognitive research enhances the decision-making processes and critical thinking put into course design and learning opportunities. All in all, this book's practicality and solid theoretical supports make it a valuable addition to any education professional's reading repertoire, one of which Miller hopes "inspires you to use future technologies-tools we can't even envision yet-to serve, inform, and inspire students, all with an eye to the cognitive process that drive learning" [196].

REFERENCES

- [1] N. G. Carr, *The Shallows*. New York, NY: W.W. Norton, 2010.
- [2] M. Miller, *Minds Online: Teaching Effectively with Technology*. Cambridge, MA: Harvard University Press, 2014, pp. xii, 41, 45, 149, 196.

Angela S. Wagner became a member of IEEE in 2017. She holds a Master's degree in curriculum and instruction from the Pennsylvania State University, University Park, PA, 2011 and a Bachelor's of Education in secondary education from Athens State University, Athens, AL, 2004.

Angela currently teaches secondary English and Advanced Placement Literature and Composition at Athens High School. She is currently study educational technology in the Boise State Ed.D. program. Angela's focused interests include online and blended learning, gaming and apps in education, and emerging technologies.

Angela Wagner is a member of Association for Educational Communications & Technologies, the International Association for K12 Online Learning, and the Online Learning Consortium. She is also a member of Delta Kappa Gamma.