Multimedia Learning Principles at Scale Predict Quiz Performance

Anita B. Delahay
Carnegie Mellon University
Pittsburgh, PA 15213, USA
adelahay@andrew.cmu.edu

Marsha C. Lovett
Carnegie Mellon University
Pittsburgh, PA 15213, USA
lovett@andrew.cmu.edu

ABSTRACT
UPDATED—14 June 2018. Empirically supported multimedia learning (MML) principles [1] suggest effective ways to design instruction, generally for elements on the order of a graphic or an activity. We examined whether the positive impact of MML could be detected in larger instructional units from a MOOC. We coded instructional design (ID) features corresponding to MML principles, mapped quiz items to these features and their use by MOOC participants, and attempted to predict quiz performance. We found that instructional features related to MML, namely practice problems with high-quality examples and text that is concisely written, were positively predictive. We argue it is possible to predict quiz item performance from features of the instructional materials and suggest ways to extend this method to additional aspects of the ID.

Author Keywords
Multimedia Learning; Cognitive Load; Coherence Principle; Instructional Design; Online Learning.

ACM Classification Keywords
H.5.1 Multimedia information systems: Evaluation/ methodology;
K.3.1 Computer uses in education: Computer-assisted instruction (CAI); Distance learning.

INTRODUCTION

Multimedia learning (MML), defined as learning from words and media (e.g., pictures, graphics, simulations, lecture capture, videos), has been extensively studied, resulting in empirically demonstrated principles that can enhance learning [1]. According to these MML principles, adding visuals to text generally improves learning. However, visuals that cause learners to split their attention or concurrently process competing types of information (e.g., auditory and visual) often detract from learning. In addition, several MML principles are based on the idea that cognitive load that is extraneous (i.e., engages mental resources with no instructional benefit), leads to slower or worse learning [2].

L@S 2018, June 26–28, 2018, London, United Kingdom
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5886-6/18/06 $15.00
https://doi.org/10.1145/3231644.3231694

The principles make clear recommendations for ID, and yet are based on research conducted in the psychologist's lab (rather than real classes) comparing "single variable" contrasts (rather than realistic instruction) [1]. Thus, it is less clear how the principles combine into a coherent ID and how well the principles' predictions are met in real educational settings.

In this study, we analyzed MML principles in the context of use data from a fully online course with a large learner population. We hypothesized that ID features that adhered to MML principles would predict better performance on quiz items but only as a function of use – because the presence of MML-consistent instruction can only promote learning if students engage with it (e.g., read materials or do practice).

MML AT SCALE

To operationalize MML in a natural instructional design, we analyzed features of a fully online homework system deployed in the "Introduction to Psychology as a Science" MOOC offered by the Georgia Tech through Coursera in 2013. The Coursera platform housed video lectures, links to multimedia and interactive practice activities developed at the Open Learning Initiative (OLI) at Carnegie Mellon University, and weekly quizzes. Log files contain behavioral (e.g., time-on-task) and performance (on practice and quiz items) data. Five content modules and two related quizzes were reviewed. These modules were selected to include both technical topics (e.g., research design) and subject-oriented topics (e.g., trait theories) from both the beginning and end of the course.

To analyze MML, we identified content and practice activities that covered the knowledge and/or skills needed to answer each quiz item based on the course designers' skill map. This "backwards design" process resulted in a mapping of quiz items to specific instructional materials and practice activities, and allowed us to quantify the MML features associated with each quiz item as well as student use of those features.

MML Principles in an ID

We operationalized two MML Principles in MOOC lessons. Measurements quantify aspects of the instructional materials (e.g., text, graphics, practice) that support a given quiz item.

Multimedia Principle People learn more deeply from words and pictures than from words alone [1]

Measured By:
(1) *multimedia quantity*: the number of graphical elements (e.g., graphs, tables);
(2) *multimedia complexity*: a ratio that captures whether graphics show one or more instructional concept(s)

Coherence Principle People learn better when extraneous material is excluded rather than included [1]

Measured By:
(1) *text verbosity*: the number of paragraphs of text
(2) *high-quality examples in practice problems* (see *Figure 2* for a further description of this measure)
(3) *high-quality examples in text*

Figure 1. Instructional features related to MML Principles that were applied to instructional materials (e.g., text, graphics, and practice activities). Measures quantify features in the instructional materials that support a given quiz item.

METHOD

Sample

We included the subset of the MOOC learners who registered for the OLI content (9,075 of 27,720), then limited the OLI-registered sample to 783 people who completed all eleven weekly quizzes plus the post-test. See [3] for a description of the entire sample.

Dependent Measure

Performance on sixteen quiz items taken from the two quizzes (four items related only to the Coursera videos were excluded).

Independent Measures

We operationalized instructional features related to MML in text, graphics, and practice activities by coding five variables: (1) *multimedia quantity*: the number of graphical elements (e.g., graphs, figures, tables); (2) *multimedia complexity*: a ratio of number of graphical elements to instructional concepts depicted in the elements. A 1:1 ratio indicates *low complexity multimedia*, whereas less than one (multiple concepts per graphic) means higher conceptual complexity; (3) *text verbosity:* the number of paragraphs of text. A higher number might indicate a higher percentage of tangential or extraneous details and less succinct writing. Although not a perfect measure, we reasoned that topics in an introductory, survey course in psychology would generally have similar complexity and should require similar depth of treatment; (4) the number of *high-quality examples* in practice problems and (5) in paragraphs of text. We measured the percentage of each that contained high-quality, versus no or low-quality, examples. See *Figure 1* for more details.

Two measures of use were computed: Read Time (sum of minutes spent on pages relevant to the quiz item) and Do Time (sum of minutes spent on relevant practice activities). Usually entire text pages and practice problems mapped to a single quiz question. In rare cases where a page or practice problem mapped to two quiz items, the minutes were allotted to both.

Initial Analyses and Transformations

An initial analysis of use data revealed large outliers. High outliers likely resulted from leaving a browser page unattended. Do Times were capped at 15 minutes and Read Times were capped at 20 minutes. Measures of zero seconds for Do (451 rows) and fewer than 15 seconds for Read were deleted (8,117 rows). Also, any learner who spent less than 15 minutes total on all available read and do activities or accessed fewer than five pages was removed. This resulted in 689 learners. Finally, based on the distributional properties of the variables, we coded instructional features via a median split and transformed use data via cube root function.

RESULTS

We regressed quiz item accuracy on MML features, their use, and the interaction of features and use. Percent of practice with high-quality examples was crossed with do time. The other four features were crossed with read time. Student was a random effect in the model.

The regression analysis showed that quiz item accuracy was predicted by spending more time practicing problems with high-quality examples compared to low-quality or no examples ($\beta = 0.28$, $p < 0.001$) and by spending more time reading when there were fewer paragraphs compared to more ($\beta = -0.38$, $p < 0.001$). For each additional minute of practice where the percent of practice problems containing high-quality examples was greater than the median split, the learner was 2.28 times more likely to get a quiz item correct. See *Figure 2* for examples of the use of high- and low-quality examples in practice problems. Furthermore, for each additional minute of reading where the number of paragraphs was greater than the median split, the learner was 0.33 times less likely to get a quiz item correct.

What is a High-Quality Example in a Practice Problem?

Practice activities with higher-quality examples positively predicted quiz performance. High-quality examples illustrate, extend, or contextualize an instructional concept; low-quality examples do not.

In many low-quality examples, the scenario or scene could be removed and the question would still make sense.

High-Quality Examples

(1) Ninth-grade students and teachers were surveyed to determine the level of bullying that the students experienced. The researchers were given permission to access scores of the students on several standardized tests, with topics including algebra, earth science, and world history. The researchers found that the more bullying a student experienced, the lower the student's grades on the standardized tests.

Was there a positive correlation, a negative correlation, or no correlation?

(2) Doug considers John to be a very honest person. In fact, in a psychology class in which the consistency of traits was being discussed, Doug cited his friend as an example of consistency.

It may be that Doug is overestimating John's consistency because Doug ________________.

(3) According to psychodynamic theory, when the ego cannot find a realistic solution to the conflict between the id and the superego, it relies on defense mechanisms as a way to reduce anxiety by unconsciously distorting reality. Complete the table below by dragging each defense mechanism to the example that corresponds.

Low-Quality Examples

(1) Dr. Kellman finds a near-zero correlation between identical twins' scores on a measure of a personality trait. Dr. Kellman will probably attribute variability in the trait to __________________.

(2) Identify the theory that forms the basis for each of the following individuals' approach to assessing personality.

Dr. Grey assumes that personality can be measured by assessing the patterns of bumps on people's skulls.

Figure 2. Examples of high- and low-quality examples. Leaners' use of practice problems with higher-quality examples positively predicted quiz performance.

DISCUSSION

In this study, we analyzed MML in a large-scale, online course based on a natural ID and student use data. The two significant predictors conformed with theoretical predictions. Namely, MML features in practice and text that might require extraneous processing were associated with worse quiz performance.

We extended MML principles, typically applied to multimedia, to learning from text and practice. Recent research [3] has reported that learners above the median in initiating practice activities (i.e., the "doers") scored best on quizzes, regardless of how much they read text or watched videos. We wondered whether MML hypotheses related to cognitive load might explain the greater effect of doing compared to reading. As reading constitutes mental activity [4], we found load could be problematic in both reading and doing.

This work also contributes to MML theory by suggesting an alternative measure of cognitive load, which is typically measured by self-report after task completion [5]. Live courses with competing elements have natural variability in MML features related to cognitive load. One criticism is that MML studies do not actually tax cognitive load [5]. Looking at MML features in aggregate may give better measures of additive load, and interactions with use of multiplicative load, that more authentically tax working memory resources. See *Figure 3* for additional MML principles that may be found in a large-scale instructional design.

This study had several limitations. First, it was correlational in nature. Experimental manipulation of these variables would be valuable and could use a "lesioning" approach — manipulating the presence or absence of particular multimedia features or activities — to evaluate the impact of MML components on learning. Relatedly, there is certainly unique "noise" in the measurements from this specific population and environment. Therefore, we do not claim that this study reveals universal principles of MML at scale. Analyses of additional datasets are needed to determine whether these effects replicate and generalize to further contexts, and to refine the operationalization of MML principles at scale.

Despite this, it seems useful to investigate the most coherent presentation of multimedia elements and text that are combined within a lesson and used over time. Instructors and instructional designers would benefit from a better understanding of how to maintain learners' attention and help them process the large quantity of content they often encounter.

Additional MML Principles in a Full-Scale ID

We found evidence of theoretically interesting measures related to MML principles (see [1]) that did not vary enough in our learning materials to include in our analyses. Depending on the design of a course or instructional platform, these measures may be possible:

Coherence: all words and pictures relevant to an instructional goal (Note: our coherence measures focused on the words in text and practice)

Signaling: cues highlighting the organization of key information in text and graphics via formatting

Segmenting: learner-controlled pacing through segmented instruction

Personalization: conversational style in multimedia and text

Feedback: explanative feedback on performance

Figure 3. Additional MML Principles that may be measurable in a full-scale instructional design.

ACKNOWLEDGEMENTS

The research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305B150008 to Carnegie Mellon 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

1. Richard E. Mayer (Ed.). 2014. The Cambridge Handbook of Multimedia Learning (2nd ed.). Cambridge University Press, New York. https://doi.org/10.1017/CBO9781139547369
2. Ruth C. Clark, Frank Nguyen, John Sweller. 2006. Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load. Pfeiffer, San Francisco.
3. Kenneth R. Koedinger, Jihee Kim, Julianna Zhuxin Jia, Elizabeth A. McLaughlin, Norman L. Bier. 2015. Learning is Not a Spectator Sport: Doing is Better than Watching for Learning from a MOOC. In *Proceedings of the 2nd ACM Conference on Learning @ Scale*, 111-120. DOI:10.1145/2724660.2724681
4. Ruth C. Clark, Richard E. Mayer. 2008. Learning by viewing versus learning by doing: Evidence-based guidelines for principled learning environments. *Performance Improvement.* 47, 9: 5-13. DOI: 10.1002/pfi.20028
5. Ton De Jong. 2010. Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science.* 38, 2: 105-134. DOI: 10.1007/s11251-009-9110-0