Tailoring learning experiences for adults through data analytics in e-Learning

The type of work modern humans conduct is rapidly changing. The way we teach humans to interact with technology (and the data it leaves in its wake) needs to be updated. Technological development necessitates changing both the content of adult learning opportunities, and the methods used to design, distribute, and assess these educational experiences. Data science is at the intersection of using computational tools in the workplace and gleaning insights from the data being rapidly acquired by developing technologies. The democratization of data science education has blossomed at scale in the past few years (e.g Kross, Peng, Caffo, Gooding, & Leek, 2017), piggybacking on the emergence of educational data science in the mid 2000s (Piety, Hickey, & Bishop, 2014) and the success of Massive Open Online Courses (MOOCs) in a variety of subjects. However, the use of data analytics to tailor the experience to online adult learners, with their diverse motives for engaging with data science learning materials, is lacking. Developments in adult learning, a shift in focus from completion rates to learning outcomes in e-Learning, a better understanding of learners' motivators, and refinements to typical data science curricula all indicate that the ability to leverage one's data will soon be accessible to more fingertips than ever.

What motivates MOOC learners? Even a few years ago, edX had 2 million users, spanning 8 to 95 years of age, from all over the world (Agarwal, 2014). What most users had in common was a desire to learn, to connect globally, and to consume online content. Most had little interest in earning a certificate. Learners can be driven by at least four types of motivator: goals pertaining to learning, performance, identity, and social affinity. Common means of motivating learners – gamification, badging, and community – can differentially target all of these (Magnifico, Olmanson, & Cope, 2013).

Of course, differences in specific motivators exist based on course topic. In an analysis of the same Introduction to Data Science course described by Kross et al., researchers found that most learners were driven by their interest in the course topic/content, even when they are already professionals in the field. Many were also motivated to prepare for future careers or for role progression (Milligan & Littlejohn, 2017).

Cluster analysis of MOOC learners on the basis of engagement with material and community, demographic characteristics, and learning outcomes is revealing. Most (65%) of users are tasters, quickly abandoning an online course (Kahan, Soffer, & Nachmias, 2017). Downloaders (9%) quickly grab course content but do not engage, whereas disengagers (12%) participate fully for a while before dropping out. The remaining students are engaged, watching most videos and taking the final exam, but those who do in-video quizzes and engage heavily on discussion forums end up with the highest grades.

Following up on the success of horizontally-focused e-Learning companies like Coursera and edX, DataCamp created a company with a depth-based, vertical focus for data science. This structure allows individuals with skill gaps at different levels to allocate their time accordingly (Cornilessen, 2018). Across both broadly-focused MOOC platforms and more tailored online learning environments, what outcomes are being measured?

Four conventional outcome variables were evaluated in a 2012 MOOC: enrollment, participation, curriculum, and achievement (DeBoer, Ho, Stump, & Breslow, 2014). Upon analysis, the authors suggest that theoretical adjustments are necessary in this educational context. Enrollment blends the experience of the committed active learner with the casual content user. Participation can be broken down by attendance, clicks, hours, and assessment across multiple timescales and material types. Pairwise rank correlations among these metrics are low. Participation can take on many forms, and neither posting on forums nor completing videos is necessary to remain engaged in and successfully complete a course (Balakrishnan & Coetzee, n.d.). Curriculum traversal is highly asynchronous, with users pursuing their own pathways. Releasing course content sequentially is not beneficial enough to warrant limiting learners' flexibility (Mullaney & Reich, 2014). Some students actively participate without attempting any assessment. How should their learning be measured?

In a study measuring engagement, mind wandering, interest, and learning outcomes across numerous course structures, those with assessment yielded high drop out rates, but assessment's negative effects could be mitigated with feedback and explanation (Thomas, Türkay, & Parker, 2017). Not all content creation is beneficial, though: for every 100 characters of instruction in DataCamp courses, completion rates drop 3.7% (Vaidyanathan, 2018). Subjective ratings suggest that the feeling of confusion is a stronger indicator of comprehension than feelings of difficulty or understanding (Thomas et al., 2017). Difficulty is known not to relate to assignment grades (Thille et al., 2014).

Newly-defined fields of study always go through a turbulent pedagogical period initially, while the boundaries of the field are being established and adopting a beginner's perspective is more difficult. Data science benefits from combining elements from fields which have focused heavily on education already, such as computer programming, statistics, and communication. One helpful strategy to adopt from programming is Parsons problems (Wilson, 2018). These problems require learners to rearrange lines of code so that they execute in an expected fashion. This shifts the focus from learning syntax or vocabulary to thinking about control flow. A similar approach, applied at various levels of granularity, could be applied to data science analyses.

Data science education subsumes so many disciplines that a given student could excel in some but not others, making personalization important. While typical MOOC completion rates hover around 5-10%, DataCamp has succeeded in reaching 30% completion rates for its free courses, and up to 80% for paid ones, thanks in part to offering deliberate practice of weak topic areas for users (Vaidyanathan, 2017), with hopes of building adaptive coursework in the future. Coursera also prioritizes adding practice to improve courses (Urban, 2017).

Should data science even be taught online? One study took the same graduate-level univariate statistics course and assigned students to either attend in-person or online versions of the course, and found that online students performed 7% worse (~1 standard deviation) on an assessment at the end of the course. One interesting finding was that males outperformed females in person, but females performed numerically better in the online setting. Online education may be able to somewhat ameliorate gender disparities in traditionally gender-imbalanced fields (Christmann, 2017), and allow opportunities for active participation across diverse age groups and ethnicities.

Three prominent adult learning theories offer insights into the design of lifelong data science education opportunities. Malcom Knowles developed andragogy theory in the 1970s. His focus was on integrating adults' prior experience into their education, prioritizing task-based learning, and working with learners' internal motives. Transformational learning theory

(Cranton, 2002) emphasizes adults' ability to learn through insights gained by ascribing new meaning to a previous experience, or reinterpreting old meanings. Experiential learning theory, on the other hand, focuses on adults' need to be actively involved in doing, reflecting, conceptualizing, and experimenting (Gutierrez, 2018). Challenges to andragogy theory have been raised (Blondy, 2007), but online courses should allow adults to establish their own learning goals and have flexibility with assessment within the bounds of course objectives, ideally via diagnostic experiences helping them assess gaps in their knowledge to further direct their learning.

A prominent cognitive neuroscientist – Lila Davachi – recently crafted empirically-grounded guidelines for adult learning. Increased sensitivity to distraction means learning should occur in a focused manner, without multi-taskiing. Adults should engage in more self-directed learning, as their attention becomes more selective. Learning experiences should be spaced in order for content to be remembered (ideally coupled with experiences of emotion to learn even more effectively) (Gutierrez, 2015).

In closing, many advances have been made in disseminating data science education, but open questions remain. How should learners' preferences for certain material be balanced with what they need to work through to accomplish their goals (Willems, 2017)? How will private sector control of large-scale learning change the political economy of education (Williamson, 2017)? How can we develop skill models and use partial completion trajectories (Thille et al., 2014) across individual courses to generalize to future learning experiences? Could we apply education data science to documentation, vignettes, and other software-related materials that learners seek out in an entirely self-directed fashion?

Top 3

- Wilson 2018 – Really nice article detailing the thinking behind DataCamp exercises. Clear examples illustrate their points.

  https://www.datacamp.com/community/blog/exercise-types-programming

- Kahan et al., 2017 – Deep dive into MOOC learner motivations via clustering based on demographics, participation, and achievement. I took away insights into what types of learners to expect in a MOOC, and how age and gender shape certain e-Learning behaviors.

  https://doi.org/10.19173/irrodl.v18i6.3087

- Thille et al., 2014 – Definitely the most methodologically innovative and impactful. Offers three diverse case studies in which data analytics enriched educational assessment. Aside from outlining approaches which could be creatively applied in other online learning contexts, also provides insights into latent factors contributing to learner behavior.

  https://web.stanford.edu/~cpiech/bio/papers/futureOfAssessment.pdf

Bibliography

Agarwal. (2014). MOOC Learners: Who They Are, What Motivates Them. Retrieved from https://www.huffingtonpost.com/anant-agarwal/mooc-learners-who-they-ar_b_4934941.html

Balakrishnan, G. K., & Coetzee, D. (n.d.). Predicting Student Retention in Massive Open Online Courses using Hidden Markov Models, 13.

Blondy, L. (2007). Evaluation and Application of Andragogical Assumptions to the Adult Online Learning Environment. *Journal of Interactive Online Learning*, *6*(2), 116–130.

Christmann, E. P. (2017). A comparison of the achievement of statistics students enrolled in online and face-to-face settings. *E-Learning and Digital Media*, *14*(6), 323–330. https://doi.org/10.1177/2042753017752925

Cornilessen. (2018). Context on Why We Started DataCamp. Retrieved from https://www.datacamp.com/community/blog/datacamp-reason-why

Cranton, P. (2002). Teaching for Transformation. *New Directions for Adult and Continuing Education*, *2002*(93), 63–72. https://doi.org/10.1002/ace.50

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing “Course”: Reconceptualizing Educational Variables for Massive Open Online Courses. *Educational Researcher*, *43*(2), 74–84. https://doi.org/10.3102/0013189X14523038

Gutierrez. (2015). 4 Elements to Effective Adult Learning. Retrieved from https://www.shiftelearning.com/blog/effective-adult-learning-neuroscience

Gutierrez, K. (2018). 3 Adult Learning Theories Every E-Learning Designer Must Know. Retrieved from https://www.td.org/insights/3-adult-learning-theories-every-e-learning-designer-must-know

Kahan, T., Soffer, T., & Nachmias, R. (2017). Types of Participant Behavior in a Massive Open Online Course. *The International Review of Research in Open and Distributed Learning*, *18*(6). https://doi.org/10.19173/irrodl.v18i6.3087

Kross, S., Peng, R., Caffo, B., Gooding, I., & Leek, J. (2017). The democratization of data science education. *PeerJ*.

Magnifico, A. M., Olmanson, J., & Cope, B. (2013). New Pedagogies of Motivation: Reconstructing and Repositioning Motivational Constructs in the Design of Learning Technologies. *E-Learning and Digital Media*, *10*(4), 483–511. https://doi.org/10.2304/elea.2013.10.4.483

Milligan, C., & Littlejohn, A. (2017). Why Study on a MOOC? The Motives of Students and Professionals. *The International Review of Research in Open and Distributed Learning*, *18*(2). https://doi.org/10.19173/irrodl.v18i2.3033

Mullaney, T., & Reich, J. (2014). Staggered versus All-at-Once Content Release in Massive Open Online Courses: Evaluating a Natural Experiment. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2499729

Piety, P. J., Hickey, D. T., & Bishop, M. J. (2014). Educational data sciences: framing emergent practices for analytics of learning, organizations, and systems (pp. 193–202). ACM Press. https://doi.org/10.1145/2567574.2567582

Thille, C., Schneider, E., Kizilcec, R., Piech, C., Halawa, S., & Greene, D. (2014). The Future of Data–Enriched Assessment. *Research and Practice in Assessment*, *9*, 5–16.

Thomas, M. P., Türkay, S., & Parker, M. (2017). Explanations and Interactives Improve Subjective Experiences in Online Courseware. *The International Review of Research in Open and Distributed Learning*, *18*(7). https://doi.org/10.19173/irrodl.v18i7.3076

Urban. (2017). Using data to transform the learning experience. Retrieved from https://blog.coursera.org/using-data-transform-learning-experience/

Vaidyanathan. (2017). Learn, Practice, Apply Data Science! Retrieved from https://www.datacamp.com/community/blog/learn-data-science-practice-apply

Vaidyanathan. (2018). Using Analytics To Drive Content Quality. Retrieved from https://www.datacamp.com/community/blog/analytics-content-quality

Willems. (2017). What We Learned From Teaching 1M People Data Science. Retrieved from https://www.datacamp.com/community/blog/what-we-learned-from-teaching-1m-people-data-science

Williamson, B. (2017). Who owns educational theory? Big data, algorithms and the expert power of education data science. *E-Learning and Digital Media*, *14*(3), 105–122. https://doi.org/10.1177/2042753017731238

Wilson. (2018). Exercise Types to Teach Programming. Retrieved from a.

https://www.datacamp.com/community/blog/exercise-types-programming