

EDUC 545: Big Data, Education, and Society
Spring 2018
Professor Ryan Baker

SYLLABUS

Instructor Info

Email: rybaker@upenn.edu

Office: GSE 439

Office hours: Wednesdays, 930am-1030am

Course time: Wednesdays, 2pm-350pm

Required Texts:

- Trochim, W.M.K., Donnelly, J.P. (2007) *The Research Methods Knowledge Base*.

Information on how to obtain course readings (aside from these texts) will be provided in class.

Course Goals: The growth of learning analytics and educational data mining has been met with both optimism and concern. Excitement about the possibilities of individualized, personalized, adaptive learning have emerged. But concerns that student privacy will be jeopardized, and that student futures will be forever shaped by data from long ago – or warped by an errant prediction about the student years into the future – have emerged as well.

In this class, we will discuss what learning analytics can do, what it has the potential to do for good, and what the potential is for harm. We will discuss multiple uses and applications of analytics, where simple steps can mitigate risk, the relationship between validity and risk, and where risk mitigation will do more harm than good. We will do so in the context of real-world educational systems, challenges, problems, and with reference to original sources as much as possible.

Course Pre-requisites: None, but some prior experience with statistics or data mining recommended.

Assignments:

This class will have one primary assignment with multiple sub-assignments.

In this project, students will propose a learning analytics application. In the first sub-project, due on February 27, they will propose the application. In the second sub-project, due on March 27, they will perform a needs assessment targeted towards articulating what societal or educational need the application addresses. In the third sub-project, due on April 17, they will discuss the risks and challenges inherent in their solution and how they can be mitigated. In the fourth sub-project, due on May 8, they will present their solution in the form of storyboards, mock-ups, or an interactive prototype.

Extensions for the assignments will only be available in case of instructor error or extreme circumstances (assignments in other classes, research studies, and so on do not count as extreme circumstances; serious injury, illness, or death in the family do count as extreme circumstances). Outside of these circumstances, late hand-ins will not be accepted (e.g. zero credit will be given).

No examinations will be given in this class.

Class participation involves both attendance and active (and constructive) participation in classroom discussions. While it is not expected that you will memorize every paper assigned for the class, it is expected that you will have studied the readings to the degree that you can participate actively in discussions.

Grading

- Project Proposal 20%
- Needs Assessment 20%
- Risks and Challenges 20%
- Final Project 20%
- Class Participation 20%

Course Schedule

Big Data, Education, and Society

Professor Ryan S. Baker

Wed, Jan. 17

The Emerging Era of Big Data in Education

Readings

- None

Wed, Jan. 24

What is Learning Analytics?

Readings

- Baker, R.S.J.d., Yacef, K. (2009) The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining, 1* (1), 3-17.
- Baker, R., Siemens, G. (2014) Educational data mining and learning analytics. In Sawyer, K. (Ed.) *Cambridge Handbook of the Learning Sciences: 2nd Edition*, pp. 253-274.

Wed, Jan. 31

Automated Intervention

Readings

- Corbett, A. (2001) Cognitive computer tutors: Solving the two-sigma problem. *UM2001, User Modeling: Proceedings of the Eighth International Conference*, 137-147.
- Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, 24(4), 427-469

Wed, Feb. 7

At-Risk Prediction

Readings

- Arnold, K.E. (2010). Signals: Applying academic analytics. *Educause Quarterly*, 33, 1-10.

- Bowers, A.J., Sprott, R., Taff, S. (2013) Do We Know Who Will Drop Out? A Review of the Predictors of Dropping out of High School: Precision, Sensitivity, and Specificity. *The High School Journal*, 96
- Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and action analytics: Three case studies to consider. *Research & Practice in Assessment*, 9.

Wed, Feb. 14

Reports for School Personnel

Readings

- Duncan, C., Hoxie, A-M., Miklasz, K. (2016) Understanding Data Dashboards: Where Needs and Designs Meet. White Paper. bit.ly/UnderstandingDataDash
- Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the assistent system. *Technology Instruction Cognition and Learning*, 3(1/2), 63
- Ocumpaugh, J., Baker, R., Slater, S., San Pedro, M.O., Heffernan, N, Heffernan, C., Hawn, A. (2017) Guidance Counselor Reports of the Assistments College Prediction Model (ACPM). *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 479-488.
- Holstein, K., McLaren, B. M., & Aleven, V. (2017). Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 257-266.

Wed, Feb. 21

Reports for Parents

Readings

- Broderick, Z., DeNolf, K., Dufault, J., Heffernan, N. & Heffernan, C. (2012). Increasing Parent Engagement in Student Learning Using an Intelligent Tutoring System with Automated Messages. *Journal of Interactive Learning Research*.
- Bergman, P. (n.d.) Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment Investment. White paper. <http://www.columbia.edu/~psb2101/BergmanSubmission.pdf>

Tues, Feb. 27

Assignments Due: Project Proposal

Wed, Feb. 28

Validity

Readings

- Trochim, W.M.K., Donnelly, J.P. (2007) *The Research Methods Knowledge Base*. Ch. 3-1, 7-1, 12-1
- Kane, M.K. (2001) Current Concerns in Validity Theory. *Journal of Educational Measurement*, 38 (4), 319-342.
- Hand, D.J. (1998) Data Mining: Statistics and More? *The American Statistician*, 52 (2), 112-118.
- Hand, D.J., Blunt, G., Kelly, M.G., Adams, N.M. (2000) Data Mining for Fun and Profit. *Statistical Science*, 15 (2), 111-126.

Fri, Mar. 2

Responses Due: Project Proposal

(Responses Prior to Mar. 6 will receive full credit)

Wed, Mar. 7

NO CLASS SPRING BREAK

Wed, Mar. 14

Rational Modeling and Model Validity

Readings

- Kovanovic, V. Gasevic, D., Dawson, S., Joksimovic, S., Baker, R.S., Hatala, M. (2016) Does Time-on-task Estimation Matter? Implications on Validity of Learning Analytics Findings. *Journal of Learning Analytics*, 2 (3), 81-110.
- Muldner, K., Burleson, W., Van de Sande, B., & VanLehn, K. (2011). An analysis of students' gaming behaviors in an intelligent tutoring system: predictors and impacts. *User modeling and user-adapted interaction*, 21(1), 99-135
- Paquette, L., de Carvalho, A.M.J.A., Baker, R.S. (2014) Towards Understanding Expert Coding of Student Disengagement in Online Learning. *Proceedings of the 36th Annual Cognitive Science Conference*, 1126-1131.

Wed, Mar. 21

Class Cancelled due to Snowstorm; University Closing

Wed, Mar. 28

Generalizability

Readings

- Trochim, W.M.K., Donnelly, J.P. (2007) *The Research Methods Knowledge Base*. Ch. 2-1
- Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., Heffernan, C. (2014) Population validity for Educational Data Mining models: A case study in affect detection. *British Journal of Educational Technology*, 45 (3), 487-501.
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68-84

Thurs, Mar. 29 EXTENSION

Assignments Due: Needs Assessment

Wed, Apr. 4

Adoption and Implementation Fidelity

Readings

- Ferguson, R. (2012) Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning (IJTEL)*, 4 (5/6), 304-317.
- Feng, M., Roschelle, J., Heffernan, N., Fairman, J., & Murphy, R. (2014). Implementation of an intelligent tutoring system for online homework support in an efficacy trial. *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 561-566).

Wed, Apr. 11

Privacy and Perceptions of Malevolence

Readings

- Arnold, K.E., Sclater, N. (2017). Student Perceptions of their Privacy in Learning Analytics Applications. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*.
- Rotenberg, M. and Barnes, K. (2013). Amassing student data and dissipating student privacy rights. *Educause Review*, Jan-Feb.
- Khalil, M., & Ebner, M. (2016). De-identification in learning analytics. *Journal of Learning Analytics*, 3(1), 129-138

Wed, Apr. 18

Discrimination and the Perpetuation of Bias

Readings

- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510-1529
- Garcia, M. (2016). Racist in the Machine: The Disturbing Implications of Algorithmic Bias. *World Policy Journal*, 33(4), 111-117

Thurs, Apr. 19 EXTENSION
Assignments Due: Risks and Challenges

Wed, Apr. 25
Beneficence

Readings

- Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room: the obligation to act. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*.
- Kraemer, F., van Overveld, K., and Peterson, M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology*, 13(3), 251-260.

Wed, May. 2
Data Sharing and Privacy

s

Readings

- Koedinger, K. R., Baker, R. S., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. *Handbook of educational data mining*.
- Brooks, C., Baker, R., Andres, J.M.L. (2017) Infrastructure for Replication in Learning Analytics. *Proceedings of the Workshop of the Methodology in Learning Analytics Bloc*.

Tues, May 8
Assignments Due: Final Project

Supplementary Readings
Big Data, Big Science, and Longitudinal Follow-up

Readings

- San Pedro, M.O.Z., Baker, R.S.J.d., Bowers, A.J., Heffernan, N.T. (2013) Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School. *Proceedings of the 6th International Conference on Educational Data Mining*, 177-184.

- Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4), 470-497.
- Andres, J.M., Baker, R., Gasevic, D., Siemens, G., Crossley, S., Joksimovic, S. (in press) Using the MOOC Replication Framework to Examine Course Completion. *Proceedings of the International Conference on Learning Analytics and Knowledge*.