EDUC 623: Big Data, Education, and Society
Spring 2020
Professor Ryan Baker

SYLLABUS

Instructor Info
Email: rybaker@upenn.edu
Office: GSE 439
Office hours: Thursdays, 230pm-330pm
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 21, they will propose the application. In the second sub-project, due on March 20, 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 3, they will discuss the risks and challenges inherent in their solution and how they can be mitigated. In the fourth sub-project, due on April 29, they will present their project as if they were submitting it to a potential funder.

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. 22
The Emerging Era of Big Data in Education

Readings

- None

Wed, Jan. 29
What is Learning Analytics?

Readings

- 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.
- Wise, A. F. (2019). Learning Analytics: Using Data-Informed Decision-Making to Improve Teaching and Learning. In *Contemporary Technologies in Education* (pp. 119-143). Palgrave Macmillan, Cham.

Wed, Feb. 5
At-Risk Prediction

Readings

- Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and action analytics: Three case studies to consider. *Research & Practice in Assessment*, *9*.
- Dawson, S., Jovanovic, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 474-478). ACM.
- Coleman, C., Baker, R., Stephenson, S. (2019) A Better Cold-Start for Early Prediction of Student At-Risk Status in New School Districts. *Proceedings of the 12th International Conference on Educational Data Mining*, 732-737.

Wed, Feb. 12
Reports for School Personnel

Readings

- Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the assistment system. *Technology Instruction Cognition and Learning*, *3*(1/2), 63
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-Designing a Real-Time Classroom Orchestration Tool to Support Teacher-AI Complementarity. *Journal of Learning Analytics*, *6*(2), 27-52.
- Wise, A. F., & Jung, Y. (2019). Teaching with Analytics: Towards a Situated Model of Instructional Decision-Making. *Journal of Learning Analytics*, *6*(2), 53-69.

Wed, Feb. 19
Reports for Parents

Readings

- Broderick, Z., O'Connor, C., Mulcahy, Heffernan, N. & Heffernan, C. (2011). Increasing Parent Engagement in Student Learning Using an Intelligent Tutoring System. *Journal of Interactive Learning Research*, 22(4)*, 523-550.
- Bergman, P. (in press) Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment Investment. To appear in *Journal of Political Economy.*
- Sousa, D. A., Luze, G., & Hughes-Belding, K. (2014). Preferences and attitudes toward progress reporting methods of parents from diverse backgrounds. *Journal of Research in Childhood Education*, *28*(4), 499-512.

Fri, Feb. 21
Assignments Due: Project Proposal

Wed, Feb. 26
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, Mar. 4
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.

Wed, Mar. 11
NO CLASS SPRING BREAK

Wed, Mar. 18
Rational Modeling and Model Validity

Readings

- 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.

Fri, Mar. 20
Assignments Due: Needs Assessment

Wed, Mar. 25
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

Wed, Apr. 1
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).
- Khachatryan, G. A., Romashov, A. V., Khachatryan, A. R., Gaudino, S. J., Khachatryan, J. M., Guarian, K. R., & Yufa, N. V. (2014). Reasoning Mind Genie 2: An intelligent tutoring system as a vehicle for international transfer of instructional methods in mathematics. *International Journal of Artificial Intelligence in Education*, *24*(3), 333-382.

Wed, Apr. 8
Data Sharing and Privacy

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.*
- 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.*
- Gardner, J., Andres-Bray, M., Brooks, C., Baker, R. (2018) MORF: A Framework for Predictive Modeling and Replication at Scale With Privacy-Restricted MOOC Data. *Proceedings of the 3rd Workshop on Open Science in Big Data.*
- The ASSISTments project data set webpage
  https://sites.google.com/site/assistmentsstudies/useourdata

Fri, Apr. 10 NEW DEADLINE
Assignments Due: Risks and Challenges

Wed, Apr. 15
Discrimination and the Perpetuation of Bias

Readings

- Garcia, M. (2016). Racist in the Machine: The Disturbing Implications of Algorithmic Bias. *World Policy Journal*, *33*(4), 111-117
- Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, *28*(2), 127-203.

Wed, Apr. 22
<u>Beneficence</u>

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.*
- Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology*.

Wed, Apr. 29
Class Presentations of Assignment 4: Final Project
Class will run an extra hour long this session

_____

Supplementary Readings

<u>Big Data, Big Science, and Longitudinal Follow-up</u>

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.L., Baker, R.S., Gasevic, D., Siemens, G., Crossley, S.A., Joksimovic, S.(2018) Studying MOOC Completion at Scale Using the MOOC Replication Framework. In *Proceedings of the International Conference on Learning Analytics and Knowledge*, 71-78.