

Program Dates

Module	Topic	Instructor	Week Starts	Week Ends
1	Introduction, Challenges, and Framework	Ryan S. Baker	8/24/2026	8/28/2026
2	Data and Measurement Validity	Wendy Chan	8/31/2026	9/4/2026
3	Feature Extraction and Feature Engineering	Haiying Li	9/7/2026	9/11/2026
4	Prediction Modeling and Metrics	Anthony Botelho	9/14/2026	9/18/2026
5	Causal Reasoning	Walter Leite	9/21/2026	9/25/2026
6	Ethics, Equity, and Algorithmic Bias	Shamya Karumbaiah	9/28/2026	10/2/2026
7	Cluster Analysis	Alex Bowers	10/5/2026	10/9/2026
8	Network Analysis	Bodong Chen	10/12/2026	10/16/2026
9	Sequential Pattern Mining and Temporal Analysis	Bodong Chen	10/19/2026	10/23/2026
10	Neural Networks and Deep Learning	Anthony Botelho	10/26/2026	10/30/2026
11	Natural Language Processing	Scott Crossley	11/2/2026	11/6/2026
12	Transformer and Foundation Models	Jinnie Shin	11/9/2026	11/13/2026
13	Knowledge Tracing	Ryan S. Baker	11/16/2026	11/20/2026
14	Knowledge Graphs	Seth Adjei	11/30/2026	12/4/2026
15	Data Visualization	Jaclyn Ocumpaugh	12/7/2026	12/11/2026
16	Data Management and Database Access	Michael Mogessie	12/14/2026	12/18/2026

Module	1. Introduction, Challenges, and Framework
Instructor(s)	Ryan Baker
Dates	8/24/2026 - 8/28/2026
AMA Sessions	- Zoom link for both sessions: https://adelaide.zoom.us/j/81318982670?pwd=l7VsxtqnisgaFfuaDlj6Qu2XmirwfW.1 - Monday, August 31, 4:00–4:30 PM ET

	Thursday, September 3, 5:00–5:30 PM ET
Topics	- Introduction to the course - Overall framework on learning analytics
Readings (if any)	- Baker, R., Siemens, G. (2022) Educational data mining and learning analytics. Sawyer, K. (Ed.) Cambridge Handbook of the Learning Sciences: 3rd Edition. - 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.
Assignments (if any)	- Discussion Forum Questions - ASSISTments

Module	2. Data and Measurement Validity
Instructor(s)	Wendy Chan
Dates	8/31/2026 - 9/4/2026
AMA Sessions	- Zoom link: https://upenn.zoom.us/j/4505681299?pwd=Y3luT1BXy3VXRE15YWp6ODJyOW1idz09 - Thursday, September 3rd from 4:00- 4:30pm - Wednesday, September 9th from 1:30-2:00pm
Topics	- Construct validity - Statistical conclusion and other types of validity - Internal validity and causality - External validity and generalizability
Readings (if any)	- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. <i>Educational measurement: Issues and practice</i>, 14(4), 5-8. - Cook, T.D. & Campbell, D.T. (1978). <i>Quasi-experimentation</i>. New York: Houghton Mifflin. - Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). <i>Experimental and Quasi-Experimental Designs for Generalized Causal Inference</i>. New York: Houghton Mifflin. - Coe, R., Waring, M., Hedges, L.V., & Ashley, L.D. <i>Research Methods & Methodologies in Education</i>. Los Angeles: SAGE.
Assignments (if any)	- Basic assignment: conceptual questions about the different types of validity - Discussion Forum Questions

Module	3. Feature Extraction and Feature Engineering
Instructor(s)	Haiying Li
Dates	9/7/2026 - 9/11/2026
AMA Sessions	- Tuesday, September 8, 6:00–6:30 PM ET (https://upenn.zoom.us/j/94606358097) - Thursday, September 10, 6:00–6:30 PM ET (https://upenn.zoom.us/j/94606358097)
Topics	- What is Feature Extraction and Feature Engineering? Why are they important? - Numeric Data - Categorical Data
Readings (if any)	- Romero, C., Romero, J.R., Ventura, S. (2014) A survey on pre-processing educational data. <i>Educational Data Mining: Applications and Trends</i>, 29-64. - Seeniselvi, T., & Nirmala, M. (2019). A survey on data preparation and feature engineering in machine learning. <i>International Journal of Advanced Research in Computer Engineering & Technology</i>, 8(5).
Assignments (if any)	- Discussion Forum Questions - ASSISTments – Feature extraction and feature engineering project

Module	4. Prediction Modeling and Metrics
Instructor(s)	Anthony Botelho
Dates	9/14/2026 - 9/18/2026
AMA Sessions	- Tuesday, September 15, 6:00-6:30 PM ET (https://ufl.zoom.us/j/5936742721) - Friday, September 18, 11:00-11:30 AM ET (https://ufl.zoom.us/j/5936742721)
Topics	- Prediction Modeling - Linear and Non-linear Models - Cross-Validation and Evaluation Methods - Metrics of Evaluation
Readings (if any)	- Bosch, N., & Paquette, L. (2018). Metrics for discrete student models: Chance levels, comparisons, and use cases. <i>Journal of Learning Analytics</i>, 5(2), 86-104. - Baker, R.S.J.d, Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., & Rossi, L. (2012). Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. In <i>Proceedings of the 5th International Conference on Educational Data Mining</i>, 126-133.
Assignments	- Discussion Forum Questions - ASSISTments - Prediction Model Comparisons

(if any)	
----------	--

Module	5. Causal Reasoning
Instructor(s)	Walter Leite
Dates	9/21/2026 – 9/25/2026
AMA Sessions	- Zoom link for both sessions: https://ufl.zoom.us/j/92175723512?pwd=ffPdobpm2lJPWEa8f7SztEmkCDbd33.1&from=addon - Tuesday, September 29, 6:00–6:30 PM ET - Tuesday, October 1, 6:00–6:30 PM ET
Topics	- Potential outcomes framework - Propensity score methods - Regression discontinuity designs - Heterogeneity of treatment effects
Readings (if any)	- Chapter 1 of Leite, W. L. (2017). Practical propensity score methods using R. Sage Publishing. - Chapter 5 of Leite, W. L. (2017). Practical propensity score methods using R. Sage Publishing. - Bloom, H. S. (2012). Modern Regression Discontinuity Analysis. Journal of Research on Educational Effectiveness, 5(1), 43-82. https://doi.org/10.1080/19345747.2011.578707 - Jacob, D. (2021). CATE meets ML. Digital Finance, 3(2), 99-148. https://doi.org/10.1007/s42521-021-00033-7
Assignments (if any)	- Discussion Forum Questions - ASSISTments: - Assignment 1 in Assistments: Causal Reasoning - Assignment 2 in Assistments: Propensity Score Analysis - Assignment 3 in Assistments: Regression Discontinuity Designs - Assignment 4 in Assistments: Heterogeneity of Treatment Effects

Module	6. Ethics, Equity, and Algorithmic Bias
Instructor(s)	Shamya Karumbaiah
Dates	9/28/2026 - 10/2/2026
AMA Sessions	- Tuesday, September 29, 2:00–2:30 AM ET (https://uwmadison.zoom.us/j/7012932689) - Thursday, October 1, 2:00–2:30 PM ET (https://uwmadison.zoom.us/j/7012932689)

Topics	• Algorithmic bias: approaches and limitations • Socio-technical systems in which AIED systems are embedded • Structural injustices that AIED systems reproduce
Readings (if any)	• Baker, R. S., & Hawn, A. (2021). Algorithmic bias in education. <i>International Journal of Artificial Intelligence in Education</i>, 1-41. • Holstein, K., & Doroudi, S. (2021). Equity and Artificial Intelligence in Education: Will "AIED" Amplify or Alleviate Inequities in Education?. In <i>The Ethics of Artificial Intelligence in Education</i>. Routledge. • Levinson, M., Geron, T., & Brighthouse, H. (2022). Conceptions of educational equity. <i>AERA Open</i>, 8, 23328584221121344. • Suresh, H., & Gutttag, J. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In <i>Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization</i>
Assignments (if any)	• Discussion Forum Questions

Module	7. Cluster Analysis
Instructor(s)	Alex Bowers
Dates	10/5/2026 - 10/9/2026
AMA Sessions	• Tuesday, October 6, 4:00–4:30 PM ET (https://upenn.zoom.us/j/93808023810) • Wednesday, October 7, 4:00–4:30 PM ET (https://upenn.zoom.us/j/98174076411)
Topics	• Cluster analysis combined with heatmaps • Hierarchical cluster analysis(HCA) and k-means • HCA Heatmaps in R using ComplexHeatmap
Readings (if any)	• Bowers, A.J., Zhao, Y., & Ho, E. (2022). Towards Hierarchical Cluster Analysis Heatmaps as Visual Data Analysis of Entire Student Cohort Longitudinal Trajectories and Outcomes from Grade 9 through College. <i>The High School Journal</i>, 106(1), 5-36. https://doi.org/10.1353/hsj.2022.a906700 HCA Heatmap open access R code tutorial: https://doi.org/10.7916/r1mg-yn37 • Gu, Z. (2021) <i>ComplexHeatmap Complete Reference</i>. Chapters 1-3. https://jokergoo.github.io/ComplexHeatmap-reference/book/ • Zhe, A.Y. (2022) K-Means Clustering An Explorable Explainer. https://k-means-explorable.vercel.app/ • Xu, R., & Wunsch, D. (2005). Survey of Clustering Algorithms. <i>IEEE Trans On Neural Networks</i>, 16. https://doi.org/10.1109/tnn.2005.845141
Assignments (if any)	• Replicate Bowers et al. (2022) HCA heatmap R markdown tutorial https://doi.org/10.7916/r1mg-yn37 • Discussion Forum Questions • ASSISTments

Module	8. Network Analysis
Instructor(s)	Bodong Chen
Dates	10/12/2026 - 10/16/2026
AMA Sessions	- Tuesday, October 13, 1:30–2:00 PM ET (https://upenn.zoom.us/j/2675369666) - Thursday, October 15, 10:00–10:30 PM ET (https://upenn.zoom.us/j/2675369666)
Topics	- Foundations of social network analysis - Techniques for collecting, organizing, and visualizing network data - Basic network measures - Practical guidelines on conducting network analysis in education
Readings (if any)	- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. <i>Science</i>, 323(5916), 892–895. https://doi.org/10.1126/science.1165821 - Poquet, O., & Joksimović, S. (2022). Cacophony of networks in learning analytics. In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), <i>The handbook of learning analytics</i> (2nd ed., pp. 38–45). SoLAR. https://www.solaresearch.org/publications/hla-22/hla22-chapter4/ [Optional] Scott, J. (2017). <i>Social network analysis</i> (4th edition). SAGE Publications.
Assignments (if any)	- Discussion Forum Questions - ASSISTments

Module	9. Sequential Pattern Mining and Temporal Analysis
Instructor(s)	Bodong Chen
Dates	10/19/2026 - 10/23/2026
AMA Sessions	- Tuesday, October 20, 2:00–2:30 PM ET (https://upenn.zoom.us/j/2675369666) - Thursday, October 22, 10:00–10:30 PM ET (https://upenn.zoom.us/j/2675369666)
Topics	Data in digital learning systems are often rich in temporal information that enables us to explore the dynamic changes in learning behavior at different time granularities. This module will engage participants to explore: - Basic concepts in temporal analysis and why temporality matters - Data manipulation with sequences - Exploratory sequence analysis - Association rules and frequent sequence mining - Critical issues in applying temporal analysis to digital learning data
Readings (if any)	Chen, B., Knight, S., & Wise, A. F. (2018). Critical Issues in Designing and Implementing Temporal Analytics. <i>Journal of Learning Analytics</i>, 5(1), 1–9.

	https://doi.org/10.18608/jla.2018.53.1 - Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. <i>Journal of Statistical Software</i>, 40(1), Article 1. https://doi.org/10.18637/jss.v040.i04 Technical manuals (no need to read sequentially) Mining sequence data in R with the TraMineR package (Chapters 1-8) Introduction to arules
Assignments (if any)	- Discussion Forum Questions - Self-paced sequential analysis lab

Module	10. Neural Networks and Deep Learning
Instructor(s)	Anthony Botelho
Dates	10/26/2026 - 10/30/2026
AMA Sessions	- Tuesday, October 27, 6:00-6:30 PM ET (https://ufl.zoom.us/j/5936742721) - Friday, October 30, 11:00-11:30 AM ET (https://ufl.zoom.us/j/5936742721)
Topics	- Multi-Layer Perceptrons and Neural Networks - Feature Embeddings and Model Training - Recurrent Neural Networks - Convolutional Neural Networks
Readings (if any)	- Jiang, Y., Bosch, N., Baker, R. S., Paquette, L., Ocumpaugh, J., Andres, J. M. A. L., ... & Biswas, G. (2018). Expert feature-engineering vs. deep neural networks: which is better for sensor-free affect detection? . In <i>Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part I 19</i> (pp. 198-211). Springer International Publishing. - Botelho, A. F., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S. A., & Beck, J. E. (2019). Developing early detectors of student attrition and wheel spinning using deep learning . <i>IEEE Transactions on Learning Technologies</i>, 12(2), 158-170. Optional Suggested Readings: - Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications . <i>Computer Science Review</i>, 40, 100379. - Aung, A. M., Ramakrishnan, A., & Whitehill, J. R. (2018). Who Are They Looking At? Automatic Eye Gaze Following for Classroom Observation Video Analysis . In <i>Proceedings of the Eleventh International Conference on Educational Data Mining</i>, 252-258.
Assignments (if any)	- Discussion Forum Questions - ASSISTments - Neural Network Training and Evaluation

Module	11. Natural Language Processing
Instructor(s)	Scott Crossley
Dates	11/2/2026 - 11/6/2026
AMA Sessions	- Zoom for both sessions: https://vanderbilt.zoom.us/j/92797764545?pwd=XnEZb0XUtMGT2FmolyPW8jnT9t5Ub3.1 - Wednesday, November 11, 3:00–3:30 PM ET - Thursday, November 12, 10:00–10:30 PM ET
Topics	- Tidyttext: Videos 1_tt to 3_tt - Tokenization: Video 4_tt - Word frequency: Videos 5_tt to 6_tt - Sentiment Analysis: Videos 1_sent to 5_sent
Readings (if any)	- Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment Analysis and Social Cognition Engine (SEANCE): An Automatic Tool for Sentiment, Social Cognition, and Social Order Analysis. <i>Behavior Research Methods</i>, 49 (3), 803-821. - Crossley, S. A., Paquette, L., Dascalu, M., McNamara, D., & Baker, R. (2016). Combining Click-Stream Data with NLP Tools to Better Understand MOOC Completion. In Gasevic, D., & Lynch, G. (eds.). <i>Proceedings of the 6th International Learning Analytics and Knowledge (LAK) Conference</i>. (pp. 6-14). New York, NY: ACM. doi: 10.1145/2883851.2883931 - Silge, J., & Robinson, D. (2017). Text Mining with R: A Tidy Approach. O'Reilly Media, Sebastopol, CA
Assignments (if any)	- Discussion Forum Questions - Sentiment analysis project (Kaggle): https://www.kaggle.com/t/3de26754a8464d49a12f7db70cecc676

Module	12. Transformer and Foundation Models
Instructor(s)	Jinnie Shin
Dates	11/9/2026 - 11/13/2026
AMA Sessions	- Zoom link for both sessions: https://ufl.zoom.us/s/6018466346#success - Monday, November 11, 11:00–11:30 AM ET - Wednesday, November 12, 2:00–2:30 PM ET
Topics	Language Models in NLP

	• Properties and Characteristics of Transformer models • Example Transformer Models
Readings (if any)	• Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. <i>Advances in neural information processing systems</i>, 30. • Devlin, Chang, Lee, & Toutanova (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://aclanthology.org/N19-1423.pdf
Assignments (if any)	• Discussion Forum Questions • Text classification model project

Module	13. Knowledge Tracing
Instructor(s)	Ryan Baker
Dates	11/16/2026 - 11/20/2026
AMA Sessions	• Zoom link for both sessions: https://adelaide.zoom.us/j/81318982670?pwd=l7VsxtqnisgaFuaDlj6Qu2XmirwfW.1 • Monday, November 16, 2:30–3:00 PM ET • Wednesday, November 18, 2:30–3:00 PM ET
Topics	• What is Knowledge Tracing? Why is it used? What is it used for? • Bayesian Knowledge Tracing • Logistic Knowledge Tracing/Performance Factor Analysis • Item Response Theory and Elo • Deep Knowledge Tracing • Memory Algorithms
Readings (if any)	• Pelánek, R. (2017). Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. <i>User Modeling and User-Adapted Interaction</i>, 27, 313-350.
Assignments (if any)	• Bayesian Knowledge Tracing exploratory activity (in Excel) • ASSISTments • Discussion Forum Questions

Module	14. Knowledge Graphs
Instructor(s)	Seth Adjei
Dates	11/30/2026 - 12/4/2026

AMA Sessions	- Monday, November 30, 2:00–2:30 PM ET (https://upenn.zoom.us/j/92096031100) - Thursday, December 3, 5:00–5:30 PM ET (https://upenn.zoom.us/j/97491515518)
Topics	- What are knowledge graphs? How do they apply to student learning? - Techniques for Building Knowledge graphs from students performance data - LFA, POKS, Q-Matrix, etc.
Readings (if any)	- Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. 2021. Knowledge Graphs. ACM Comput. Surv. 54, 4, Article 71 (May 2022), 37 pages. https://doi.org/10.1145/3447772 - Cen, H., Koedinger, K., Junker, B. (2006). Learning Factors Analysis – A General Method for Cognitive Model Evaluation and Improvement. In: Ikeda, M., Ashley, K.D., Chan, T.W. (eds) Intelligent Tutoring Systems. ITS 2006. Lecture Notes in Computer Science, vol 4053. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11774303_17 - Desmarais, Michel C., Xiaoming Pu, and Jean-Guy Blais. "Partial Order Knowledge Structures for CAT Applications." Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing. Retrieved [12/21/2023] from www.psych.umn.edu/psylabs/CATCentral . 2007. - Barnes, Tiffany. "The q-matrix method: Mining student response data for knowledge." American association for artificial intelligence 2005 educational data mining workshop. AAAI Press, Pittsburgh, PA, USA, 2005. Other Resources: Introduction to Knowledge Graphs - Q-Matrices [YouTube] [pptx] - Other Approaches [YouTube] [pptx] - Learning Curves [YouTube] [pptx]
Assignments (if any)	- Discussion Forum Questions - ASSISTments

Module	15. Data Visualization
Instructor(s)	Jaclyn Ocumpaugh
Dates	12/7/2026 - 12/11/2026
AMA Sessions	- Monday, December 7, 7:00–7:30 AM ET (https://upenn.zoom.us/j/95727101847) - Thursday, December 10, 4:00–4:30 PM ET (https://upenn.zoom.us/j/98007087436)
Topics	- What data do we have?/What improves learning? - How do all people perceive data? The art and the science of data vis and

	visualization sciences. • How do we know if we're helping?
Readings (if any)	• Wise, A. F. (2014, March). Designing pedagogical interventions to support student use of learning analytics. In <i>Proceedings of the 4th International Conference on Learning Analytics and Knowledge</i> (pp. 203-211). • Rogers, T., Gašević, D., & Dawson, S. (2016). Learning analytics and the imperative for theory driven research. <i>The SAGE Handbook of E-Learning Research</i>, 232-250. • Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. <i>Psychological Science in the Public Interest</i>, 22(3), 110-161. • Additional Resources available here: https://sites.google.com/view/discoverylearndashboards/home
Assignments (if any)	• Discussion Forum Questions • ASSISTments

Module	16. Data Management and Database Access
Instructor(s)	Michael Mogessie
Dates	12/14/2026 - 12/18/2026
AMA Sessions	• Thursday, December 17, 8:00PM-8:30PM EST (https://upenn.zoom.us/j/2687221525) • Friday, December 18, 8:00PM-8:30PM EST (https://upenn.zoom.us/j/2687221525)
Topics	• Understanding Databases • Database servers and clients • Accessing databases programmatically • Getting the information you need from a database • Observations, tips and tricks
Readings (Useful Links)	• DBeaver: https://dbeaver.io/download/ • Python: https://www.python.org/downloads/ • Jupyter Notebook: https://jupyter.org/install • MySQL Connector/Python: https://dev.mysql.com/doc/connector-python/en/ • SQLAlchemy: https://www.sqlalchemy.org/
Assignments (if any)	• ASSISTments

Grading Policy: The DSMDLP program uses a Pass/No Pass grading policy. To receive a passing outcome, participants must complete at least 70% of required program activities. This includes the assignments (i.e., ASSISTments and Kaggle assignment) and discussions, whether completed synchronously (in Canvas) or asynchronously (through vivi-SD). For asynchronous discussion, vivi-SD activities will be graded based on full **participation**, rather than attendance (video recordings may be used as needed to verify participation).