

# EDM in a Complex and Changing World

Ryan S.J.d. Baker  
Teachers College, Columbia University  
525 W. 120<sup>th</sup> Street  
New York, NY 10027 USA  
+1 (212) 678-8329  
Baker2@exchange.tc.columbia.edu

## ABSTRACT

We've started to answer the questions of what we can model through EDM, and we're getting better and better at modeling each year. We publish papers that present solid numbers under reasonably stringent cross-validation, and we find that our models don't just agree with training labels, but can predict future performance and engagement as well. We're making progress as a field in figuring out how to use these models to drive and support intervention, although there's a whole lot more to learn.

But when and where can we trust our models? One of the greatest powers of EDM models is that we can use them outside the contexts in which they were originally developed, but how can we trust that we're doing so wisely and safely? Theory from machine learning and statistics can be used to study generalizability, and we know empirically that models developed with explicit attention to generalizability and construct validity are more likely to generalize and to be valid. But our conceptions and characterizations of population and context remain insufficient to fully answer the question of whether a model will be valid where will apply it. What's worse, the world is constantly changing; the model that works today may not work tomorrow, if the context changes in important ways, and we don't know yet which changes matter.

In this talk, I will illustrate these issues by discussing our work to develop models that generalize across urban, rural, and suburban settings in the United States, and to study model generalizability internationally. I will discuss work from other groups that starts to think more carefully about characterizing context and population in a concrete and precise fashion; where this work is successful, and where it remains incomplete. By considering these issues more thoroughly, we can become increasingly confident in the applicability, validity, and usefulness of our models for broad and general use, a necessity for using EDM in a complex and changing world.

## SHORT BIO

Ryan Shaun Joazeiro de Baker is the Julius and Rosa Sachs Distinguished Lecturer at Teachers College, Columbia University. He earned his Ph.D. in Human-Computer Interaction from Carnegie Mellon University. Baker was previously Assistant Professor of Psychology and the Learning Sciences at Worcester Polytechnic Institute, and he served as the first Technical Director of the Pittsburgh Science of Learning Center DataShop, the largest public repository for data on the interaction between learners and educational software.

He is currently serving as the founding President of the International Educational Data Mining Society, and as Associate Editor of the Journal of Educational Data Mining. His research combines educational data mining and quantitative field observation methods in order to better understand how students respond to educational software, and how these responses impact their learning. He studies these issues within intelligent tutors, simulations, multi-user virtual environments, and educational games.