Data-Driven Instructional Practices: Why They Work

By Ryan S. Baker, University of Pennsylvania

Table of Contents

Ryan Baker is Associate Professor at the University of Pennsylvania and Director of the Penn Center for Learning Analytics. His lab researches engagement and robust learning to find indicators that can predict future student outcomes. Baker was the founding president of the International Educational Data Mining Society and serves as an Associate Editor of two journals. Baker has co-authored published papers with over 250 colleagues.

Introduction: What We Don't Know About Students

Envision a package containing an action figure—a gift for a 7-year-old perhaps—making its way from a warehouse in Memphis, Tennessee to a mailbox in Dayton, Ohio. At every step, as the package moves from a warehouse shelf in Memphis to the packaging area to the shipping truck to a depot in Dayton to the truck driver who will eventually ensure it reaches the child's excited hands, someone knows the exact location of that action figure.

The shipping company places such importance on that action figure's condition and location that it tracks every detail of the process. The toy is shipped in a truck that is appropriate for this type of cargo (as opposed to a refrigerated truck for medication, for example, or a truck full of industrial chemicals). It is delivered to the correct place (usually), and at high speed. This process has been mastered to the point that the shipping company can easily repeat it over and over every day. And yet, in many schools, we have no way of knowing as much about a child as the shipping company knows about a child's action figure.

As educators, we are missing important information. We don't know nearly as much as we'd like to about where each child is, in terms of his or her knowledge, or where each child is going, in terms of his or her progress. And even when we do know something important about a child, not everybody knows it. Key information may be hidden in a gradebook or a child's folder, for example, and that information may not make it to the parent or to a guidance counselor or school psychologist who is trying to help the child past a challenge. That information may not even make it to the child's next teacher.

Fortunately, this situation is changing. Data-driven instructional practices that have emerged over the last few years give teachers access to much better data about their students: data that they can use as they strive to make a difference in every child's life. Teachers have always been passionate about improving their students' lives, but when there are 30 homework assignments to grade, the time just isn't there. Now that we can automatically distill what a student knows from their work in online learning systems like *ALEKS* or Redbird, teachers have more opportunities to achieve their goals.

Data to Support Teachers

Imagine a group of children learning to work with algebraic equations. In a class of 24 students all working to solve for X in problems like X + 2X = 12 and 3X + 2X = 20, perhaps five of the students will have difficulty understanding that X represents 1X, and that X+2X therefore equals 3X.

How should a teacher respond to this possibility? It's not very efficient for the teacher to talk to the entire class about a difficulty that less than 20% of students might have. And what if, by chance, none of the students have that difficulty this year? Valuable instructional time may be wasted. At the same time, ignoring a potential difficulty students might have isn't a very good solution either.

Alternatively, the teacher can wait until she or he has better information about the exact struggles of each student. But in a traditional classroom, that might be fairly late in the learning process to return to the topic (for instance, not until after homework has been completed or even at the point of the exam). The teacher can certainly write feedback to the student on their exam or homework assignment when grading it. But will the student remember what they were thinking several days earlier? Will the student even read the feedback?

Contrast this situation with one in which the student is learning with a software package like *ALEKS* (Canfield, 2001) or Redbird (Suppes & Zanotti, 1996) that provides teachers with reports on student performance. First of all, *ALEKS* and Redbird (and most online learning systems nowadays) provide students with immediate feedback that, where appropriate, explains to students why their answers were wrong. Students can even obtain a worked example explaining how to solve the problem. Equally important, the system can inform the teacher about the struggles students are having.

This information can drive new teaching practices. Take, for example, a classroom that is actively working within a blended learning environment, one student to one computer. The teacher can obtain up-to-the-minute data on student performance, and when she/he looks at that information, she/he sees that five students have made the same error: X=6 rather than X = 4, for X+2X=12. This teacher recognizes the mistake that these students are making, and pulls them out of their regular activity to discuss why they made this error and explain how to solve this problem more effectively. In the meantime, the rest of the class continues with their ongoing learning activity without disruption. In 2015, Neal Miller and his colleagues termed this approach "proactive remediation."

In 2006, Mingyu Feng and Neil Heffernan wrote about another practice that was new at the time. The practice involved teachers using automated reports from online learning systems to check each student's degree of progress. Today, this type of data is commonly available to teachers in a range of online platforms. This enables a teacher to identify the degree of progress each student is making in order to determine if any students are falling behind. By discovering that a student is struggling more quickly than was previously feasible and take action early, there is more opportunity to make a difference.

Systems like *ALEKS* not only provide data on the progress a student is making, but can also let teachers know which students started the year already behind. These systems can also tell teachers which prerequisite learning a student must do in order to have a reasonable chance of learning the new material they are studying that year (Canfield, 2001). Not every student arrives on the first day of a new school year prepared to learn that year's material. When a teacher knows which prerequisites are missing, a student has a better chance of catching up than if he or she was simply launched into material for which he/she is not prepared.

Clearly, the information that modern online learning and blended learning programs can collect about student learning, knowledge and performance provides teachers with more opportunities to support their students. An additional trend that has become prominent in higher education, but has not yet emerged in most K–12 settings, is the use of data to determine not just what students *know*, but how *engaged* they are as well.

Disengagement matters for learning: If a student is not turning in their work on time, is not participating in class activities or is simply procrastinating and starting late, they are at considerable risk for performing poorly. Through technology, it is now feasible to automatically identify many forms of disengagement. Systems now in use by hundreds of universities and colleges identify which students are becoming disengaged, and present this information to instructors along with suggestions for how to support their students in re-engaging.

This type of approach has been shown in several studies (including a landmark study in 2014 by Milliron, Malcolm, and Kil) to improve the likelihood that students will pass classes and stay in college. The technology needed to implement this type of approach in K–12 already exists, but most schools that use it primarily focus on indicators such as disciplinary incidents like fighting or skipping class, a fairly late stage in the process of disengagement. Still, there is potential for leveraging the rich data available on K–12 students to help re-engage them as well. I personally expect to see these technologies developed and adopted within K–12 to a much greater degree within the next few years.

Why Data-Driven Approaches Work

Some potential questions about the use of online learning technologies include:

> How do we know the data is accurate?
>
> How does a computer know that a specific student is making progress?
>
> How can a computer tell when a student is struggling or with what?
>
> Should we trust a report coming from a computer?

It's important to remember that we never know for certain what is going on in a student's mind. Any mistake could stem from multiple causes, including not knowing how to solve the problem, having a misconception, or even carelessness on the part of the student.

For a classroom teacher working on his or her own without the support of an online platform, it's difficult to gain insightful information about what each student knows. A student might answer a question in class, but is that enough information to really know what's going on in the student's head?

Even in instances when there is clear information (for instance, immediately after a student turns in his or her homework), a teacher may not have enough time to study each student's pattern of responses to understand the implications. By comparison, a computer can look at all of the student's answers over time. As the student responds to questions, the system compiles the evidence into a profile of what the student knows.

School teachers can't be expected to determine whether a computer program like Bayesian Knowledge Tracing is better than, for example, a Recurrent Neural Network. Fortunately, they don't have to. There is now a very active area of scientific research, which compares computer programs that attempt to determine what the student knows or doesn't know. Researchers at dozens of universities and education companies are publicly debating which approaches are best.

While debate continues in journals and at scientific conferences, several studies (most recently an award-winning scientific paper by Khajah, Lindsey, and Mozer at the University of Colorado) have repeatedly reached an interesting conclusion: The difference between the best and most recent computer models and those from twenty years ago is surprisingly small—around 10%. With a small number of exceptions, whichever online learning software you are using to measure what students know, as long as it is measuring their correctness while learning (instead of just testing them and not helping them learn), it is probably good enough to be useful to you.

This is true because ***the real power of data lies in having a lot of it.*** A student using an online learning platform like *ALEKS* or Redbird could enter between 30 and 100 answers an hour. That's a ton of information. In that hour, a student may complete 10 different mathematics problems that shed light on the same mathematical skill or competency. If the student demonstrates the skill correctly 9 of 10 times, it does not take Google-quality artificial intelligence to determine that the student probably knows the skill. Similarly, if the student demonstrates the same misconception four times, it is easy to tell that the student needs learning support.

In other words, computers do not need to be able to tell things about a child that a teacher can't. They don't need to be as smart as an expert teacher. And they aren't. A computer program simply needs to be able to look at all of the information that a time-limited teacher cannot and deliver that information to the teacher. With that support, the teacher can then focus her or his time and effort where it will make the most difference: working with a struggling student or group of students to understand why they are struggling and helping them get back on track.

Summary

In this article, I have reviewed some of the recent work being done to support classroom practice through data. I have discussed the types of data that are now available to instructors through online learning platforms, as well as the types of new teaching practices that become possible through this data. I have also discussed why data-driven approaches work.

Ultimately, the takeaway is simple: Teachers are busy. A lot needs to happen in the classroom, and they do not have infinite time to do it. When computers can take over the job of combing student work to figure out what students know and don't know, teachers are able to focus their time on a much harder task—helping students learn.

References

Canfield, W. (2001). ALEKS: A Web-based intelligent tutoring system. *Mathematics and Computer Education,* 35(2), 152-158.

Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the assistment system. T*echnology Instruction Cognition and Learning,* 3(1/2), 63.

Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016) How Deep is Knowledge Tracing?. *Proceedings of the International Conference on Educational Data Mining*.

Miller, W.L., Baker, R., Labrum, M., Petsche, K., Liu, Y-H., Wagner, A. (2015) Automated Detection of Proactive Remediation by Teachers in Reasoning Mind Classrooms. *Proceedings of the 5th International Learning Analytics and Knowledge Conference,* 290-294.

Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and action analytics: Three case studies to consider. *Research & Practice in Assessment,* 9.

Suppes, P., Zanotti, M. (1996) Mastery Learning in Elementary Mathematics: Theory and Data. In P. Suppes & M. Zanotti (Eds.), *Foundations of Probability with Applications.* New York: Cambridge University Press, pp. 149-188.