Cutting-Edge Strategies from Other Sectors

Bryan C. Hassel

Bryan C. Hassel is co-founder of Public Impact[1]

Introduction

K-12 public education is poised to make great strides in how data are amassed and used by a variety of audiences. With the advent of annual testing in reading and math, and increasing capacity in states and districts to track individual students' progress over time, the quantity and quality of information available to everyone from parents and students, to teachers and administrators to policymakers and the general public, are already increasing dramatically. Hundreds of software packages and websites have emerged that aim to help these different audiences tap into this new pipeline of information. Nearly every classroom now houses one or more computers with an internet connection.

As exciting as these prospects are, it's important to place them in the context of what is happening more broadly in the world of data accumulation and use. While most of the developments under way in K-12 education are absolutely necessary and worthy of encouragement, their overall effect will be to bring U.S. public education data systems barely into the 21st century, if that. As states have crept toward milestones like assigning a unique identification number to each student, other sectors have raced ahead with stunning advances in how data and information are gathered, aggregated, packaged and used for a variety of purposes. The vast expansion of the internet and broadband access to its resources have revolutionized the world of data and information in business, nonprofits, and even in other governmental sectors. So even as K-12 moves its necessary incremental advances forward, now is a good time to pause and look at some of these more quantum leaps that have happened elsewhere—and could be relevant for K-12.

The potential payoff is large for many different actors in the K-12 world. Parents, for starters, are hungry for better information, whether they are choosing schools for their children, figuring out how to intervene with a school when it is not working well for their kids, or joining forces to press for policy change. What if parents could tap into not just the static, once-a-year, fairly uni-dimensional results that are now available for schools, but into a dynamic, multi-dimensional stream of data? What if that stream of information were so rich it could answer the myriad of questions that different parents have, such as "how do children like mine do at this school?" or "what can I do at home to help my children with such and such specific problem?" Teachers are another information-hungry audience. Every day, they struggle to figure out the best way to help different children overcome the diverse barriers they face to achievement. The amount of research-based information about "what works" is paltry relative to the number of questions they have. What if the data inherent in millions of daily teacher and student interactions could be harnessed to give teachers real insight into whether method X or Y is better? What if they could pose questions and obtain practical, data-based answers quickly? What if assessment results came back to teachers along with research-based suggestions for how to help each student overcome her shortfalls? And education leaders, from principals to district officials to state policymakers, need much better information as well. They currently have only blunt instruments for measuring how well individual schools and teachers are doing; they have virtually no instruments for predicting the trajectory of schools into the future. What if leaders had access to an evolving pool of data that illuminated the questions they need answered, like which of my teachers are doing what is needed to improve their instruction, or which of my district's schools are most in need of an infusion of new leadership? What if policymakers could more readily distinguish between schools that are on track to get better and those that are likely to languish without new action?

The good news is that in other sectors, leading organizations are finding ways to answer these sorts of questions and create these sorts of dynamic information streams. This paper examines two major trends in data and information and speculates about their applications in K-12 education. The first is "data-mining," which refers to applying sophisticated analytical tools to the wealth of data that organizations now have on their customers, suppliers, and markets. The second is what has been called "the wisdom of crowds," or tapping the implicit, collective information that resides in the heads of thousands or millions of individuals.[2] These trends involve the use of new technologies, but more fundamentally they involve a change in the way people think about "data." The word brings to mind official statistics, gathered by having people fill out forms and send them in to centralized repositories. These new trends, by contrast, seek to convert very different kinds of information into data. Data-mining, for example, often relies on information generated almost automatically through the day-to-day activities of employees and the users of products and services. Rather than filling out forms, people are contributing to this data mountain simply by doing what they do. The wisdom of crowds trend aims to tap a very different kind of information—knowledge and insights that would otherwise remain in the minds of isolated individuals.

This same kind of information exists, of course, in K-12 education. Every day, teachers, students, and parents engage collectively in billions of activities that are full of information content, if only that information could be collected and used. Students and teachers tap away at the computers that have now been installed in so many classrooms, leaving behind "click-trails" that reveal a lot about how they learn and process information. Those same people harbor myriad bits of knowledge and insight, information that can't generally be seen or used by others. Just as other sectors have figured out ways to turn this kind of information into data and use them for transformative purposes, K-12 education could as well.

To explore how, this paper describes each of these two trends in more detail, providing examples of how they have transformed activity in a sector other than K-12. It then speculates about potential applications of the ideas within K-12 public education. How could public education begin to gather—and then use—data in these radically different ways? And what sort of positive difference could that make to educators, parents, policymakers, and ultimately students? In some cases, this paper discusses nascent attempts to develop such applications in K-12 education, though this is by no means an exhaustive survey of what may well be bubbling along in "laboratories" across the world. Finally, the paper explores why these ideas may have slow uptake in public education, and what policymakers, funders and others might do to accelerate experimentation and adoption of the most promising developments.

Trend #1: Mining Insights from the Data Mountain

Both trends discussed in this paper have been made possible by the enormous expansion in networking in the last quarter century—connecting computers to many others within the same office or globally through the internet. Most organizations now link their employees' computers together in local networks, making it possible for them to share information with each other and maintain sources of information that are available to everyone. More radically, the internet has connected computers across not just organizational lines, but national lines. Anyone with internet access can communicate with others around the world instantaneously, and tap into growing stores of information on almost any topic. K-12 schools are fully part of this networked system. In 2005, according to the National Center for Education Statistics, 94 percent of U.S. public schools had access to the internet in instructional classrooms, up from 3 percent in 1994. Almost all of these schools (97 percent) enjoyed fast connections known as "broadband," enabling them to take advantage of all that the Internet has to offer. Teachers and students everywhere are putting their networked computers to good use. They are using the internet to find information; using email to communicate with experts and with each other; using district data warehouses to obtain more timely assessment results; and the like. But what K-12 has not done yet is take full advantage of this networked existence in the way other sectors have. Instruction more or less continues as it always has, with technological tools taking the place of more traditional resources, but playing the same roles.

One activity networking has enabled in other sectors is "data-mining"—gleaning insights from the mountains of information that are now available to companies, governments and, increasingly, individuals. Networking propelled data-mining in two ways. First, information that once resided in one person's file cabinet, ledger sheet, or spiral notebook can now be easily made available to an entire work unit, organization, or the public at large via the internet. Of course, organizations have always made efforts to gather and aggregate data from far-flung sources in order to make better decisions, but the networked world has made the process immensely easier and, in some cases, automatic. When we buy a bottle of salad dressing at the grocery store, the cashier's act of checking us out instantly shares information about the purchase, which can then be used in short-term ways: to notify the stockroom when salad dressing is getting thin on the shelves, or to trigger the next shipment of salad dressing to your neighborhood store.

But the cashier's act also creates data that can be mined for a variety of longer-term purposes: to inform buying strategies by the chain, for example, or (to the chagrin of some consumers) to increase understanding of our own buying habits, which can in turn suggest different ways of marketing, packaging, and otherwise encouraging us to spend more at the store over time. Without an electronic network that shares data instantly and high-powered computers to process them, this kind of mining wouldn't be impossible, but it would be sufficiently cumbersome that it would only happen rarely. Now, it can happen more or less constantly.

The other way networking has facilitated data-mining is by making it possible for consumers to buy products and engage in other activities on the internet, which in turn enables owners of websites to observe what people do when they are shopping for products, looking for information, seeking a mate, or carrying out any of the innumerable list of activities that we now engage in at our computers. As we click, we leave "click-trails" which become part of the data mountain, data that can be mined by whoever can see the trail.

It is useful to think of two different ways that organizations are able to dig into to the piles of information they are accumulating.[3] One is after-the-fact data-mining: taking the reams of information an organization generates and analyzing it to discern patterns and correlations. When a hurricane is barreling toward a town, what kinds of products should Wal-Mart stock more heavily in anticipation? Some guesses are obvious, like batteries and flashlights, but Wal-Mart's data-mining systems made it possible for the retailing giant to know for sure as it prepared for Hurricane Frances in 2004. At that time, some experts estimated that Wal-Mart's data warehouse held more than twice as much data as the entire internet.[4] By analyzing patterns from prior hurricanes, Wal-Mart was able to determine that batteries and flashlights weren't the only items customers would want to stockpile. Pop-Tarts were also high on the list, and at the very top was beer.

The hurricane analysis is just one example of how Wal-Mart puts its massive data warehouse to work. The company uses the constant stream of information flowing in from its stores to decide which suppliers to keep and toss; to press suppliers for faster deliveries or fewer defects; to plan new store openings; and to decide what to put on its shelves even when a hurricane isn't on its way. Wal-Mart also uses data to make sure its "everyday low prices" are not any lower than they need to be. Its information systems tell managers which items typically end up together in shoppers' carts; price-setters can use that data to mix lower and higher prices within a given likely basket of goods.

Another avid data-miner is Harrah's, the casino chain. As Michael Lewis made famous in *Moneyball*, his book about management by numbers in baseball, the conventional wisdom within an industry about what drives performance can often be off the mark. In the casino world, many assume that glitzy facilities, free steak dinners, and free hotel rooms are key attractors, especially for the "high rollers" that casinos think they want to cultivate. Harrah's, by contrast, embarked on an effort in the late 1990s to mine actual data about its customers in order to determine what would draw more of them in more of the time.[5] The company already had a frequent visitors' program, through which thousands of customers were routinely swiping their membership cards at Harrah's properties, thereby generating reams of data about everything they were doing. Every time a customer plays a game, checks into a hotel room, or has a meal at a Harrah's property, she or a casino employee swipes her card through a card-reading machine. Harrah's thus gains a record of what the customer is doing or buying, at which property, at what time of day, during what part of the year. It can "follow" the customer over the course of a visit, or across visits, to amass information about patterns of spending, gaming, and eating. And through the membership sign-up process, it knows other information about the customer: where she lives, for example, as well as other data she may have provided, like her income range or her interests. Harrah's aggregates all of this information in a central database, which analysts can then use to discover patterns and correlations that give Harrah's important insights.

By supplementing the card-swipe stream with survey and focus group data, Harrah's was able to learn a great deal. About a quarter of its customers were responsible for 82 percent of the company's revenue, but these generally were not the stereotypical "high rollers." Instead, they tended to be middle class retirees, people with some disposable income but also time to spend gambling. Most stopped by casinos near their own homes for a few hours; they weren't making big trips to far away destinations. Slot machines were their top activity. And they weren't all that loyal to Harrah's: they spent only 36 percent of the gambling dollars on average in the company's casinos.

These results, and others like them, dramatically changed Harrah's strategies. To encourage members to spend more of their money at Harrah's, they created a tiered membership system based on spending, with very visible differential benefits for higher-tier members, such as shorter waiting lines. They revamped the incentives that they offered customers, shifting the focus from hotel stays and meals (not highly valued by most of their top customers) to free chips for slot machines (highly valued). Realizing the centrality of slot machines, they invested heavily in figuring out how to increase slot revenue.

As data-mining has spread, so has the use of sophisticated techniques of statistical analysis to get the most value from the data. One example comes from the bane of all of our mailboxes: the seemingly never-ending flood of credit card offers. Though it may seem that credit card companies just send every offer to every address, in fact they try to make rational decisions about which mailings are worth sending to which customers. In its basic form, this kind of decision making relies on "regression analysis"—determining how influential a range of variables (age, zip code, occupation, income, previous credit history, and so on) are in individuals' decisions to accept a given kind of credit card offer. With the results of such a regression analysis, marketers can direct their mailings to people who are statistically more likely to say "yes." In theory, the result is fewer offer letters in recycling bins. They can also use this kind of analysis to figure out what kinds of marketing approaches work best with different kinds of customers, examining everything from what color envelopes are used to how sales pitches are worded. The result is an increasing ability to target specific customers in ways likely to get results.

Data-mining strategies like these are not limited to the for-profit sector. Urban police forces around the country, for example, have in the last decade adopted versions of the Compstat data system that many believe played a central role in New York City's dramatic reduction in crime rates. In a Compstat-style system, the data generated by daily police activity are immediately logged into a computer-based system: every arrest, but also every call from a citizen reporting a crime, every complaint, every ticket, every report filed by an officer, and so on. The resulting data can be mapped by location, resulting in vivid displays that make it easier for precinct commanders and top brass to see trouble spots. The data can be analyzed to identify trends and correlations, such as what time of day different crimes are most likely to occur. Top brass can use comparative information to hold precinct commanders accountable for crime in their jurisdiction, as New York and other cities now do through weekly Compstat meetings. And by looking city-wide, officials can make more informed decisions about how to allocate resources, such as New York City's massive increase of funding and staff for narcotics enforcement.

After-the-fact data-mining yields a lot of valuable information for organizations, but it has limits. Organizations often want answers to "what if" questions—if we tried this marketing approach or that one, which one would work better? While it's possible to look at after-the-fact data for answers (as in the credit card case), the best method for answering "what if" questions is the technique known as "randomized experimental design." In this method, some subjects are randomly assigned to a "treatment" group that receives some kind of intervention—a new medicine, or a new form of advertising. Other subjects, in the "control" group, do not get the intervention. Researchers can then track each group's outcomes: their health gets better or worse, they buy more or less. Since people were randomly assigned to one group or the other, if the two groups' outcomes are different, it must be due to the fact that the treatment group received the treatment, and the control group did not.

The advent of the internet has made it possible for companies to use randomization like this to conduct a constant stream of experiments on the users of their websites at a very low cost, and in a way that is very unobtrusive from the users' point of view. They can then use the results of these experiments to change their products, services, and user interface in ways that yield more sales (or whatever behavior they want to induce). As Babson College technology professors Bala Iyer and Thomas H. Davenport put it, "It's relatively easy to perform randomized experiments on the Internet: Simply offer multiple versions of a page design, an ad, or a word choice."[6]

If we visit an Amazon.com page for a particular book, for example, we will almost certainly see a small picture of the book's cover. But this hasn't always been true on Amazon: in the early days, Amazon faced an enormous task if it was going to scan in or otherwise acquire digital images of literally millions of book covers. To make that kind of investment, Amazon had to believe doing so would be worth it. They used randomization to find out. When visitors came to some books' pages, they were randomly offered a thumbnail picture of the book's cover, or not. Amazon could then track subsequent purchase behavior of cover viewers and non-cover viewers. As it turns out, people are more likely to buy when they see a cover, and so now covers are ubiquitous on Amazon. The same process helped Amazon make countless other decisions, such as the move to encourage publishers to let surfers "search inside" of books and view excerpted content.

Many other companies have used randomization as well. "Every day," say Iyer and Davenport, "Google does thousands of experiments for their own benefit." Google even allows the customers of its web-advertising service to run their own experiments, testing which ads and search terms yield the best results.[7] Another heavy user of web experiments is the credit card and financial services purveyor Capital One, which runs thousands of experiments annually related to new products, web layouts, and any number of other variables, discarding 99 percent of what they try based on poor response, but benefiting greatly from the 1 percent that succeed.[8] Randomization is even possible in non-internet settings. When Harrah's launched a campaign to woo back former customers to its casinos, it wasn't sure which offers or incentives would work best. So its telemarketers randomly made different offers (a steak dinner, free casino chips, hotel stays) and recorded the results, which fed into Harrah's revamped customer loyalty program described above.

What kind of expanded role could data-mining play in K-12 education? Of course, analytic techniques such as regression and randomization are already well-known in K-12, with researchers all over the world running regression analyses on available data sets and a heightened interest at the federal level in recent years in randomized experimental design as the "gold standard" of casual research. What's different about the efforts in other sectors described above is their ongoing nature within the lives of the organizations that use them. Amazon's random "study" of the use of book covers was not a multi-year academic study that went through peer review before landing in a scholarly journal. Rather, it was a research activity undertaken by Amazon in the course of doing business, an activity that is repeated over and over to answer different questions. The same goes for the regression-based data-mining efforts: the organizations making the most of them have institutionalized systems in which, as more data come in, the organizations continuously apply an evolving set of analytical techniques in order to generate an ongoing stream of insights that inform practice. They are not just looking at the equivalent of end-of-year test scores over the summer and using that one snapshot as the basis for their planning and decision making.

Though public schools are clearly different in many ways from organizations like Wal-Mart, Harrah's, and Amazon.com, they do have exactly the same sort of ongoing stream of experiences that could potentially be fodder for data-mining. Every day, teachers explain concepts to students using a variety of different techniques. Students answer questions, fill out worksheets, work problems on the whiteboard, read aloud, work with manipulatives, play educational games, and so on. Teachers respond to student effort in verbal, written, and other ways. Students interact with each other as they engage in learning. Collectively, across all public schools in the U.S., this activity amounts to billions of data points that could, in theory, be mined to answer vital questions of practice.

Individual teachers may learn from their own repeated experience when they notice patterns in how students respond to their instructional techniques. But this kind of learning is inherently limited in two ways. First, a teacher's own learning is potentially biased because of the idiosyncrasies of her specific students and because of random chance. She may think that technique X works, but in fact it may just have appeared to work in a few instances that really should not be generalized. Second, even accurate learning by individual teachers remains largely unshared with others, especially on a large scale. There may be ways to foster more large-scale sharing like this—a topic discussed in the next section. But it would be vastly more efficient and valuable if somehow all of this experience could be "mined" in the way that the organizations above are mining their data piles.

The obvious problem is that unlike Amazon's customers, who leave click-trails behind them, or Wal-Mart shoppers, who at least have to record their purchases with the cashier before they leave the store, public education has no ready way to capture most of these data. They exist only in the fleeting interactions that go on within the walls of schools. And somehow trying to capture these interactions from outside the stream, through third-party observation or teacher logs or the like, would be ridiculously costly. The beauty of the Wal-Mart, Amazon, and Harrah's systems is that they collect data in the natural course of activity, rather than as some kind of costly, add-on task that requires enormous ongoing effort.

Rather than throw up our hands at this point, it seems worthwhile to ask the question: what changes in K-12 would make it possible to capture more of this daily activity data? Here are two ideas worth exploring. First, a dramatic expansion of handheld devices by both teachers and students could capture and share much of the daily-experience information that currently evaporates into the ether. A company called Wireless Generation, for example, has pioneered the use of handhelds by teachers to administer live reading assessments, with the results instantly uploaded and available for analysis by the teacher, his peers, the school principal, and potentially, higher-ups. In a more conventional setting, a teacher might administer a set of questions to a student and record the student's response on paper. Later, the teacher or some other person would have to go back and grade the test, going answer-by-answer and marking each right or wrong. Only after this grading process was complete would the teacher have the student's results. And to conduct any kind of analysis of how a student is progressing over time, or what kinds of patterns are emerging across a class or grade level, someone would need to enter the results manually into a computer, and then manually generate reports.

With the handheld system, the device guides the teacher through the live assessment, telling her what to ask when and keeping time if the test has time limits. As soon as the assessment is done, the device can "grade" the test and show the teacher the results. And when connected to a computer, the handheld automatically transfers the data to the network, making it available to the teacher, administrators, and potentially parents. With the growing prevalence of wireless networks, even this transfer step will become automatic: a student's completion of an assessment will immediately register her results in the larger system. Pre-established reports make it easy to look at individuals' progress over time or patterns across classrooms or grade levels. Increasingly, the system can also give teachers ideas for activities that can be used to remedy specific challenges revealed by the data.

The potential for expanded use of this basic platform is significant. If teachers and students used handheld devices to record more and more of their interactions, the amount of information captured would be richer. And if the data from these devices could be aggregated at higher and higher levels (with appropriate safeguards for student and teacher privacy), the sort of data-mining described above could become much more common and valuable. Instead of just running regressions based on year-old, end-of-grade test data, analysts would have a rich array of real-time data at their fingertips. The devices could even be used to conduct the kind of randomization experiments Amazon and Capital One use. A handheld could prompt a teacher randomly to try one or another method of explaining some material, and then capture the results of a quick assessment to see what students had learned. Across thousands of similar experiments, analysts could see which approaches generated better results, and feed that information back to teachers through their handhelds as well.

A second, related idea is to make much better use—from a data-generating point of view—of all of the time that students now spend in front of computers for educational purposes. Computers are nearly ubiquitous now in U.S. schools, with 95 percent of fourth graders having school-based access to computers. Schools average about one computer for every 3.8 students. In addition, the prevalence of online coursework (through entire virtual schools, or through one-off online courses) has increased dramatically. Half of the nation's states have established a virtual school.[9] And according to the North American Council for Online Learning, U.S. students enrolled in about 1 million online courses in 2007–08.[10] Clayton Christensen and Michael Horn project that by 2013, 10 percent of all courses will be computer-based, with the percentage reaching 50 percent by 2019.[11]

Though scholars have debated the instructional value of these machines and courses, here the question is different: whether they can be harnessed as a data-generating engine for K-12 education. As with customers on Amazon, students sitting at these terminals working through computer-based instructional modules generate click-trails. They select activities, give answers, and respond to feedback. They sit and think (and don't click), or they forge ahead quickly. They learn from their mistakes (or don't). All of this activity could, in theory, be captured and analyzed in the same way that Amazon and Google capture and analyze data about what their customers are doing online.

These interactions also hold the most obvious opportunities for randomization. As noted above, the country has seen increased interest in randomized experimental design in recent years, but gold standard randomization studies are very expensive and time-consuming. Even at the accelerated rate we now see, it is unlikely that such studies will address more than a tiny fraction of all the problems of practice educators face daily. Randomization on students' computers, by contrast, could yield insights much more quickly on a wide array of problems, such as the best ways to present different kinds of material, the best ways to present it to students with different learning approaches, the best ways to respond to students when they don't understand a concept or face some barrier, the best ways to motivate children to take on challenging work, and so on. If thousands of children were having repeated interactions with such a system, with data on their experiences, behavior, and outcomes rolled up into a central warehouse, we could learn much faster what works best, and what works best with different children. And the learning could be much more dynamic and continuously improved, versus the current cycle of spending years developing approach X, and then years testing it, and then years disseminating it. Techniques that proved effective with a certain type of student would automatically be used more frequently with that kind of student going forward.

Admittedly, our nation's schools are a long way from being able to harness this kind of data. Though computer use in schools has become much more prevalent, it is greatly fragmented. In contrast to Amazon.com, which is a single organization that can observe its customers' actions within a unified platform, decisions about what kinds of educational software to use are made by individual school districts, schools, teachers, and even students. Though millions of school children may be sitting in front of computers at any one time, they are by no means all creating similar or comparable click-trails that are ready for analysis. To overcome this challenge, a major effort would be required on one or both of two fronts: creating incentives for educators and students to use platforms that are set up to feed data into an analytical engine like the one envisioned here (and thus making click-trails less fragmented); or developing mechanisms by which click-trail data from these now diverse sources can be aggregated and analyzed despite their different origins. Either would require substantial investment and innovation.

As this last point demonstrates, it would be an understatement to say that implementing these ideas would require overcoming significant technical hurdles, as well as cultural obstacles. The paper will return to these hurdles in the concluding section. For now, the important point is that K-12 education is currently letting a great deal of valuable information slip through its hands. Substantial hurdles notwithstanding, it's worth applying the nation's considerable technical and entrepreneurial talent to letting less of it slip away.

Trend #2: Tapping the "Wisdom of Crowds "

Trend #1 is still hierarchical in nature—some kind of central intelligence is monitoring behavior, generating experiments, and then crunching data to glean insights. Also enabled by the networking revolution are technologies that tap into what author James Suroweicki called “the wisdom of crowds” in the title of his best-selling book. The idea behind the wisdom of crowds is that the collective knowledge of large numbers of people is often more accurate or “wise” than the analysis of a single expert. For example, when retailer Best Buy asked experts to forecast its gift card sales for February 2005, the experts’ estimate was 95 percent accurate. When they emailed hundreds of employees the same question and averaged the 190 responses that came in, the “crowd” estimate was 99.5 percent accurate.[12] It’s not that most of the 190 responders, as individuals, were smarter than the experts. It’s that the collective information of the 190, which incorporated the many perspectives that individuals throughout a complex organization had on the question, added up to a smarter average than what the experts could generate.

The expansion of electronic networks has enabled people to tap into the wisdom of crowds in several ways. One way is by facilitating actual collective work among far-flung people. A common expression of this approach is the “wiki,” which is basically a website that users can edit. The most famous wiki, Wikipedia, has created an enormous encyclopedia from user submissions, which are modified over time by users who visit the entries. Complex software packages, even entire computer operating systems, are now routinely being written through “open source” programming, in which far-flung, voluntary networks of programmers contribute “code” to a larger software development effort, the results of which are freely available for users to see, and improve upon further.[13]

While these applications are interesting, more relevant to this paper are efforts to tap crowd-wisdom to generate “data,” which can then be analyzed and used. Two developments in particular are worth exploring. One is a crowd-oriented version of the data-mining described above. Whenever we view or buy an item on Amazon.com, Amazon takes note. Since Amazon tracks this over time, and over millions of visitors, its data systems capture patterns of viewing and buying that Amazon then shares with users in various ways on the site. For example, it displays a list of other items that were viewed or bought by others who viewed or bought the items we are considering. It offers customized “recommendations,” again based on what other users who appear to have similar preferences have bought. It sends emails to us letting us know about releases of books that are linked in this way to our apparent interests. And of course, Amazon is just one of many examples of this kind of system.

Most online retailers have some version of the Amazon method. Most news and information sites, from *The New York Times* to the Fordham Institute's *Education Gadfly,* supply users with lists of "most e-mailed articles." Google's underlying method for ranking a webpage is based in large part on how many links to that page come from other pages (especially pages that are themselves highly ranked). The "crowd" in this case is all the people who have created other webpages. What's important about these systems is that all of these suggestions and recommendations and rankings are not manufactured by some expert who has analyzed the data and come to these conclusions. Instead, they are generated by the behavior (and one hopes, the wisdom) of the crowd.

The examples in the previous paragraph happen behind the scenes, in the sense that users are not necessarily aware that they are contributing to the collective wisdom as they peruse and buy items on Amazon.com, or look for DVDs to rent on Netflix, or share *Gadfly* tidbits with their colleagues. Like honeybees going after nectar and inadvertently spreading pollen, users going about their own shopping and information-seeking are contributing to crowd wisdom unintentionally.

In other cases, sites ask users to contribute explicitly. Amazon users can rate items on a star system and write reviews that other users can read; average star ratings feature prominently in each item's display. eBay encourages buyers and sellers to provide feedback on each other's behavior in a transaction. Zagat elicits consumers' reviews of restaurants, hotels, and attractions as a supplement to its own expert ratings. Digg.com provides a way for Internet users to indicate that they "digg" certain online news stories, and then assembles its own news site based on what these users are telling them through their clicks. Tripadvisor.com invites users to rate hotels and attractions, and then displays those rating for other travelers. Over 15 million reviews now populate the site.

Crowd wisdom, like in these examples, has important potential advantages, especially when there is a strong subjective component to the information users are seeking. In the case of a restaurant, for example, one might be interested in the opinion of the one reviewer that the *New York Times* sends to file a report. But that reviewer inevitably has particular tastes, which may differ from yours. She may be on the lookout for certain features of the dining experience that may be more or less important to you. And she will have had one particular experience that may or may not be a good indicator of what diners can generally expect: one particular set of appetizers, entrees and desserts on one particular night. With crowd wisdom, in theory it is possible to focus on reviews of others who have similar tastes (i.e., those who like other restaurants that you like) or who are seeking similar features (e.g., kid-friendliness, a romantic atmosphere). And by aggregating over dozens or even hundreds of meals, the collection of ratings has less chance of being skewed by particularly good or bad experiences that are not the norm.

A second example of crowd wisdom tapped by technology is the emergence of "prediction markets." Think back to the Best Buy sales forecasting example above, in which employees in the aggregate did a much better job than experts of projecting holiday card sales. In that example, employees had no real reason to give their best estimates; they had no stake in the outcome. They also had no wide scale way of getting cues from each other about whether their estimates were high or low: they just emailed them in and were done with it. Prediction markets seek to improve on crowd-based forecasting by giving predictors an incentive to predict well. The resulting market "prices" provide useful information that predictors can use over time to make better predictions. Most people are already familiar with some existing prediction markets. Stock markets, for example, serve the purpose of raising and allocating capital for companies, but in the process, all the buying and selling reveal information about how highly the market values different companies. Another example is betting at the horse track—before a race, the shifting odds on different horses represent the collective judgment of all the bettors about each horse's probability of winning.

Less familiar are efforts to create prediction markets to serve some organizational or social purpose. One set of examples of such prediction markets resides at the Iowa Electronic Markets (IEM), a project of the University of Iowa's College of Business.[14] IEM has established a handful of prediction markets, including several related to upcoming presidential elections. As of this writing, for example, IEM was running a market designed to predict the outcome of the 2008 general election. Users could buy securities that would pay $1.00 times the percentage of the vote a given party receives in November 2008. So if you were holding a Democratic Party security and the Democratic candidate won 49 percent of the two-party vote on Election Day, you would receive 49 cents. On April 24, 2008, Democratic shares closed at 53.3 cents, versus 47.6 cents for Republican shares. In markets like this, the price of a party's shares can be interpreted as the market's estimate of the vote-share it will win. This kind of prediction market has proven remarkably accurate when compared to the other primary method for predicting election outcomes: stratified random polls. According to one study of 12 years of elections markets, the average market missed the true vote share by 1.49 to 1.55 percentage points, compared to an average of 1.93 points for polls in the same elections.[15]

Some organizations have started using prediction markets internally, such as Google. At the time of a McKinsey roundtable discussion published in April 2008, Google had used prediction markets to elicit forecasts on 275 questions on subjects ranging from demand for Google's products ("how many people will use [Google's email service] Gmail in the next three months?") to the company's performance (will a certain project meet a deadline?) to major events in the industry (such as mergers and acquisitions of other companies).[16]

One value of this kind of prediction market—as well as "real" markets such as stock exchanges—is that their existence creates an incentive for predictors and traders to find information that helps them make good predictions and trades, which in turn creates incentives for others to amass and provide helpful information to them. Witness the profusion of websites, newsletters, books, television talk shows, and other sources of information about stocks. Investors stand to gain from being well-informed; information providers stand to gain, through advertising or subscription revenue, from providing data and insights that investors value. So while the markets themselves provide one form of data (stock prices, predictions about outcomes of elections and other real-world events), they also stimulate the creation and dissemination of other forms of data. This secondary data-eliciting effect is arguably the most powerful information-generating aspect of trading markets.

Do these wisdom of crowds trends have any potential value in K-12 education? The most obvious fit is the first set of examples discussed above, in which website visitors rate products (implicitly or explicitly), and those ratings are then shared with other visitors. Numerous potential uses of this technology come to mind in the education setting. One where there is already considerable action is in websites designed to help parents evaluate and choose schools. GreatSchools.net, for example, provides detailed information about schools, from street address and phone number to demographics to test results. Increasingly, it has sought to supplement this top-down data with bottom-up parent ratings and comments. From a prominent spot on any school's page, users can click "Rate it!"—calling up a window that asks for a one to five star rating and gives space for a 10-150 word narrative comment. Each school's page then shows the average parent rating alongside GreatSchools' own rating, with a link to the narrative comments.[17] The Savvy Source seeks to provide a similar service for parents interested in preschools, inviting parents to fill out a survey on a given preschool that includes the ubiquitous five star rating as well as a series of other questions.[18]

Some nascent websites even aim to bring students' perceptions into the mix. RateMyTeachers.com enables students and their parents to rate K-12 teachers on "easiness," "helpfulness" and "clarity." As of June 2008, this site contains 10 million reviews of 1.5 million teachers nationwide. RateMyProfessors.com offers the same for higher education, claiming 6 million reviews of a million professors at 6,000 colleges and universities.

Another potential use is to help teachers in their ongoing quest for useful lesson plans, instructional materials, and advice in general about how to address problems of practice. There are few if any tasks that Ms. Jones in a Dayton elementary school encounters that haven't been encountered hundreds or thousands of times by other similarly situated teachers. What teachers don't have is any way of seeing how their peers have dealt with these common tasks and, vitally, any way of knowing which of the strategies their peers have tried have actually been effective. In theory, this could be enabled by the technology that is now on most teachers' desks.

Indeed, the internet is already replete with websites that offer material for teachers, but there are few mechanisms (aside from general purpose search engines like Google) to help educators separate the wheat from the chaff—which is the real potential value of the wisdom of crowds. If a critical mass of teachers began using the internet for this purpose, it is possible to imagine the collective wisdom being mobilized to help educators. If it were easy for teachers to rate the resources they find online, either explicitly (assigning a one to five star rating, posting a comment) or implicitly (revealing their preferences as lesson plans or student materials are clicked on), teachers would at least have a window into which resources were most popular. If services used Amazon-style matching technology to discern users' own "shopping" patterns over time, the "advice" users received could become even more powerful. It would generate a whole new form of data about the perceived quality and value of different instructional approaches and resources, data amassed from the opinions expressed by individual users.

Nascent efforts are underway on this idea as well. Yahoo! is set to release Yahoo! Teachers at some point in 2008. Teachers will be able to share lesson plans and projects, which other teachers can then search and pull into their own online "portfolios." As of this writing, Yahoo! Teachers did not contain much information about how crowd wisdom would be mobilized to help users tell wheat from chaff.[19] By contrast, TeacherTube, launched in March 2007, utilizes many of the crowd-wisdom techniques described above. This service allows teachers to upload videos to the web, either to demonstrate how they carry out instruction on a certain topic, or to post educational videos designed for students. The more popular videos rise to the top. Users can also rate videos, though the usual five stars are replaced by what looks like a cross between an apple and an old-fashioned TV set with rabbit ears. Users can "tag" videos—label them by subject or other keywords, which enables others to find relevant videos more easily.

Beyond these parent, student, and teacher examples, there are many other potential ways crowd ratings could be used to point K-12 actors in the right direction. Teachers could rate the education schools and licensing programs they attended to become teachers, or even specific courses or instructors. They could rate the professional development offerings in which they engage. The same goes for school leaders, whose wisdom could be enlisted on administrator preparation programs or summer leadership institutes. Teachers and principals could rate their schools and districts on metrics related to what it is like to work there. And so on.

One challenge with crowd-driven rating systems is that by their nature they tend to emphasize popularity, which may or may not be a good proxy for quality when it comes to something like a school, a teacher, or an instructional approach. RateMyProfessors.com, for example, has been challenged by a Central Michigan University analysis showing that professors whose courses are easier or who are rated as better looking rise to the top of the website's rankings.[20] Though another study showed that the site's ratings are highly correlated with the results of a widely used student evaluation system that includes more quality controls, questions persist about the value of RateMyProfessors, at least for students who are looking for high-quality instructors rather than attractive ones. The most promising approaches to tapping crowd wisdom, therefore, may involve combining crowd judgments with expert assessments and objective data. On GreatSchools.net, for example, a parent can read peer reviews, but can also see the site's own rating of the school (expert assessment) and direct information about the school's performance on state tests (objective data).

Exacerbating the quality question is another serious challenge all of these efforts face: getting sufficient volume of users to make them work in the ways described above. The wisdom of crowds requires, yes, a crowd, and these sites have generally struggled to attract them. GreatSchools.net's parent ratings area for this author's children's elementary school, for example, had just nine parent ratings in April 2008, several of which provided mostly vague praise. Parents considering the school would learn little of value from these ratings. RateMyTeachers.com reviewed only one of the school's teachers. The teacher sites as well tend to lack critical mass. The most viewed video on TeacherTube as of April 2008 had been watched over 500,000 times, but the numbers ramp down precipitously: the 101st most watched had only about 12,000 views, the 501st about 4,000. Low numbers create a vicious cycle: with few users' wisdom being tapped, the sites can't dispense much in the way of crowd wisdom; but without crowd wisdom, it is difficult for them to become the kind of go-to resources that attract large numbers of reviewers.

Is this cycle reversible? It seems too early to say. With brand heavyweights like Yahoo! entering the fray, presumably with some marketing budget behind it, it seems plausible that uptake could increase. Another encouraging development is that in other sectors, an early finding of analyses of such crowd-wisdom efforts is that they can work well even if just a small but highly motivated and capable corps of people take part. Not every teacher has to post and rate videos for a video-sharing site to be useful. In fact, a McKinsey study of video sharing found that just 3-6 percent of users posted 75 percent of the content, and less than 2 percent posted more than half.[21] A deliberate effort to recruit and cultivate power users could yield dividends. It may also be possible to create incentives for people to share through these mechanisms, perhaps by making sharing the "price of admission" to something of value. For example, to gain access to premium content, users might first be asked to rate some number of items listed on the site, or to share some number of lesson plans. Other incentives are promising as well. While it's easy to poke fun at RateMyProfessors.com's "hotness" ratings, for example, adding this kind of element to such a site will arguably draw more "eyeballs," at least some of which will result in serious reviews. Finally, while the idea of nationwide parent and teacher crowd networks is appealing, smaller, more focused networks (e.g., for teachers implementing certain instructional models) may have more chance of reversing the vicious cycle. One of the problems with the broad-gauged teacher sites, for example, is that the wheat-from-chaff problem is compounded by the fact that for a given teacher, 99 percent of the content is irrelevant: it's the wrong subject, the wrong grade level, the wrong instructional approach, etc.

As for prediction markets, prediction for the sake of forecasting is not the most compelling application. There could be some quantities worth using prediction markets to estimate, such as enrollment trends, but the more interesting question is whether the prediction market structure could be used to generate more robust information about the quality of districts, schools, and teachers than we receive from current measures. Our current measures of quality are, in fact, very weak. Mostly what we have is school and district student achievement data that represent a snapshot at a point in time. Even if longitudinal data continue to advance and "value-added" measures of school and teacher effectiveness become more widespread, these are still fairly narrow measures of quality, even if they are important slices. In addition to narrowness, all of these measures are inherently backward-looking, lagging indicators of performance. They provide little insight about how a school is likely to do next semester, next year, or in three years. Prediction markets, by contrast, encourage analysts to look to the future. Forward-looking indicators are important for parents, who want to know how a school is likely to do in the near term. They could also be valuable to districts and states with have limited resources to apply to school interventions; any indicators that helped them distinguish between schools on track to get better and schools more likely to languish could be valuable.

A prediction market of sorts related to school quality already exists via the market for homes. Economists and families alike have long known that home prices reflect in part the perceived quality of local public schools. Home shoppers with school-aged children are willing to pay more, all else equal, for a house in a better school district. This market, however, is of limited value as a source of useful data about schools and school quality. While home buyers may know that home prices reflect perceived school quality, the information they can obtain from home prices is very blunt and general. They can glean a broad sense that district X is preferred to district Y, but not more detailed and useful information about specific schools, or how their own children are likely to fare.

What if more explicit prediction markets existed for long- and short-term outcome measures for individual schools and districts, such as graduation rates, test scores, and growth in test scores? "Investors" could buy and sell securities related to these measures, which would then pay off at some future time based on the outcome. For example, suppose a state calculated a test-score-growth index for every school each year that ranged between zero and one. Securities could be created for each school that paid $1 times the school's growth index. The market price for the security would represent investors' collective prediction about the school's test score growth for the year.

As noted above, the potential power of such a market would not be the predictions themselves. Instead, the hope would be that the existence of the markets would lead "investors" to seek out good information to help them understand schools' growth potential. That search would in turn lead others to provide information that investors valued. There would be no telling in advance what this information would look like: information-providers might survey parents, or conduct school visits to produce ratings, or analyze previous test score data more finely, or ask experts to rate schools, or develop new ways to rate schools that this paper can't envision. And that, in fact, would be the chief reason to have the market in the first place: to create an engine for better forms of data about school effectiveness.

What would be the value of data like these? Parents choosing schools (or considering whether to stay put) could use them to make better decisions. Teachers could use them to decide where to seek employment. As noted above, district and state leaders could make better decisions about where to place scarce resources for school intervention—both financial resources and "human capital" such as turnaround leaders or teachers. Researchers could use the new measures to identify schools that are likely to be performance outliers, and then start observing those schools' practices earlier rather than later. In general, "leading indicators" would provide a range of decision makers with much better predictions about the future, which would enable them to make better decisions.

The notion is admittedly far-fetched, and replete with potential problems. The largest, as with other wisdom-of-crowds ideas, is how to generate enough volume of trading to make a market function. Volume is essential not only to make trading possible, but also to create incentives for information providers to come forward with their offerings. They need eyeballs on their information in order to make it worthwhile, either financially or otherwise, to gather and post it. Another challenge relates to "insider trading"—the people in the best position to predict a school's outcomes are the staff and parents involved in the school. Yet, allowing them to bet on the school's outcomes raises serious ethical questions. Even if they are not allowed to bet, having them emerge as information providers would also be problematic. One way companies like Google have minimized insider trading problems is by making the financial stakes of the markets fairly low. Successful traders receive some financial rewards, but the biggest payoff is in recognition. But this example highlights the inherent tradeoff: the higher the stakes, the more likely a market is to attract a volume of serious investors and the consequent inflow of information providers. Lower stakes minimize insider trading issues but also make it less likely that the process will have the data-generating effects that are possible with richer markets.

Another way prediction markets could work in education is the creation of futures contracts linked to individual teachers or groups of teachers. Suppose, for example, that individual elementary school teachers received contracts that would pay off according to how well their students did on state tests in future years relative to expectations based on their starting points. So if a teacher's students did better than expected, the teacher's contract would be worth more. In this basic form, the system would just be a form of performance-based pay, without any relevance to the question of data in education. But what if teachers could sell their contracts to investors? A market would form in which the price of each teacher's contract would reflect investors' collective estimate of the teacher's effectiveness. As with the example above, the hope would be that the existence of such a market would spur the creation of information flows to help investors rate teachers and make good trades. This information would be forward-looking, aiming to educate investors about likely future success rates of teachers. That kind of information could be enormously valuable to administrators, teacher developers, and ultimately teachers themselves as they plan for professional growth and make staffing decisions over time.

The idea of using prediction markets to elicit better data and information about schools and teachers needs a lot more development. Perhaps the best way to accomplish that would be through pilot projects, in which philanthropists prime the pump of prediction markets in order to get trading and information flowing. Once they are underway, it would be possible to see how prediction markets like this might work in K-12, what problems emerge, and what kinds of valuable data and information they begin to elicit.

Discussion

In education circles, data-driven instruction is the subject of innumerable papers and professional development workshops. It is the mantra of most every school principal and district superintendent. If anyone in a discussion of K-12 improvement suggests that more and better use of data is vital, everyone will nod and murmur their assent. So why is it that K-12 data systems ignore the kinds of opportunities sketched above?

Of course, the first culprit is lack of funding. Companies like Amazon and Wal-Mart invest huge sums in developing the kinds of systems described in this paper. They do so because they expect even more enormous profits to flow as a result of these investments. In public education, as in all public sector activities, it is clearly more difficult to make that kind of investment, because even socially valuable investments will not necessarily generate the financial return needed to pay for the investment in the first place. It is likely that higher levels of government, such as state and federal agencies, will need to be the source of investment capital for these developments. Schools and all but the largest districts are unlikely to be able to bankroll the kinds of investments needed to build and scale the kinds of systems needed to conduct data-mining and wisdom-of-crowds applications.

But funding is only part of what's needed. A related missing piece is the intense drive that the leading data-using organizations described in this paper have to find new ways to exploit and use data. Netflix, for example, is engaged in an all-out war against Blockbuster for market share, not to mention the rapidly growing sector of video-on-demand providers who can deliver movies even more quickly than they arrive in Netflix's famous red envelopes. Netflix must distinguish itself or be extinguished. As Blockbuster moves to compete and catch up, Netflix has to leapfrog ahead again. And so on.

To capitalize on these same opportunities in K-12 education, that same kind of leap-frogging spirit needs to be infused. Already, there are numerous education ventures afoot that have formed to build the kind of data systems K-12 need to succeed, some of them described in this paper. The fact that many of them, such as Greatschools.net, are nonprofits suggests that is not just the profit motive that can generate the drive to leapfrog. But nonprofit or for-profit, the nation needs a much larger and more robust set of organizations out to win the race to create the next great data application for K-12 education. Private philanthropy—and public investment—can usefully be applied to creating and growing this sector.

Public policy can play a role as well, primarily by promoting the kinds of accountability policies that create a demand on the part of schools, districts, parents, and others for the kinds of services that these data-oriented organizations might offer. Ultimately, these organizations need "customers" who will pay for their services, either directly through fees, or indirectly by providing the eyeballs that advertisers covet and are thus willing to pay for. By ramping up their insistence that schools and districts report out ever more refined data on their performance, that districts take action when schools are truly languishing, that teachers be evaluated more meaningfully and granted tenure only when deemed effective, that families have options among public schools, policymakers can increase the demand from parents, teachers, school leaders, districts, and others for ever-better sources of data to act on these new imperatives.

Endnotes

1 The author wishes to thank Public Impact's Emily Ayscue Hassel for her many insights that informed this paper, especially with regard to the potential of technology to help teachers diagnose and respond to students' specific challenges; Public Impact's Dana Brinson for invaluable research assistance; Fordham's Eric Osberg, Coby Loup, Marci Kanstoroom, and Chester E. Finn Jr., for their guidance and helpful comments; and the authors of other chapters in this volume for their feedback and suggestions.

2 Suroweicki, James. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Doubleday, 2004.

3 Ayres, Ian. *Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart*. New York: Bantam, 2007. Davenport, Thomas H. and Jeanne G. Harris. *Competing on Analytics: The New Science of Winning*. Cambridge: Harvard Business Press, 2007.

4 This and subsequent references to Wal-Mart's data systems: Hays, Constance L. "What Wal-Mart Knows about Customers' Habits." *New York Times,* November 14, 2004. http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html.

5 Loveman, Gary. "Diamonds in the Data Mine." *Harvard Business Review,* May 2003.

6 Iyer, Bala and Thomas H. Davenport. "Reverse Engineering Google's Innovation Machine." *Harvard Business Review*, April 2008.

7 Iyer, Bala and Thomas H. Davenport. "Reverse Engineering Google's Innovation Machine." *Harvard Business Review*, April 2008.

8 Davenport, Tom. "Entrepreneurial Execution: The Future of Strategy." *The Next Big Thing* (a Harvard Business Publishing blog), December 11, 2007. http://discussionleader.hbsp.com/davenport/2007/12/entrepreneurial_execution_the_1.html; See also Farson, Richard and Ralph Keyes. "The Failure-Tolerant Leader." *Harvard Business Review,* August 2002.

9 Data table in *Technology Counts 2008: STEM*, Education Week, 2008. http://www.edweek.org/media/ew/tc/2008/30trends.h27.pdf

10 Christensen, Clayton and Michael Horn. "How Do We Transform Our Schools?" *Education Next*, Summer 2008.

11 Christensen and Horn.

12 Dye, Renée. "The Promise of Prediction Markets: A Roundtable." *The McKinsey Quarterly* No. 2, 2008.

13 For a fascinating overview of open source technology, with an emphasis on how open source projects manager to elicit such wide scale voluntary contributions, see Weber, Steven. *The Success of Open Source*. Cambridge: Harvard University Press, 2004.

14 Iowa Electronic Markets. http://www.biz.uiowa.edu/iem/

15 Berg, Joyce, Robert Forsythe, Forrest Nelson and Thomas Rietz. "Results from a Dozen Years of Election Futures Markets Research," forthcoming in C.A. Plott and V. Smith (Eds.) *Handbook of Experimental Economic Results,* 2001. http://www.biz.uiowa.edu/iem/archive/BFNR_2000.pdf.

16 Dye, Renée. "The Promise of Prediction Markets: A Roundtable." *The McKinsey Quarterly* No. 2, 2008.

17 "Frequently Asked Questions: Writing a Review of Your Child's School." Greatschools.net. http://www.greatschools.net/cgi-bin/static/guidelines.html/nc/

18 About Us." The Savvy Source. http://www.savvysource.com/savvy/jsp/aboutus/aboutUs.jsp Disclaimer: the author's organization has provided content for both GreatSchools.net and The Savvy Source.

19 Yahoo! Teachers. http://teachers.yahoo.com/

20 "'Hotness' and Quality." Insidehighered.com, May 8, 2006. http://www.insidehighered.com/news/2006/05/08/rateprof

21 Bugin, Jacques. "How Companies Can Make the Most of User-Generated Content." *The McKinsey Quarterly,* August 2007.