Data Analytics and the Assessment of Student Writing

Darrell M. West and Joshua Bleiberg

INTRODUCTION

Darrell M. West is vice president and director of Governance Studies and founding director of the Center for Technology Innovation at Brookings. His studies include technology policy, electronic government, and mass media.

Joshua Bleiberg is a research assistant in Governance Studies and Center Coordinator for the Center for Technology Innovation at Brookings. He has a Masters Degree in Education Policy and Management from the Harvard Graduate School of Education.

Experts have long discussed the power of big data to revolutionize educational assessment. Analysts note the ability of data analytics to provide real-time evaluation and improve decision-making on the part of students, teachers, parents and administrators.

Yet confidence about the utility of analytics is undercut by difficulty in gaining access to actual data. Real world demonstrations of how big data would fit into a classroom environment have been few and far between. Much of the information is proprietary and not available to researchers and existing projects have focused on core skills like math and reading. This omission is problematic because we need to determine ways to assess complex skills such as writing.

This paper seeks to fill that void with a practical demonstration of data analytics in regard to student writing. We rely upon information provided by the literacy instruction program In2Books (I2B) at http://In2Books.EPals.com. It seeks to improve writing proficiency through a combination of reading books, corresponding with trained adult pen-pals, and coaching by teachers.[1]

Our goal is to show how to deploy an automated writing assessment system in a classroom that adds value to students and teachers. The data set comes from fourth-grade classroom in a school district on a Native American reservation in Arizona. We analyze student writing for a full year using an online tool measuring readability.[2] This project enables quick calculation of a number of different writing analytics.

In our analysis, we found that data analytics offer helpful means of ranking students and assessing classrooms as a whole. Automated tools can provide dashboards for teachers and real-time feedback and evaluation for students. There remain a number of challenges in big data applications, but there are promising tools available for classroom usage.

Personalized Learning and Data Analytics

Imagine the following classroom scenario. A teacher has divided the class into small groups based on the student's reading proficiency. Some students are reading new books while others are writing journal entries on the texts they have read. A group uses tablets or mobile devices to correspond with pen pals who have read the same books.

Except for a small group working with the teacher, pupils engage in self-directed learning. Each individual works at his or her own pace and gets feedback on progress and proficiency. Teachers can look at a dashboard that lists each student by name and see where he or she is in completing particular reading and writing assignments. Both students and teachers receive immediate feedback on educational progress.

This is not a fantasy, but now is possible in many classrooms around the world. Personalized instruction helps both struggling and strong students alike. Those who need support have the time and teaching assistance necessary. Those students who make rapid progress can take on more challenging assignments that help them develop their abilities. Digital technology allows teachers to become more productive and effective at delivering personalized instruction. Combining educational content with real-time assessment offers tremendous potential to reshape and improve the learning experience.

> *This is not a fantasy, but now is possible in many classrooms around the world. Personalized instruction helps both struggling and strong students alike.*

The Study of an Apache Classroom

School districts in areas of extreme poverty face numerous challenges. They have limited financial resources. They deal with students coming from families with difficult circumstances. In some cases, their homes have experienced divorce, job losses and/or substance abuse problems. This makes it difficult for students to focus on their own instruction.

The district we examined is the Whiteriver Unified School District in Whiteriver, Arizona. It is located on the Fort Apache Indian Reservation in northeastern Arizona. The district educates around 2,000 students each year from preschool to the twelfth grade. Nearly all of its students are eligible for free or reduced price lunch, indicating high poverty.[3] The counties have high levels of alcoholism and unemployment.[4]

Even for the best of the students and teachers, it is a challenging educational environment. Pupils struggle to keep up with their studies and poverty forces some to change schools during the course of the academic year. Research has shown that students in poverty experience more difficulty developing literacy due to funding inequity, student mobility, and limited social capital.[5]

Yet for these students, the development of literacy skills is critical to escaping the cycle of poverty. For the rural poor, it is vital that schools teach needed reading and writing skills as well as offer the possibility of a brighter future. Getting a decent education may be the only route out of difficult life circumstances for these individuals.[6]

The In2Books Program

I2B's creators designed a platform through the ePals educational company to improve literacy skills for students from low income communities. Their goal is to encourage students to develop reading and writing skills, engage the community through real world literacy activities, and delve into challenging high quality texts.

The core of I2B is a pen pal program where students exchange letters with adults in the community. The pen pals engage the students in conversations about books to motivate and encourage students to develop literacy skills. Educators recruit the adult pen pals from local towns including business leaders, government workers, and other community leaders. They receive training on how to write letters to students about the books they are reading. The pen pals engaged the students in questions about the texts, drew the students out in terms of what they learned from each book, and encouraged students to write up their own reactions to the books.

A survey of students participating in the I2B program found a number of positive features. When asked what they did that they liked the best, students named: having a pen pal (57 percent), reading the books (37 percent), writing letters to an adult pen pal (33 percent), getting the books (20 percent), and communicating with others (17 percent).[7]

Another program evaluation of I2B found a positive impact on elementary school student reading scores. Researchers studied SAT-9 Reading Test scores for second through fourth graders in the District of Columbia Public Schools. The study compared students in I2B classrooms with those who participated in another literacy instruction program.

Classrooms were divided into 4 groups: veteran I2B, first-year I2B, total I2B, and non-I2b. Veteran I2B included teachers who taught for I2B for two years or more. First-year I2B teachers taught the program for the first time in the year of study. Total I2B included all Veteran I2B and first-year I2b classroom. Non-I2B classrooms did not use the program. In all there were 2,219 students in I2B classrooms across all grades and 8,634 students in non-I2B classrooms.

Researchers analyzed student performance and found in grades 2, 3, and 4 that veteran I2B classrooms showed the greatest improvement. Students in those courses had statistically significant benefits in performance compared to their non-I2B peers. The differences between I2B and non-I2B classrooms were statistically significant in third and fourth grades.

This study showed that I2B helped disadvantaged students learn reading and writing. The environment was very challenging because about eight out of every ten students in the classrooms were from Title I schools, indicating serious economic deprivation.[8] If technology can be beneficial in these areas, it offers promise for other schools around the country and the world.

Teachers who used In2Books have lauded the program. For example, a teacher in Wyoming commented, "The integration of reading and writing provides for a seamless and authentic use of literacy skills. This is real-life literacy the way my students will use it in their daily lives. The purpose is authentic, not contrived for classroom instruction."[9]

In the case of the Whiteriver Unified School District, teacher Susan Rodriguez explained why she liked to use In2Books in her classroom. She said, "I use In2Books because it is the real-world application of writing for a purpose. So many times, the only writing students do is for a grade…. Where In2Books is different, is that the students are matched up with adults. Adults who have (or have had) careers. They talk about their careers to the kids. They talk about college. Unfortunately, my students don't get those conversations at home."

In describing the most beneficial aspects of In2Books, Rodriguez noted that "it is the seamless integration of Reading, Writing, and Technology (along with Social Studies and Science). Teachers are being asked to teach more and more, with less and less time to prep, let alone collect materials. I have found In2Books to be a perfect way to integrate the Science content into Reading, Writing, and Technology. The same for Social Studies. In2Books has done a wonderful job over the years of adding lesson plans for teachers, complete with graphic organizers, prompts, rubrics, etc."

Teaching Reading and Writing

In this project, we focused on 24 students in the Whiteriver Elementary School fourth grade taught by Susan Rodriguez, STEM curriculum developer for the district. During the 2012-13 academic year, she used the In2Books program to have students write an introductory letter about themselves to their adult pen-pal and then read books from five different genres: fiction (where they read *Dear Mr. Blueberry* by Simon James), social studies (featuring the book *Me on the Map* by Joan Sweeney), biography (*George Washington Carver: The Peanut Wizard* by Laura Driscoll), traditional tales (*Brer Rabbit* by Robert Roosevelt for more advanced students and *The Empty Pot* by Demi for other students), and science (*Tsunamis and Other Natural Disasters* by Mary Pope Osborne and Natalie Pope Boyce). These assignments formed six modules for the academic year and students wrote six letters based on these tasks.

For each module, students followed a number of steps. They would read the book, have Book Club discussions, receive a letter about the book from their pen-pal, write a rough draft of a letter to their pen-pal containing their thoughts about the book, and type their letter on the In2Books website, which then would send it to the teacher and pen-pal.

Students were encouraged to write what they learned from the book, what its moral was, and descriptions of the main character and his or her central traits. The teacher also employed a "question matrix" developed by Chuck Weiderhold that contains a number of questions concerning what, where, when, which, who, why, and how. That allows students to think about the past, present, and future, and develop creative reactions to each volume.

In literacy studies, researchers examine the quantity and quality of writing expression and sophistication in the use of words, sentences and paragraphs. We looked at student writing assignments and used automated tools to analyze the number of words and sentences that were written. For higher-level analysis, we computed the average number

of words per sentence and sentences per paragraph.[10] By dividing the number of words per sentence by the sentences per paragraph, we computed the Flesch-Kincaid Grade Level measure for each student and each writing assignment.[11] These basic calculations allow us to rank individual students during the course of the academic year and assess class performance as a whole.

Rating the Students

The 24 students completed 111 of the 144 writing assignments. Four students moved away from the district during the course of the school year and there were others who did not complete all of their writing assignments. Twenty-four pupils completed the first writing assignment introducing themselves to their pen-pal, while 21 students finished the fiction assignment, 17 did so on social studies, 18 wrote about their biography book, 13 finished traditional tales, and 18 wrote the science assignment.

Overall, the students wrote 10,732 words and 1,037 sentences across their six writing assignments. The average number of words per sentence was 11.9, while the average number of syllables per word was 1.32. According to the Flesch Kincaid readability assessment, the fourth grades wrote at a 4.6 grade level, where is exactly where their age cohort lies.

But for each of these metrics, there was considerable variation around the mean. Table 1 shows writing metrics for the 24 students in the fourth grade class. It includes the total number of words for the academic year, total number of sentences, average syllables per word, average words per sentence, and the Flesch Kincaid Grade Level. Some individuals performed above the fourth grade level, while others were well below that level.

TABLE 1: INDIVIDUAL STUDENT WRITING METRICS, 2012-13

Student	Total Words	Total Sentences	Ave Syllables Per Word	Ave Words Per Sentence	Flesch Kincaid Grade Level
1	596	39	1.36	17.9	7.4
2	532	45	1.30	15.2	5.6
3	202	22	1.35	9.3	3.9
4	289	37	1.29	7.7	2.7
5	201	17	1.28	14.1	5.0
6	691	65	1.36	10.7	4.6
7	550	66	1.34	9.3	3.8
8	255	26	1.36	13.6	5.7
9	132	8	1.43	18.7	8.6
10	183	9	1.18	19.9	6.1
11	526	53	1.29	9.6	3.4
12	320	39	1.36	8.0	3.6
13	644	63	1.34	10.6	4.4
14	752	74	1.30	10.7	3.9
15	238	25	1.25	11.0	3.5
16	578	24	1.28	25.9	9.6
17	438	32	1.26	13.7	4.6
18	483	44	1.27	11.5	3.9
19	671	70	1.34	9.8	4.0
20	396	50	1.46	7.9	4.8
21	529	69	1.33	7.9	3.2
22	648	71	1.29	10.2	3.6
23	304	36	1.25	8.7	2.6
24	574	53	1.30	14.2	5.2
Class Summary	*10,732*	*1,037*	*1.32*	*11.9*	*4.6*

Writing Metrics by Type of Educational Module

Table 2 shows how the class progressed through each of the six modules. In general, students performed the best on science (5.2 grade) and fiction (5.7 grade), and worst on the introduction (3.6 grade) and traditional tales (3.8 grade). There were medium-level performances on social studies (5.1 grade) and biography (4.3 grade).

TABLE 2: WRITING METRICS BY TYPE OF EDUCATIONAL MODULE, 2012-13

Module	Total Words	Total Sentences	Ave Syllables Per Word	Ave Words Per Sentence	Flesch Kincaid Grade Level
1 (Intro)	1,644	183	1.30	10.0	3.6
2 (Fiction)	2,575	212	1.29	15.5	5.7
3 (Social Studies)	1,625	147	1.32	13.1	5.1
4 (Biography)	1,776	183	1.34	10.4	4.3
5 (Traditional Tales)	1,266	148	1.31	10.2	3.8
6 (Science)	1,846	164	1.37	11.8	5.2

Gender Differences

In this class, there were 13 females and 11 males. The girls completed 61 writing assignments compared to 50 for the boys. There has been considerable discussion about gender differences in writing, and which group performs at a higher level.

Table 3 shows the writing metrics by gender. In general, males performed a little better than females. They wrote at a slightly higher grade level and produced more complex sentence structures. But befitting the fact there were more girls in the class, the females wrote more in terms of total words (6,734) compared to males (3,998).

TABLE 3: WRITING METRICS BY GENDER, 2012-13

Gender	Total Words	Total Sentences	Ave Syllables Per Word	Ave Words Per Sentence	Flesch Kincaid Grade Level
Female	6,734	673	1.32	11.2	4.3
Male	3,998	364	1.33	12.7	5.0

Predicting Performance

The most basic goal of a fourth grade education is to move students from fourth to fifth grade competency. With the U.S. system predicated on yearly grade promotion, it is vital that individuals keep up with their age cohort and master the skills required of their grade level. Automated assessment can help predict which qualities are most associated with grade level attainment.

Table 4 shows the correlations between grade level and various writing metrics. The writing qualities most highly correlated with grade level is the average number of words per sentence. Those traits correlated at the .92 level, indicating they were highly connected. Students who write complex sentences are able to perform at a higher grade level than those who lack that skill. In addition, there were high correlation between the percent of readable sentences (.45), the total number of sentences (-.40), and the number of sentences needing to be rewritten (-.38).

TABLE 4: CORRELATIONS OF GRADE LEVEL AND WRITING METRICS, 2012-13

	Flesch Kincaid Grade Level Attainment
Average Words Per Sentence	.92***
Percent of Readable Sentences	.45***
Total Number of Sentences	-.40***
Number of Sentences Needed To Be Rewritten	-.38***
Average Syllables Per Word	.22*
Total Number of Characters	.20*
Average Characters Per Word	.20*
Total Number of Words	.18
Note: *** indicates statistical significance at the .001 level of probability, while * indicates significance at the .05 level of probability.	

The Crucial Role of the Teacher

Some educators have anxiety incorporating new assessment technologies into their classrooms. They fear big data will diminish their autonomy, be impersonal, or inhibit their ability to help students learn.

But it doesn't have to operate that way. Big data analytics and dashboards should have the opposite effect. They should empower teachers to have more freedom and to improve instruction by providing real-time data and more nuanced performance metrics. Big data should supplement the role of the teacher, not replace it.

> *Some educators have anxiety incorporating new assessment technologies into their classrooms. They fear big data will diminish their autonomy, be impersonal, or inhibit their ability to help students learn.*

Teachers face a number of constraints in their effort to evaluate students. Educators have so many responsibilities in addition to teaching that dedicating hours to formative assessment preparation is difficult. Grading is a time consuming and tedious process. Reading assignments for the purpose of developing formative assessments is all the more difficult.

A further difficulty is developing reliable and valid assessments. Evaluation metrics must measure what they claim to assess. Writing questions that are both reliable and valid is extremely difficult and time consuming.

Data analytics address many of these concerns. Computer graded assessments save teachers' time. Teachers do not have to read and assess student work because the learning system does it for them. This allows them to focus on higher level instructional activities and coaching those students who need extra help.

Performance Dashboards

Automated assessments help teachers develop dashboards and target instructional interventions on particular students. The explosion of assessment data has not corresponded with an increase in useful teaching tools. The sheer amount of data renders it useless for any purpose other than testing and accountability. In some districts standardized testing reaches teachers months or even a full year after the students take a test.

One way to solve this issue is through teacher dashboards. Digital technology enables the aggregation and presentation of data in a simplified form that is instructive and visually compelling. Data arrives in real-time and is available through a mobile app or a web interface. Teacher dashboards represent a way to convert mountains of data into a format that is useful for educators and administrators.

Using data from our automated writing assessment tool, it is possible to set up a dashboard for teachers to monitor student performance. The dashboard can color students green (making above average progress), orange (making normal progress), or red (making below average progress) based on various metrics. These types of displays can inform a teacher who in the course is struggling or needs extra help. Seeing lots of red on the tracking screen indicates a serious class-wide problem.

Ideally, students will start the fourth grade with a Flesch Kincaid Grade Level of 4 and by the end of the year have reached at least a 5 grade level. The dashboard can track student achievement to see if each person is on track to reach the desired attainment level.

In addition, there are instructional benefits as well. Observing differences in student performance during the academic year provides evidence on how students are responding to actual teacher interventions. Providing real-time feedback allows instructors to see which teaching techniques work and helps them address instructional problems right away.

Low Stakes, High Yield Assessment

Lee Shulman of the Carnegie Foundation for the Advancement of Teaching has argued that American education focuses too much on "high stakes, low yield" evaluation.[12] Schools measure performance through once a year standardized tests and rate schools by their overall results. An increasing number of districts have tied teacher pay to student test performance. This creates incentives for teachers to "teach to the test" and even to cheat.

An alternative is "low stakes, high yield" assessment that features "running records" of student performance. Rather than once a year testing, pupils are evaluated on an ongoing basis and given feedback in real-time. Several researchers have argued that timely and relevant feedback improves students' ability to learn.[13] Digital assessment is impactful when it delivers immediate feedback. More importantly, it does so in a manner

that is more nuanced than standardized tests. This helps guide students and teachers to better learning outcomes.

Some schools have unfortunately narrowed the curriculum in order to prepare students for annual testing. Sociologist Donald Campbell has noted that "the more any quantitative social indicator (or even some qualitative indicator) is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."[14] In an education context as school and teachers were held responsible for the test scores of their students, some educators spent a disproportionate amount of time on instruction tailored to specific test items.

This is not always a bad thing. Accountability advocates intended for a certain degree of refocusing on core reading and math skills. Teaching to the test and high quality teaching do overlap with each other. The point at which teaching to the test becomes undesirable is when it drives out more conceptually oriented thinking or crucial problem-solving skills.

Data analytics can serve as a high quality formative assessment and counteract teaching to the test. Formative assessments help guide instruction and allow instructors to focus on techniques that improve learning. It is not just a question of what students know at the end of a lesson, but how they acquire information and reach the desired knowledge. A prompt guiding a student to write a more complex sentence helps that individual develop new skills and learn more efficiently.

One of the virtues of the In2Books program is that it allows schools to use data to inform teaching practices and educational performance. It presents a novel combination of social collaboration, teacher coaching, adult pen-pals, and data-driven innovation. Their model integrates digital content with regular assessment to promote student learning. With over 14 million participants around the world, they have a demonstrated ability to engage students and improve performance in reading and writing.

The Challenges of Big Data

Automated scoring represents one way to make digital assessment scalable to millions of students. Massive open, online courses (MOOCs) employ automated tools as do a growing number of private educational companies. This takes the form of style checks, spelling checkers, grammar tools, vocabulary tests, and pop-up quizzes, among other

things. The best of the tools identify what is wrong but also help students figure out how to fix their mistakes. When combined with user profiles, these types of analytical tools enable researchers to parse student learning by year, gender, subject area, styles of learning, and many other dimensions.

Big data analytics don't answer every educational challenge. So far, they do best when applied to tasks such as mathematics or multiple choice tests where machines can easily distinguish correct from incorrect answers. But it has been more challenging to assess higher-level functions such as problem-solving or critical thinking. Those tasks require context and nuances that remain difficult to assess through automated tools.

Increasingly, though, instruction is combining digital content delivery with embedded assessment. That helps students see how much progress they have made and enables teachers to determine where they should devote their efforts. It is this progress towards personalized learning that represents the most promising development in computerized assessment.

Students learn in very different ways so we need metrics to evaluate those differences. One of the virtues of online systems is that they encourage students to engage literature at a personal level. The designers of educational programs encourage students to read books, interact with friends, and find their own voice in the process. We should develop curricula that encourage higher order skills. Appropriately designed formative assessments can fulfill the promise of accountability reforms without narrowing instruction or spoiling the success of high quality educational programs.

Endnotes

Note: We wish to thank Susan Rodriguez of the Whiteriver Unified School District and her fourth-grade students for making this research possible. Neither she nor her school district is responsible for the interpretations presented here.

1. For a description of how the program works, see William Teale, Katie Lyons, Linda Gambrell, Nina Zolt, Rebecca Olien, and Donald Leu, "An Online Learning Community as Support for At-Risk Students' Literacy Growth", unpublished paper, 2011.
2. See writing assessment tool at http://www.online-utility.org/english/readability_test_and_improve.jsp.
3. Author calculation from NCES at http://nces.ed.gov/ccd/schoolsearch/school_list.asp?Search=1&DistrictID=0409160.
4. Unemployment numbers for CDP Whiteriver Arizona.
5. James Flood and Patricia L. Anders. *Literacy Development of Students in Urban Schools: Research and Policy*. Newark, Delaware: International Reading Association, 2005.
6. William Teale, et al. "Getting children I2B: Engagement in authentic reading, writing, and thinking." *Phi Delta Kappan*, Volume 88 (2007): 498-502.
7. Linda Gambrell, Elizabeth Hughes, Leah Calvert, Jacquelynn Malloy and Brent Igo, "Authentic Reading, Writing, and Discussion", *Elementary School Journal*, Volume 112, number 2, 2011, p. 245.
8. William Teale and Linda Gambrell. "Raising urban students' literacy achievement by engaging in authentic, challenging work." *The Reading Teacher*, Volume 60 (2007): 728-739.
9. http://in2books.epals.com/content.aspx?caid=I2BCorp&divid=Wyoming_Teacher
10. For our writing assessment, we used an online readability tool available at: http://www.online-utility.org/english/readability_test_and_improve.jsp. It computed the various metrics deployed in this study.
11. There are a number of different ways to measure readability. We looked at several different approaches and found they generally correlated highly with one another. The correlation of the Flesch Kincaid Grade Level with the Gunning Fog index was .95 and it was .96 with the automated readability index, .72 with the Simple Measure of Gobbledygook, and -.86 with the Flesch Reading Ease assessment. It had a slightly lower correlation with the Coleman Liau index (.53). The latter measures readability based on characters per word as opposed to syllables per word the way most of the others do.

12.   Lee Shulman, "Counting and Recounting: Assessment and the Quest for Accountability", Carnegie Foundation for the Advancement of Teaching, *Change*, Volume 39, number 1, January/February, 2007.

13.   Lev Vygotskij, *Thought and Language*, MIT Press, 2012.

14.   Donald Campbell, *Assessing the Impact of Planned Social Change*, Hanover, New Hampshire: Dartmouth College Public Affairs Center, December, 1976.

**EMAIL YOUR COMMENTS TO
GSCOMMENTS@BROOKINGS.EDU**