Assessing Chinese Students' Reading Behavior Through the Lens of Big Data

Yun Zhang, Lingjian Wang, Jingjie Wang, Yang Qu

Abstract

In the new era of Big Data, researchers, educators, and policy makers across the world show keen interests in how to use Big Data to assess and improve students' learning. Here we presented an example to illustrate the use of Big Data in assessing Chinese students' reading behavior. 137,916 Chinese students from 86 elementary schools in Shanghai used an online reading platform, yielding to over 1,520,000 records on their reading behavior. First, detailed patterns on Chinese students' reading behavior, such as age differences in reading, time of reading, and types of books were found. Second, students' reading behavior uniquely predicted better achievement in Chinese 6 months later over and above their prior achievement. Third, schools using the reading platform (vs. not using the platform) showed better performance in Chinese 6 months later. Taken together, this study illustrates how Big Data can provide an unprecedented opportunity to examine students' learning that are unable to be addressed via traditional methods. Implications for research, practice, and policy are discussed.

Assessing Chinese Students' Reading Behavior Through the Lens of Big Data

In the new era of Big Data, researchers, educators, and policy makers across the world show keen interests in how to use Big Data to assess and improve students' learning (Baker & Siemens; 2014; Eynon, 2013; Macfadyen, Dawson, Pardo, & Gaševic, 2014; Siemens & Long, 2011). One essential basis for academic and career success is reading. It is particularly important in students' learning, such that reading provides a powerful tool, through which students acquire new knowledge and concepts (Pretorius, 2002). In addition to reading in class, a key reading behavior that receive much empirical attention is independent reading, which refers to reading that students choose to do on their own. Independent reading is also called voluntary reading (Krashen, 1993; Morrow, 1991), leisure reading (Greaney, 1980), spare time reading (Searls, 1985), and reading outside of school (Anderson, Wilson, & Fielding, 1988). However, before the rise of Big Data, the assessment of independent reading behavior heavily relies on traditional methods such as observations and self-reports.

Prior research suggests that independent reading among school-aged children is relatively low. For example, only a small proportion of preschool and primary children chose to read books during free-choice time at school (Morrow & Weinstein, 1986). Moreover, fifth graders spent only 5.4 percent of their out-of-school free time engaged in reading, and 23 percent of them chose not to read at all (Greaney, 1980). However, little is known about independent reading among students in other countries. This represents an important lacuna, because reading is essential for students' success and may play a long-term impact on country differences in academic achievement. Given that recent large-scale cross-country comparison (e.g., PISA) reveals substantial variability in students' academic achievement in different countries, it is important to examine the role of independent reading in achievement outside the United States.

To address this issue, the current research investigated students' independent reading in China, with attention to the implications for students' academic achievement.

Empirical studies consistently suggest that independent reading benefits learners in different contexts. In addition to the gains in reading proficiency and reading habits (Camiciottoli, 2001; Nash & Yuan, 1992), independent reading also leads to improvement in listening proficiency (Elley & Mangubhai, 1983), writing ability (Mason & Krashen, 1997; Tsang, 1996), reading speed (Walker, 1997), and spelling (Krashen, 1989). It is not surprising that a positive link between independent reading and academic achievement was found in prior research. For example, more time that student spent on out-of-school reading, even if it was a small amount, is positively associated with reading achievement (Anderson, Fielding, & Wilson 1988). The more students read outside of school, the higher they scored on reading achievement tests. Similarly, Greaney and Hegarty (1987) found a significant relationship between the amount of time spent on independent reading and reading achievement, verbal ability, and attitude toward reading.

Although independent reading mostly occurs outside the school, research suggests that school's encouragement of reading and positive reading environment can also facilitate students' independent reading, which leads to better reading ability (Daniels & Steres, 2011; Lu, 2012). Schools in China usually hold a variety of reading activities and tests to stimulate students' reading motivation and interests. To effectively track students' independent reading, a growing number of schools in China use online reading platform, which can provide books retrieval and help students record their reading behavior. Moreover, students can make comments on books, take notes, answer questions, and communicate with other students via reading platform. Recent research suggested that among third and fourth Chinese graders, students who used online

reading platform scored significantly higher in reading and writing tests than those who did not use (Su, 2015).

The current study aimed to utilize Big Data to provide a deeper understanding on students' reading behavior. Specifically, this study was guided by three goals. First, using an online reading platform to track students' reading behavior, we characterized detailed patterns of Chinese students' reading behavior, such as such as age differences in reading, time of reading, and types of books. Second, to examine the validity of data on reading behavior, we tested whether students' reading predicts their academic performance over and above their prior academic performance as well as other variables (i.e., subject interests, gender, and grade). Third, we further investigate whether schools using a reading platform show significantly better academic performance than those not using a reading platform after taking into account their prior academic performance.

Method

137,916 Chinese students from 86 elementary schools in Shanghai participated in the present study. These schools used an online reading platform called Psyread, which has information on 360,000 Chinese books and allows students to record their reading behavior (e.g., to find and record the books that they read), take notes, make comments on books, and answer questions). Data used in the current research was collected from 2016 to 2017, yielding to over 1,520,000 records on their reading behavior. These records were analyzed to characterize patterns of students' reading behavior.

To examine the validity of data on reading behavior, we further examined whether reading records predict students' academic achievement using a subset of participants. This sample included 3487 Chinese first and second graders (mean age = 7.10 years; 75% girls) from

13 schools. These participants took academic assessments in Chinese, math, and English in January 2017 (Time 1) and June 2017 (Time 2). Their interests in Chinese, math, and English were also collected at Time 1 and 2 as covariates.

Results

We conducted three sets of analyses. First, we used reading records of all 137,916 Chinese students to characterize detailed patterns on students' reading behavior. Second, we used a longitudinal design to examine if these reading records are predictive of students' academic achievement, focusing on those who took academic assessment across 2 time points, with 6 months apart. Third, we investigated the long-term benefits of assessing students' reading behavior, by comparing academic achievement in schools using the reading platform with those not using the platform over 6 months.

What are the patterns of Chinese students' reading behavior?

We first examined students' reading behavior. As shown in Figure 1, Chinese students read more books per year over the course of elementary school (i.e., from grade 1 to 5). For example, second graders tended to read more books than first graders, $p < .001$. Also, girls read more books compared to boys, $p < .001$. Bivariate correlations between number of books and academic performance at each time point were presented in Table 1.

Does reading behavior predict students' later academic achievement?

Next, we examined the role of students' reading in their academic achievement. To this end, we conducted regression analyses with students' academic performance at Time 1, number of books that they read from Time 1 to Time 2, interests in each subject, gender, and grade predicting students' academic achievement in each subject at Time 2. Specifically, we focused on whether reading plays a unique role in students' academic achievement at Time 2 over and

above their performance at Time 1 and interests in this subject. Consistent with previous research, our results indicated that reading more books was significantly associated with better academic achievement in Chinese at Time 2 (Table 2). This effect is specific to students' performance in Chinese, as reading more books was not associated with better academic achievement in math and English at Time 2, controlling for their performance at Time 1. Other reading behavior, such as reading pages, comments and the questions answered, yielded the similar results. These findings are in line with past literature on the benefits of reading and supports the validity of using platform to track students' reading behavior.

Does using the reading platform improve students' academic achievement in Chinese?

Given the predictive role of reading behavior in students' academic achievement in Chinese, the next goal of this study was to investigate whether schools using a reading platform show significantly better academic performance in Chinese than those not using a reading platform after taking into account their prior academic performance. To this end, we recruited additional 6572 first and second grader from 32 schools. These schools did not use the online reading platform, but took the same academic exam at Time 1 and 2. A mixed model Multivariate Analysis of Variance (MANOVA) with students' performance in Chinese at Time 1 and 2 as a repeated measured (i.e., time of assessment) was conducted to examine the effects of using the online reading platform on students' academic performance. There was a significant platform X time of assessment interaction, $F(1, 10057) = 31.87$, $p < .001$. Univariate tests indicated that schools that used or did not use the online reading platform did not differ in performance on the Chinese exam at Time 1, $F(1, 10057) = .56$, $p = .455$. However, schools that used the online reading platform showed better performance in Chinese at Time 2 compared with

schools that did no used the platform, $F(1, 10057) = 38.33$, $p < .001$ (Figure 2). This finding suggests that using an online reading platform may enhance students' performance in Chinese.

Discussion

Big Data provides an unprecedented opportunity that enable researchers and educators to have a more comprehensive understanding on learning. In line with the call to utilizing Big Data in students' learning and education, the current study analyzed data in an online reading platform to assess Chinese students' reading behavior. Findings reveal hidden patterns of students' reading behavior. For example, the number of books students read per year increases as grades increases. This highlights the importance of taking into account students' grade in developing and implementing policy on students' reading.

Moreover, we followed a subset of participants over 6 months and assessed their academic performance in Chinese, math, and English over 6-month period. Students who read more books showed better performance in Chinese over and above their prior performance. Other reading behavior, such as reading pages, comments and the questions answered, was also associated with better academic achievement in Chinese. Interestingly, the benefits of reading on students' academic achievement are specific to Chinese, but not math and English. This finding builds upon a significant body of literature highlighting the importance of independent reading in students' academic achievement. In addition, schools using the online reading platform showed significantly higher scores in Chinese six months later than those not using the platform, suggesting that using the online reading platform enhances students' academic performance in Chinese. Our findings suggest that school-based online reading platform is a valid tool to track students' reading behavior and highlight the positive role of independent reading in students' academic achievement.

Taken together, this study illustrates how Big Data can be a powerful tool to examine students' learning that are unable to be addressed via traditional methods. Effectively utilizing Big Data has a tremendous potential to make contributions to academic research, school practices, and policy making.

Table 1. Bivariate correlations between number of books that students have read, gender, grade, and academic performance at Time 1 and 2

	1	2	3	4	5	6	7	8	9
1. Number of books	--								
2. Gender	.08***	--							
3. Grade	.05**	.17***	--						
4. Academic performance in Chinese at Time1	.16***	.15***	.37***	--					
5. Academic performance in Chinese at Time2	.15***	.09***	-.06***	.59***	--				
6. Academic performance in Math at Time1	.13***	.12***	.20***	.65***	.57***	--			
7. Academic performance in Math at Time2	.13***	.13***	.50***	.66***	.60***	.67***	--		
8. Academic performance in English at Time1	.13***	.10***	.10***	.57***	.57***	.53***	.48***	--	
9. Academic performance in English at Time2	.10***	.19***	.89***	.57***	.26***	.47***	.74***	.62***	--

Note. For youth's gender, -1 = male and 1 = female. ** p <.01, *** p <.001.

Table 2. Results of the regression analyses predicting students' academic achievement in Chinese at Time 2

Predictor	*B*	*SE* (*B*)	*β*	*t*
Academic performance in Chinese at Time1	.594	.012	.692	48.32***
Gender	.407	.366	.015	1.11
Grade	-6.911	.353	-.276	-19.56***
Interests in Chinese at Time1	.191	.029	.086	6.51***
Number of books	.024	.007	.045	3.42**

Note. For youth's gender, -1 = male and 1 = female. ** $p < .01$, *** $p < .001$.

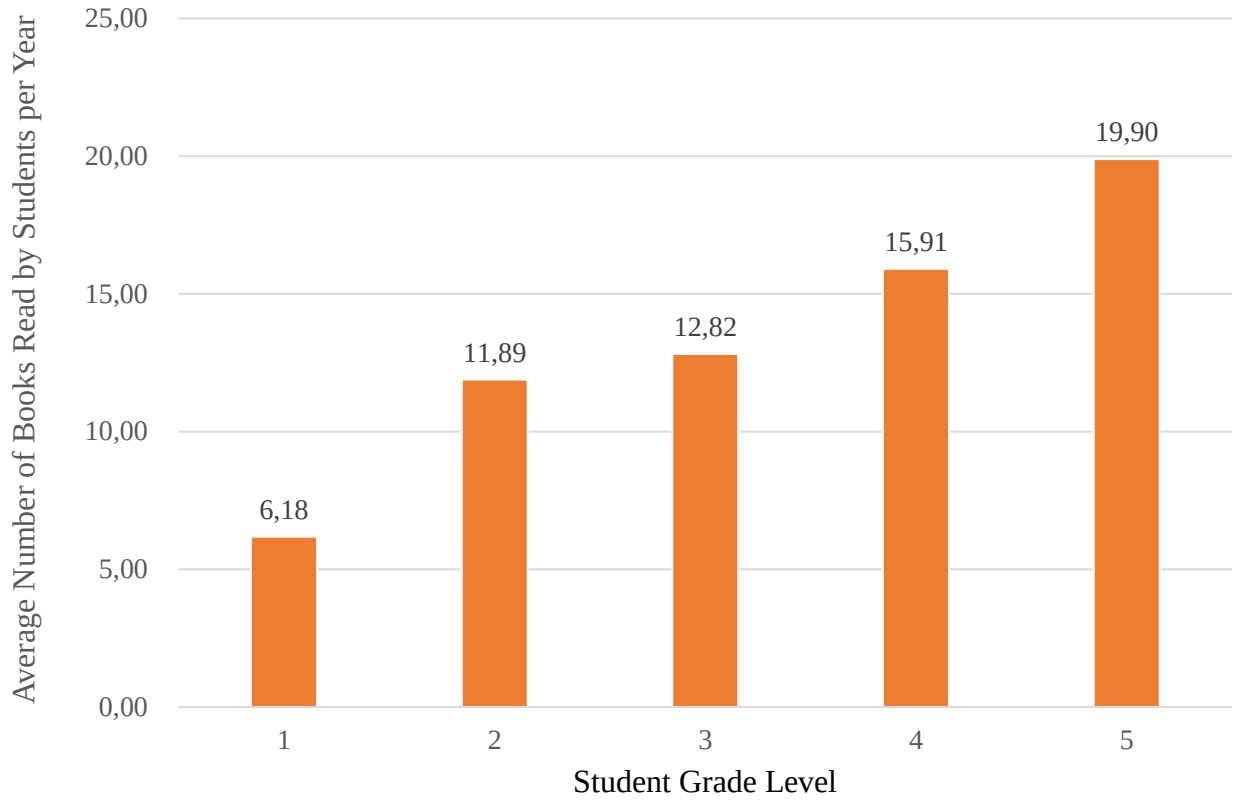

Figure 1. Grade differences in number of books read by students per year.

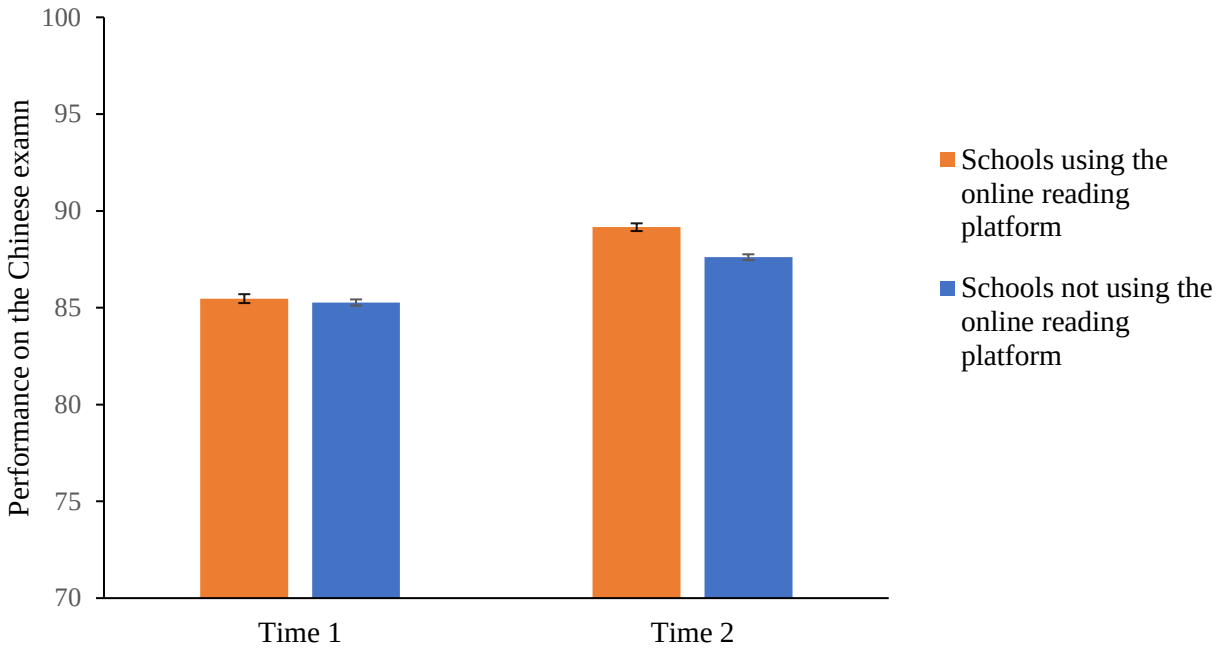

Figure 2. Despite no difference in prior performance in Chinese exam at Time 1, schools using the online reading platform showed significantly higher scores in the Chinese exam at Time 2 (six months later) than those not using the platform.

Note. Error bars represent standard errors of the means.

Reference

Anderson, R. C., Wilson, P., & Fielding. L. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly,* 23, 285-303.

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *the Cambridge handbook of the learning sciences (2nd ed.).* Cambridge University Press, New York, NY.

Camiciottoli, B. C. (2001). Extensive reading in English: Habits and attitudes of a group of Italian university EFL students. *Journal of Research in Reading, 24*(2), 135-153.

Daniels, E., & Steres, M. (2011). Examining the effects of a school-wide reading culture on the engagement of middle school students. *Research in Middle Level Education, 35*, 1-13.

Elley, W. B., & Mangubhai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly*, *6*(1), 53-67.

Eynon, R. (2013). The rise of Big Data: what does it mean for education, technology, and media research? *Learning, Media and Technology, 38*(3), 237-240.

Greaney, V. (1980). Factors related to amount and type of leisure reading. *Reading Research Quarterly, 15*, 337-357.

Krashen, S. D. (1993). *The power of reading: Insights from the research.* Englewood, Colo.: Libraries Unlimited.

Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the Input Hypothesis. *Modern Language Journal, 73*, 450-464.

Lu, J. (2012). The influence of engagement in reading and learning strategy on reading performance: Evidence-based research with Shanghai PISA 2009 data. *Education Development Research, 18*, 17-24.

Pretorius, E. J. (2002). Reading ability and academic performance in South Africa: Are we fiddling while Rome is burning? *UNISA Press, 1*(33), 169-196.

Macfadyen, L. P., Dawson, S., Pardo, A., & Gaševic, D. (2014). Embracing Big Data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment, 9*, 17-28.

Morrow, L. M. (1991). Promoting voluntary reading. In *Handbook of research on teaching the English language arts.* Edited by J. Flood, J. Jensen, D. Lapp, and J. Squire. New York: Macmillan.

Morrow, L. M., & Weinstein, C. S. (1986). Encouraging voluntary reading: The impact of a literature program on children's use of library centers. *Reading Research Quarterly, 21*(3), 330-346.

Nash, T., & Yuan, Y. (1992). Extensive reading for learning and enjoyment. *TESOL Journal, 2*(2), 27-31.

Searls, D. T., Mead, N. A. & Ward, B. (1985). The relationship of students' reading skills to TV watching, leisure time reading, and homework. *Journal of Reading, 29*, 158-162.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*, 30-40.

Su, M. (2015). The influence of digital reading platform on the extracurricular reading of primary school students. *Software Guide: Education Technology, 3*, 91-92.

Tsang, W. K. (1996). Comparing the effects of reading and writing on writing performance. *Applied Linguistics, 17*(2), 627-642.

Walker, C. (1997). A self-access extensive reading project using graded readers (with particular reference to students of English for academic purposes). *Reading in a Foreign Language, 11*(1), 121-149.