Review of *Continued Progress: Promising Evidence on Personalized Learning*

Reviewed By

William R. Penuel
Raymond Johnson
University of Colorado Boulder

January 2016

Summary of Review

A recent evaluation report from RAND focused on school-wide initiatives funded by the Bill & Melinda Gates Foundation to promote teaching approaches touted as personalized learning. These reforms generally rely on digital technology and encompass a range of strategies, such as developing learner profiles with individualized goals, and using data to provide personalized learning paths in which students have choice, get individualized support, and engage in learning outside school. The research, which includes many high-quality elements, suggests that some of the studied approaches are associated with higher scores on a common assessment (the MAP). Broad conclusions about the efficacy of technology-based personalized learning, however, are not warranted by the research. Limitations include a sample of treatment schools that is unrepresentative of the general population of schools, the lack of a threshold in the study for what qualified as implementing "personalized learning" in the treatment schools, and the reality that disruptive strategies such as competency-based progression, which require the largest departures from current practice, were rarely implemented in the studied schools.

Kevin Welner
Project Director

William Mathis
Managing Director

Alex Molnar
Publishing Director

National Education Policy Center

REVIEW OF *CONTINUED PROGRESS*: *PROMISING EVIDENCE ON PERSONALIZED LEARNING*

William R. Penuel and Raymond Johnson, University of Colorado Boulder

I. Introduction

The report reviewed here focuses on three school-wide initiatives funded by the Bill & Melinda Gates Foundation to promote *personalized learning*:[1] Next Generation Learning Challenges (NGLC), Charter School Growth Fund's Next Generation School Investments, and the Gates Foundation's Personalized Learning Pilots. In this report, RAND researchers organized personalized learning according to five different strategies:

1. Learner profiles with individualized goals using data from multiple sources that students and teachers both access;

2. Personalized learning paths, in which students have choice, get individualized support, and engage in learning outside school;

3. Competency-based progression;

4. Flexible use of time, space, and technology; and

5. Developing academic and non-academic career and college readiness skills.

A range of stakeholders embrace the concept of personalized learning as a way to move beyond "one-size-fits-all" approaches to education. It appeals to those concerned that standardization of curriculum and assessments limits opportunities for student choice in what, where, how, and when students can learn. It also appeals to advocates seeking a more robust integration of technology into instruction, including as part of blended and online learning experiences.[2] There are different definitions of personalized learning, with some advocates foregrounding digital technology, and others foregrounding other aspects. Here, we refer to the RAND authors' use of the term.

The RAND evaluation report aims to add to the evidence base by examining the effects of schoolwide efforts to promote personalized learning and the links between implementation of particular strategies and outcomes.

II. Findings and Conclusions of the Report

The report organizes its findings into four parts.

The first part, focused on student achievement results, found that two years of personalized learning had positive effects on both mathematics and reading. When compared to a matched "virtual comparison group" of students, the performance of individual students in 62 personalized learning schools was 0.27 standard deviations higher in mathematics and 0.19 standard deviations higher in reading. Effect sizes varied widely across schools, and not all schools showed increased achievement. When student results were aggregated by school, effect sizes per school ranged from -0.56 to 1.00 in mathematics and -1.27 to 0.90 in reading. Most schools in the treatment condition were charter schools; district schools did not show a significant positive effect, though the authors cautioned that the number of district schools was small.

The second part describes personalized learning implementation. The overall finding was that implementation of personalized learning "varied considerably" (p. 14), with practices similar to traditional educational approaches being the most common. For example, all schools included time for individual student support and offered variety in types of instruction (large-group, small-group, independent).

The third part related implementation to outcomes. The authors found no associations between single elements of personalized learning and schools with the greatest achievement effects. In combination, however, some elements were associated with higher outcomes. At least two of three elements were present in the 5 schools with effect sizes greater than 0.20 standard deviations (the highest achieving schools out of 32 schools used in this part of the study): (1) students were flexibly grouped and regrouped based on data, (2) student learning spaces supported rather than hindered personalized learning, and/or (3) students participated in data-based decision making about their personal learning goals. Due to the small number of schools in the analysis and potential errors in the identification of these elements in schools, the authors of the report caution against forming conclusions about necessary elements of personalized learning based on these findings.

The report concludes that its findings are "largely positive and promising" but issues some noteworthy cautions

The final part of the findings compares teacher and student survey results from personalized learning schools to results from a reference group made from a national sample. A separate organization conducted the national survey. Results here were mixed, and many differences were not statistically significant. Expected results included students in personalized learning schools feeling abler to make choices about their learning. Surprising results included more positive responses from students in the control group, who reported greater enjoyment and comfort in school, and felt their out-of-school work was more useful and connected to their in-school learning. Students in both groups responded similarly regarding their understanding of goals, tracking progress towards mastery, and how teachers helped them plan for the future.

The report concludes that its findings are "largely positive and promising" (p. 34) but issues

some noteworthy cautions. For one, strategies such as competency-based progressions were less common and more challenging to implement, compared to common strategies found in most schools. Such common strategies include using student data to support instructional decisions, offering time for individual academic support, and using technology for personalization. A second caution the authors provide is that the study was unable to separate school effects from personalized learning effects. As many of the schools in the study were awarded competitive grants to implement personalized learning, their students might have been more successful than a comparison group regardless of the use of personalized learning strategies, because they were schools that were already more effective than other schools.

III. The Report's Rationale for Its Findings and Conclusions

The study employed a matched comparison group design to compare outcomes of students in schools that implemented several features of personalized learning. Researchers analyzed student achievement scores on an online adaptive test, the Northwest Evaluation Association's Measures of Academic Progress (NWEA MAP), of roughly 11,000 students from 62 schools from 2013-14 and 2014-15. They selected a "virtual comparison group" from NWEA's database of students to match schools and students on a set of baseline characteristics. For each student in the treatment group, up to 51 comparison students were matched by school locale (e.g., urban, rural), gender, school participation in the national free and reduced-price lunch program, the days elapsed between the pretest and posttest, and achievement levels. Students with equal pretest scores were preferred, but a difference up to five points on the MAP equal-interval scale was allowed when necessary.[3]

The implementation and survey findings were derived from a smaller sample of 8,000 students from the 32 schools participating in the Next Generation Learning Challenges initiative, one of the three considered in the student achievement results. Researchers collected teacher and student surveys, teacher instructional logs, and held interviews with school administrators. Researchers also conducted one-day site visits at seven of the schools, each of which included brief, 10- to 15-minute observations in at least one mathematics and one English/Language Arts classroom, as well as separate focus groups with instructional staff and students. Teacher and student surveys administered in spring 2015 focused on implementation supports, student study skills, and attitudes toward reasoning. A different contractor, Grunwald Associates, conducted a similar survey with a national sample in summer 2015 with the intent of providing a reference group for teacher and student responses. To compare responses, RAND researchers weighted the survey responses to match the sample in their study. To analyze the relationship between implementation features and student outcomes, the researchers used an innovative strategy called Qualitative Comparative Analysis (QCA),[4] described in greater detail in Section V of this review.

IV. The Report's Use of Research Literature

The reference section of the report consists of just four citations, three referring to research methods used to collect and analyze data and one to an NWEA white paper describing Measures of Academic Progress (MAP) scale norms. There is no cited research literature focused on personalized learning.

Much of what describes the current approach to personalized learning through digitial technology is "grey literature" produced by intermediary organizations like the International Association for K-12 Online Learning (iNACOL),[5] EDUCAUSE,[6] and Marzano Research,[7] as well as by private foundations like the Bill & Melinda Gates Foundation, the Eli and Edythe Broad Foundation,[8] and the Michael & Susan Dell Foundation.[9] Talks by figures such as Salman Khan[10] (Khan Academy) have also outlined the strategies that schools are implementing today. These advocates of personalized learning often draw on research evidence to justify adoption of particular strategies, such as engaging students in conversations about their learning data and grading based on demonstration of student competency,[11] but evidence of efficacy related to the more 'disruptive' strategies that they are advocating is generally lacking in the research literature.

V. Review of the Report's Methods

The methods used in the study to evaluate the effectiveness and to study the implementation of the program were generally appropriate for the purposes of the evaluation, which were to "create a broad picture of the schools' efforts to implement personalized learning and to understand the outcomes that resulted from the adoption of these new teaching and learning practices" (p. 4).

The use of multiple sources of implementation evidence, including from students, is a strength of the methodology because it provided a means for the study team to triangulate findings across different data sources. The evidence related to instruction, however, was rather limited. The observations were too brief and too few to yield valid inferences about instructional quality. Though the use of daily instructional logs per teacher in the study might have permitted the team to draw valid conclusions about the frequency with which teachers engaged in particular practices, the number of days of instruction for which teachers returned data was not reported. Data on instruction were critical for supporting the conclusion that the presence of two of three elements of personalized learning—student grouping, learning spaces that support personalized learning, and opportunities for students to discuss their learning data with teachers—were consistently linked to positive student outcomes.

The research team used a relatively less well-known but appropriate analytic approach to develop conclusions about which characteristics of personalized learning were associated with student learning outcomes. Qualitative Comparative Analysis (QCA) combines qual-

itative and quantitative approaches to data analysis. It is based on the assumption that in research with multiple cases, there are multiple possible combinations of factors that can produce similar outcomes.[12] The analytic approach is particularly appropriate for the current study, in which personalized learning was defined by multiple components and schools' implementations of these components varied widely. By using QCA in the way that study researchers did, they were able to identify which components were associated with reliably positive student achievement gains (again, setting aside the data collection issues noted above).

The use of a matched comparison group was another strength of the study, in that the research team was able to identify students with similar levels of prior achievement and background variables to form a "virtual comparison group." The virtual comparison group is one strategy for reducing bias in an observational study, making the study a kind of quasi-experiment. However, quasi-experiments do not always produce the same results as experiments, in part because unobserved variables may be the cause of observed impact estimates,[13] as the study authors themselves concede and as others have shown. In this case, it is schools and not students that are a potential source of selection bias, and the characteristics used to match schools are not the same as those used to select schools into the initiative. There is, moreover, reason to believe that the personalized learning schools were unusual in some characteristics that might have affected the achievement levels in those schools. Evidence that this is the case comes from the national survey, which found personalized learning schools to have higher rates of collegiality, a factor known to support reform implementation.[14]

There is also good reason to suspect that treatment schools represent a skewed sample of programs within the overall initiative. Charter schools represented 90% of the sample used to analyze student achievement results (p. 9), and nearly half were part of the Charter School Growth Fund, one of the three initiatives studied. The study team conducted analyses focusing on students from other schools of choice to check the robustness of their results, but this strategy cannot eliminate the bias associated with being a school selected as part of a competitive process to be part of a program. The study presents a case where the selection mechanisms are fairly well known but not accounted for in the selection of a matched comparison group, which can lead to biased results.

There is also good reason to suspect that treatment schools represent a skewed sample of programs within the overall initiative.

Another concern is the different samples used for different analyses. The achievement analysis had only a partial overlap with the implementation analysis. The authors did not provide a reason for why the samples were different. Moreover, different samples mean that the observed associations between particular personalized learning features and outcomes might not be true for the full sample. The population for the national survey, administered by a separate research group, was not clear. Though the RAND researchers sought to weight these schools based on similarities to schools in the sample, the lack of definition for the

population makes it difficult to interpret those survey results.

VI. Review of the Validity of the Findings and Conclusions

As this study did not employ random assignment, the authors are correct to caution against interpreting their results as causal. Other cautions appear in the report as well, yet claims of "positive and promising" findings deserve further scrutiny than the report's headlines would indicate. "Promising" might accurately describe the likelihood that a similar study with similar schools, data, and methods might produce similar results, but the study should not be taken as evidence that any school can implement personalized learning in the varied ways described by the report and expect to see positive results or results of the same magnitude.

The student achievement findings yielded significant results in favor of personalized learning on average, with considerable variability across schools. However, questions remain about the selection of schools and sampling of students. The findings therefore should not be generalized to say, "personalized learning has a positive effect in schools," as the authors themselves concede (p. 34). Shifts in sampling complicate any such claims: the 62-school sample representing three personalized learning initiatives used for analyzing student achievement results was reduced to a sample of 32 schools representing a single initiative for the implementation findings. Moreover, nearly half of the schools in the implementation analysis were high schools, even though high schools showed no significant effects on student achievement when included in the larger sample.

Of further concern is the lack of a threshold for what qualified as implementing personalized learning in the treatment schools. The authors of the report acknowledge that implementation "varied considerably" (p. 14), which is not unexpected. It seems that some personalized learning schools had features of personalized learning commonly found in all schools, such as using student data to inform instruction, using varied instructional formats (small-group, large-group, individual), and flexible grouping of students. The mixed responses in the student survey results support the claim that schools in the treatment and comparison group are not so different in their use of personalized learning strategies, at least from the perspective of students. With few of the treatment schools relying on more novel approaches to personalized learning, such as competency-based progression, readers should be skeptical of what promise the report's evidence actually provides for the model of personalized learning promoted through the three initiatives studied.

VII. Usefulness of the Report for Guidance of Policy and Practice

Though the study conclusions were appropriately drawn with clear attention to limitations of the study design, the study's relevance to policy and practice is limited.

One reason is that 90 percent of the schools in the student achievement study were charter schools selected into a competitive funding program. Though the study might be relevant to charter schools that also met the criteria for inclusion in the competitive program, those criteria are not presented in the report, and so it would be difficult for school leaders—including charter school leaders—to know whether the study findings could reasonably apply to their schools.

Also, most of the schools had a 1:1 ratio of students to computers. Though there are many schools with such ratios, the average ratio in U.S. public schools in 2009, the last year for which statistics are available, was 5.3 to 1.[15] Given that models of "personalized learning" make strong claims about the importance of technology, this is an important limitation of applying the study results more broadly.

Finally, the study lacks utility for judging the value of the more disruptive and digital-technology-based personalized learning, simply because some strategies that require large departures from current practice were not implemented. Two of the factors associated with positive learning gains—student grouping and making flexible use of learning spaces—do little to distinguish these schools from many other schools that may not claim to be implementing personalized learning. Only engaging students in analyzing their own data showed a consistent relationship to positive outcomes, and so the study does not provide strong evidence for the claim that novel forms of personalized learning can improve learning outcomes.

Notes and References

1   Pane, J. F., Steiner, E.D., Baird, M.D., & Hamilton, L.S. (2015). *Continued progress: Promising evidence on personalized learning*. Santa Monica, CA: RAND Corporation. Retrieved December 15, 2015 from http://www.rand.org/pubs/research_reports/RR1365.html

2   Enyedy, N. (2014). Personalized interest: *New interest, old rhetoric, limited results, and the need for a new direction for computer-mediated learning*. Boulder, CO: National Education Policy Center. Retrieved December 26, 2015 from http://nepc.colorado.edu/publication/personalized-instruction

3   For more on how NWEA constructs virtual comparison groups for its partners, see https://www.nwea.org/content/uploads/2015/09/Virtual-Comparison-Group-VCG-FAQ-JUL15.pdf

4   Ragin, C.C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago, IL: University of Chicago Press.

5   Patrick, S., Kennedy, K., & Powell, A. (2013). *Mean what you say: Defining and integrating personalized, blended and competency education*. Vienna, VA: International Association for K-12 Online Learning (iNACOL). Retrieved December 26, 2015 from http://www.inacol.org/resource/mean-what-you-say-defining-and-integrating-personalized-blended-and-competency-education/

6   EDUCAUSE. (2009). *7 things you should know about personal learning environments*. Louisville, CO: EDUCAUSE. Retrieved December 26, 2015 from http://www.educause.edu/library/resources/7-things-you-should-know-about-personal-learning-environments

7   Marzano Research. (2015). *Competency-based education*. Bloomington, IN: Marzano Research. Retrieved December 26, 2015 from http://www.marzanoresearch.com/competency-based-education?pd_type=215&topics=210

8   The Broad Foundation. (2013). *Broad Foundation invests nearly $23 million in nonprofits working to empower teachers and parents with educational technologies that personalize learning for every student*. Retrieved December 26, 2015 from http://www.broadfoundation.org/asset/1165-130313digitallearningrelease.pdf

9   Michael & Susan Dell Foundation. (2015). *Personalized learning blog team*. Retrieved December 26, 2015 from http://www.msdf.org/blog/author/personalized-learning-blog-team/

10 Khan, S. (2014). 2014 iNACOL Symposium - Sal Khan keynote address. Presentation, Palm Springs, CA. Retrieved from https://www.youtube.com/watch?v=-9MOkgRJmhw

11 Sturgis, C. (2014). *Progress and proficiency: Redesigning grading for competency education*. Vienna, VA: International Association for K-12 Online Learning (iNACOL). Retrieved December 26, 2015 from http://www.inacol.org/resource/progress-and-proficiency-redesigning-grading-for-competency-education/

12 Ragin, C.C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago, IL: University of Chicago Press.

13 Cook, T.D., Shadish, W.R., & Wong, V.C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within study comparisons. *Journal of Policy*

Analysis and Management, 27(4), 724-750. http://doi.org/10.1002/pam.20375

14   Bryk, A.S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement.* New York, NY: Russell Sage Foundation.

15   National Center for Education Statistics. (2010). *Teachers' use of educational technology in U.S. public schools: 2009* (NCES 2010-040). Washington, DC: U.S. Department of Education. Retrieved December 26, 2015 from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010040