

DRAFT Report– Please do not cite or distribute without permission of the authors

Can a State Graduation Test be Transformed into a Formative Assessment?

Preliminary Findings from the Randomized Evaluation of

Ohio's Personalized Assessment Reporting System

Henry May and Marian A. Robinson

Consortium for Policy Research in Education

University of Pennsylvania

The research reported in this paper was conducted by the Consortium for Policy Research in Education (CPRE). Opinions expressed in this paper are those of the authors and do not necessarily reflect the views of districts, schools, or subjects studied; CPRE; or its institutional members. This paper was internally reviewed by CPRE research staff; however, it has not been externally reviewed.

Paper Presented at the Annual Conference of the

American Education Research Association

April 10, 2007

Chicago, IL

Ohio has been forging a new path in its statewide programs to promote the use of assessment data to inform instruction. What is especially interesting is that Ohio's newest initiatives are designed to promote formative use of data from the state's accountability tests. This began with Ohio's contract with American Institutes for Research (AIR) to develop enhanced test score reports for grades 3 through 8, and it continues as Ohio works with the Grow Network (McGraw-Hill) to launch a pilot of its Personalized Assessment Reporting System (PARS) in fall of 2006.

Maximizing the information obtained from state tests has become Grow Network's claim to fame. Their "Grow Reports," which provide detailed numerical and graphical accounts of a student's strengths and weaknesses in relation to standards, have been thoroughly embraced by many teachers, administrators, students, and parents in large districts (e.g., New York City, Philadelphia) and in several states (e.g., California, Florida, Texas). In Ohio, the Grow Network is taking their approach to a whole new level of personalization with the Ohio Graduation Test (OGT). In summer and fall of 2006, Grow Network provided multiple OGT score reports for each school showing the distribution of student performance by subject, by standard, and even by item. Grow also provided an OGT score report for each student that is personalized based upon the student's performance profile. The degree of customization is substantial, with literally thousands of possible variants of score reports across the population of students. In addition to the printed reports, Grow has also launched two new websites for Ohio: one website that allows educators (e.g., teachers, principals, administrators) to analyze test data, learn about instructional strategies, and access professional development tools; and a second website for students with support for those retaking the OGT including online

tutorials, and advice to all students regarding high school course selection and planning for college and career.

This paper presents early impressions from an ongoing mixed-method randomized field trial of Ohio's PARS. As data collection activities will continue through April and May 2007, and data analyses will continue through June 2007, the results presented here do not necessarily represent the final conclusions of the evaluation. The findings presented here are based upon preliminary analyses of student achievement data and qualitative data from interviews and focus groups with students and teachers.

Theoretical Framework

The PARS program may achieve positive impacts through two primary pathways. The first pathway focuses on student motivation and behavior. By providing clear and detailed information to each student about their performance relative to the state standards, along with resources for improving performance and planning for the future, the PARS approach may lead to changes in student attitudes and behaviors that are fundamental to success in high school and beyond. These areas of potential impacts on students include:

- Students' understanding of standards and expectations
- Students' motivation to improve
- College aspirations
- Targeted studying
- High school course selection (e.g., taking calculus as a senior)
- Improved communication with teachers
- Improved communication with parents

The second pathway through which the PARS program may achieve positive impacts focuses on changes in school-wide instructional policy and individual teachers'

instructional decisions. By providing clear and detailed information to teachers and administrators about student performance, along with tools for making sense of the data and resources for improving and targeting instruction, PARS has the potential to inform numerous aspects of instruction.

- Diagnosing individual student strengths and weaknesses
- Uncovering class-wide and school-wide trends
- Targeted remediation for students who need extra support
- Dynamic/flexible grouping of students for instruction
- Data-driven differentiated instruction
- Improved teacher collaboration
- Improved teacher-parent communication

Additional impacts on parents are also anticipated, primarily with regards to parents' understanding of the standards and what is expected of their child. Access to clearer information and a better understanding of expectations may also lead parents to provide better support for their children.

Methods

This evaluation began in late 2005 with the selection of a stratified sample of 60 school districts, representative of the entire state, which were then randomized to treatment and control groups. The thirty treatment districts received their first round of PARS reports and access to the PARS websites during the summer of 2006. A second round of PARS reports were delivered to treatment schools after the October retake administration of the OGT. The thirty control districts have received only the standard OGT score reports.

The randomized control design of this evaluation produces estimates of the PARS program's impact on teachers' practice and student outcomes while minimizing or

eliminating bias associated with potential confounding variables. Post randomization analyses confirmed that the treatment and control groups, prior to the PARS program, had no statistically significant differences with regards to prior student performance, student demographics, or school contextual factors (e.g., school size, location).

Data Sources

This study involves a mixed-method approach designed to accomplish two goals. Surveys are used to collect broad information from across a large sample of teachers, administrators, students, and parents; while interviews and focus groups are used to collect detailed, contextualized data from within the larger sample.

The teacher survey is designed to collect information on teachers' use of data to make instructional decisions, their perceptions of their students' abilities, their perceptions about their schools, and their utility of the information from PARS and the PARS reports. The survey of students is designed to gauge students' utility of the information from the PARS reports, their perceived self-efficacy, and their efforts to improve their academic performance. The survey of parents is designed to gauge parents' understanding of the information from the PARS reports, their reactions to the information from the PARS reports, and their efforts to help improve their child's academic performance. The surveys used in the control groups are identical, but focus on the use of information from the traditional OGT Reports instead of the PARS reports and websites.

In January 2007, surveys were mailed to 800 teachers in four subjects along with 400 students and their parents. Optical scanning of completed surveys and data quality

control processes are currently underway. Therefore, no results from analyses of survey data appear in this paper.

Interview and focus group protocols for teachers and administrators in treatment schools were developed to collect information on teachers' use of the information contained in the PARS reports, their use of PARS tools, and their ideas on how the PARS reports and websites might be improved. Interview and focus group protocols for students and parents were developed to gauge students' understanding of the information contained in the Grow reports, their thoughts on the usefulness of the information contained in the Grow reports, and their use of that information to improve academic performance. Similar interview protocols were developed for control schools, but these are limited to teacher focus groups and concentrate on the use of OGT data and the utility of the traditional OGT reports.

To date, CPRE researchers have conducted 4 focus groups with more than 20 teachers, individual interviews with 7 additional teachers, and individual interviews with 10 students. Key preliminary impressions from analyses of transcribed data from a subset of these interviews and focus groups are presented in this paper. Although these impressions are based upon only the most consistent responses across the interviews and focus groups, they do not necessarily reflect the final conclusions that will be based upon the complete set of qualitative data.

As the primary measure of student performance, student test scores from the March 2006 and October 2006 administrations of the OGT have been collected for more than 20,000 students across the 60 participating districts. The difference in retake rates

and retake passing rates for the treatment and control groups serves to estimate the impact of PARS on student OGT performance.

Challenges to the Study Design

As is usually the case in large-scale education research, there were several challenges in implementing the planned research design. Most of these were anticipated prior to the start of this evaluation; however, the extent of these potential problems was not known. As such, we have documented these problems in order to understand their potential limiting effects on this research.

Arguably the most significant factor was that the PARS reports and websites were delivered as supplements to existing OGT reports and tools. As such, the PARS intervention had to compete with previously existing and familiar resources. One might expect that this would not be a problem for the PARS intervention given that its components were designed specifically to be more useful and intuitive than existing OGT reports and tools. Unfortunately, due to development delays, the PARS reports and tools were not made available to schools until the summer of 2006, when most students were out of school (the original distribution date was in the spring of 2006). As such, some schools did not distribute the PARS reports during the summer. A principal at one of these schools explained that,

To be honest with you, when they came to the school was my biggest gripe about this. It almost made these [PARS reports] unusable because we would send out the ones from the state department to the students and to the students' parents. As soon as they came in, we'd mail those out and get them taken care of. And then these [PARS reports] would come in two to three, maybe four weeks after that. And at that point, they are much easier to read, much more user friendly, I think they're great. But, since

they were so late, we didn't want to send out a second group of reports and end up confusing parents. 'Oh, this must be the retakes. Oh no, my kid still failed.' and have them get all upset about it.

This situation makes it clear that the PARS intervention was not implemented as designed. The extent to which PARS reports were actually delivered to students is an essential question and will be estimated via data from the student survey. Plans to extend the evaluation to include more interviews with principals are being considered.

A second important factor in the implementation of PARS was the quantity and design of professional development (PD) for teachers. More specifically, the PARS project utilized a "train-the-trainers" model of PD. While this model is less expensive than direct PD, it is dependent on the capacity of districts and schools to support the local training of their teachers. The training provided to the intended trainers was intensive and well-received. Comments like "I wish I had this when I started teaching!" were heard many times in the Grow Network professional development sessions. Unfortunately, the training the most teachers received was less helpful. Limits on time and resources available to administrators and teachers appear to have contributed to sparse PD sessions where teachers were given a copy of the PARS school report and then watched a demonstration of the website without having a opportunity to use it themselves and ask questions.

A third challenge encountered in this research was a lack of resources to provide incentives for individuals to complete surveys or for schools to participate in interviews and focus groups. In recognition of this, the surveys were designed to fit easily on two pages and ask straightforward fixed-response questions. Each survey was designed to be completed in less than five minutes. Surveys continue to be returned to CPRE, although

we expect the overall response rate to be relatively low. Recruitment of schools for fieldwork has required a tremendous amount of correspondence with schools in order to get an initial response. As such, the schedule for fieldwork had to be delayed and extended to provide a representative sample of schools. Fieldwork will continue into May, 2007.

Preliminary Findings

Student Achievement: Preliminary analyses of student achievement data suggest that PARS has not had a significant effect on retake passing rates; however, PARS appears to have a positive impact on the rate by which students attempted to retake the OGT in October if they had failed in the previous March. More specifically, students in the treatment schools were 8% more likely to attempt to retake the OGT than students in the control schools ($p < .05$). This is important for two reasons. First, the positive effect of PARS on retake rates suggests a positive impact on student motivation. Second, a positive impact on retake rates may actually mask a positive effect on achievement. This is because the average score on the OGT tends to go up as lower performing students drop out of school. Planned analyses of student achievement data will use propensity scores and/or Bayesian estimation to adjust the overall PARS impact estimates for differential retake rates between the treatment and control groups.

Development of the PARS Reports and Websites: The most important findings to date concerns the nature of the relationship between the Ohio Department of Education (ODE) and the Grow Network. The development process was collaborative and iterative, involving many cycles of production and revision that pushed back the roll-out schedule, but greatly enhanced the final products. The strategic use of color and graphics for clear

visual representation of complex data trends is present in both the print reports and the websites. ODE contributed a great deal of time and effort to ensure that the written content in the reports and websites was concise and understandable. The Teaching Tools from the educator website are also the result of a substantial development effort. There are literally hundreds of pages designed to help teachers enhance their understanding of the standards and develop instructional strategies that cover every standard in every subject.

Yet the substantial effort to develop some parts of the PARS system may have contributed to a lack of resources to develop other aspects of the system. For example, while the online student tutorials appear to have a great deal of potential for enhancing students' study experiences, there were only a handful of tutorials developed prior to the October OGT retake administration. In mathematics, the five available tutorials concentrate exclusively on finding the surface area and volume of basic three-dimensional objects.

Teachers' Views About the Traditional OGT Reports: In the single control school visited to date, teachers found many of the OGT reports available to them to be too general to be helpful. Teachers reported having open access through the principal's office to OGT data reports that included the school report, individual subject reports, and an item analysis report, which showed performance across the entire school for each item. A few of these reports were passed out to teachers during an all-faculty in-service at the beginning of the school year. Of these, the item analysis report was the one report that was deemed helpful by most teachers. These teachers also agreed that this report could be much more helpful if it were available for subgroups of students, particularly for their

individual classes. Moreover, teachers wanted access to the item level data for each student so that they better diagnose students' misunderstandings in a particular standard. In this regard, one teacher noted that, "My question is why don't they hand the test back? We do not know how they thought on that question. A lot of it is the thinking process. Why did they choose B?"

A couple of teachers were also aware of a "big huge binder" in the principal's office that contained individual student reports and all other OGT data printouts, but only one teacher had spent some time looking through the binder to investigate a student's performance. Other teachers in the focus group were interested to learn about this resource, but the first teacher warned that sifting through the binder to find individual students was time intensive, perhaps "15 minutes a student."

Teachers' Views About the New PARS-OGT Reports: Teachers in the treatment schools found the PARS reports easy to understand and well-organized. They felt that the new format was much easier to understand than the traditional OGT reports. Although most teachers had only seen the school-wide report, they were familiar with the format and content of the retake and intervention reports. More importantly, most teachers said that they would prefer to get the intervention report, the retake report, and the student roster through the website. One teacher added that she would like printed reports, but that it was easy to print them from the PARS website.

When asked about the clarity of the school report, one teacher questioned the descriptions of overall school performance by saying, "I am confused and a little deterred by the wording, 'Science is not an area of particular strength or need.' It sounds confusing. Don't you think that's a little bit of weird wording?" Other teachers agreed

and a discussion of intended meaning and alternative wordings ensued. The group decided that the report was trying to distinguish between “areas of strength,” “areas that need improvement,” and “areas that need much improvement.” It was also suggested that the symbols be color coded green, yellow, and red, “We all know what green, yellow, and red mean.”

Another teacher pointed out inconsistent use of color on the school report. “On page 1, dark green is advanced, and dark brown is limited. Then I look at the first graph and it looks like everyone listed there is advanced or limited.” While she and other teachers agreed that it was easy to figure out the correct interpretation of the graph, the inconsistent use of color was potentially confusing.

Teachers’ actual use of the printed PARS reports was very limited. This appeared to be the result of two factors: receiving only a copy of the school report and the ability to access the same information and even more detailed information using the PARS website. It was very apparent that most teachers found the website to be a more useful tool than the collection of printed reports.

Teachers’ Views About the New PARS-OGT Website: A majority of teachers at the treatment schools had spent some time reviewing the PARS website either in the beginning of the school year or in preparation for the focus group. Of these, most reported spending a total of less than an hour using the website. Preliminary explanations for the minimal amount of time that teachers’ used the website include limitations on teachers’ time and little or no opportunity for training. Nonetheless, teachers agreed that the website is a potentially powerful resource.

An unanticipated aspect of our focus group was that it provided an opportunity for teachers to learn about the different website features from each other; teachers frequently asked questions of each other during the conversation to probe the website's features.

Overall, teachers were highly positive about the PARS website, found it easy to navigate, and expressed disappointment that they had not fully understood its unique features and had an opportunity at the school to visit the website during the school day and with peers.

Teachers made comments such as:

Teacher 1: I think it's useful. I think it's easy to navigate once you're in. I think it's useful, you know, if you're willing to put forth the time as an individual, as a teacher.

Teacher 2: Well, next year I will find time.

Teacher 3: It will take us time to use this and implement it and make it beneficial for our school.

One advantage for teachers was the website's capacity to serve as a one stop shop for all materials related to the OGT and teaching. The website's integration of multiple sources of information about the test – cohort performance, OGT test items, Ohio Standards and Benchmarks, and related lesson activities – could facilitate an efficient and thorough analysis. Teachers at the treatment school found the PARS website valuable for creating a complete loop from OGT performance to instructional strategy.

Teacher 1: Before when I was looking at the [student] roster and I saw an area that wasn't proficient and then I went to the report and I looked up the actual, you know the area, 'Earth and Space Science' and I went and got all the benchmarks and then I only printed off one lesson plan that I could use...I was able to flow through it nicely and I understood where I could go to get something ...On the second page it says click here – get it -- and you have a lesson right there ready for you.

Teacher 2: Yeah. If you clicked ‘Earth and Space’ it pulled up the benchmarks. Click the benchmarks it pulls up the lesson plan. I mean it doesn’t get any easier.

The value of this integrated centralized resource for all things OGT comes into high relief when compared to the frustrations of teachers at the control school in accessing OGT data. For these teachers their full use of OGT data was constrained by the time required to locate and reference all OGT materials, located in the “big paper binder” and beyond, just to interpret student performance.

In sharing the review of the PARS reports and website, teachers shared what they learned about student performance and how it identified areas of their classroom practice that could be improved. Not all anticipated uses were mentioned during the focus groups. In fact, teachers tended to interpret the data and lay out actions that reflected a fine-tuning of their existing practice rather than any radical rethinking of what or how they designed learning opportunities for their students. The actions reported tended to reflect the same set of three strategies already in use at these schools: practice taking the OGT test using old OGT items, learning and using test taking tips, and aligning curriculum in terms of selection of topics and time allotment to the OGT test topics. In general, most teachers found that the data analysis component of the website was helpful for uncovering class-wide trends in student performance, while the instructional tools section of the website served to validate or supplement their current instruction.

Teacher 1: I found a correlation with some of my lessons, you know, some of the things I don’t spend as much time on. With my kids, I could see it there... This year, I did see some things I needed to change.

Teacher 2: “When I looked through these lesson plans, I thought do my lesson plans look close to it? Am I teaching the subject matter? Although I

may teach it differently, but am I teaching it? And I found, well, I am teaching it. So it made me feel like I am doing was the state is saying or covering. Like this one here. Actually, I do that. I don't do this exact lesson, but I do a lesson that's like this one."

Although it was clear that teachers could and did use the PARS website to inform their classroom instruction to a limited extent, there was little evidence that teachers used the system to diagnose individual student's needs. Much like the teachers in the control school, the treatment teachers found the PARS website limited in that it did not provide access to individual item-level data. One teacher in particular described the frustration felt when trying to dig into the data using the PARS website.

We had a student who was a senior who was going to be taking [the OGT] again and needed to pass. And we were trying to find out the areas the he really needed to focus on. But we were having trouble finding his name and finding what he needed to work on. And there were three of us together working on this—three teachers working on this. What you get is this report, which says that they are either proficient or not proficient in the strand. But what we want to see is the actual questions they missed—what questions did they get correct; what questions don't they get correct. And I hate to say it, but at least in social studies, you can look at the strand title, and that still does not tell you what they're really missing. It says, 'History.' Well, what does that mean? Is it American History, or is it the World History questions they're missing? But you can't click on the student's name and see specifically which questions they got right and which ones they did not get right. You have to go back to the stand where you can then find the questions, but the problem is that there's a certain number of questions that fall under American History and a certain number that fall under World History. So, we can't tell what areas we need to make improvements in.

This theme was apparent in all the schools visited to date. Teachers feel that they need to have access to item-level data for each student in order to get enough information to diagnose students' needs.

Students’ Views About the New PARS-OGT Reports and Website: In both treatment schools we visited, about half of the students interviewed had never seen their PARS-OGT report. They had only received a traditional OGT report. After taking time to look through an example report, all students stated that they preferred the PARS report over the traditional report because it was easier to read and understand.

None of the students interviewed had ever visited the PARS student website. They reported that they weren’t aware that the website existed. Even though the website address was shown on the PARS student report, no students remembered seeing it. One student suggested that the website address would need to be moved toward the top of the page and increased in size in order to make it visible and prominent enough for students to notice. When students were asked if they would have used the website to prepare for the OGT had they know about it, about half of the students interviewed said “yes.”

Improving the Future Value of the PARS Reports and Website

Teachers’ feedback on the PARS reports and website, and about the overall challenge to using data and instructional resources to support efforts to prepare students for the OGT, suggest a number of improvements. Below are a few preliminary points for discussion.

1. Improve student awareness of and access to OGT resources
2. Edit the performance descriptors in the PARS school report
3. Get the word out on ODE sponsored OGT preparation resources
4. Increase support for teacher’s use of PARS website.
 - a. Emphasize the particular value of these tools compared to the other OGT preparation materials that are available in the state.

- b. Allocate resources to provide teachers time and assistance needed to fully use the website. For example, dedicate a faculty in-service day to the review of the tools on-line in pairs or teams. Better yet, implement a series of common planning periods dedicated to exploring patterns in OGT data and discussing instructional responses as a department.
5. Increase teachers' access to reports online. Rather than deliver small number of each report to the school, post copies of the reports online so that teachers can review the reports with the color enhancements and print those they find most helpful.
6. Improve the utility of PARS data analysis tool. Consider providing teachers with access to individual student responses

References

- Boruch, R. F., May, H., Lavenberg, J., Turner, H. M., Petrosino, A., De Moya, D., Grimshaw, J., & Foley, E. (2004). Estimating the effects of interventions that are deployed in many places: Place randomized trials. *American Behavioral Scientist*, 47, 608-633.
- Boudett, K. P., City, E. A. & Murnane, R. J. (2005). *Data Wise: A Step-by-Step Guide to Using Assessment Results to Improve Teaching and Learning*. Cambridge, MA: Harvard Education Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-75.
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79, 474-482.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1-14.
- Choppin, J. (2002, April). *Data use in practice: Examples from the school level*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Elawar, M. C., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. *Journal of Educational Psychology*, 77, 162-173.

- Feldman, J., & Tung, R. (2001, April). *Whole school reform: How schools use the data-based inquiry and decision making process*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*, 199-208.
- Herman, J., & Gribbons, B. (2001). *Lessons learned in using data to support school inquiry and continuous improvement: Final report to the Stuart Foundation* (CSE Technical Report. 535). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Hupert, N., Martin, W., Heinze, C., Kanaya, T., & Perez, H. (2004, June). *Trends in the use of handheld technology to support student reading assessment*. Paper presented at the National Educational Computing Conference, New Orleans, LA.
- Kluger, A., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.
- Mandinach, E.B., Honey, M., Light, D., Heinze, C., & Nudell, H. (2005, April). *Data-driven instructional decision-making using technology-based tools*. Paper presented at the Annual Meeting of the American Education Research Association, Montreal, Canada.
- Mason, S. (2002). *Turning data into knowledge: Lessons from six Milwaukee public schools* (WCER Working Paper 2002-3). Madison, WI: Wisconsin Center for Education Research.

Sharp, D. (2004). *Supporting teachers' data-driven instructional conversations: An environmental scan of reading first and STEP literacy assessments, data visualizations, and assumptions about conversations that matter*. Report to the Information Infrastructure System Project, the John D. and Catherine T. MacArthur Foundation and the William and Flora Hewlett Foundation.

Supovitz, J.A. & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement*. Philadelphia: Consortium for Policy Research in Education.