

Assessment for Learning in K-12 Science Departments: A Case Study

William J. Glenn

School of Education, Virginia Tech

ABSTRACT: This paper discusses the efforts of science departments in two secondary schools to implement data-driven decision making at the department level. It covers the first year of the process and identifies the successes and barriers to success experienced by the two departments. The catalyst for the change at each school was the department chair studying data analysis and applying the newly learned techniques to stimulate their science departments to study assessment data. The departments were surprised by some of their findings. For example, the faculty at one school found that their students performed poorly on one strand covered in the state's science assessment. However, their data analysis showed that their students scored about equally well on each strand. The findings led both schools to consider and implement various changes. Both departments saw the need to use shorter and timelier formative assessments to drive instruction throughout the year. The departments encountered several barriers to using data-driven decision making to guide instruction. They found it easier to identify issues that needed to be addressed than implementing solutions.

KEYWORDS: *evidence; data interpretation; formative evaluation; educational improvement; communities of practice; program effectiveness*

Data-driven decision making has become a cutting-edge topic in K-12 education. Schools possess more data than ever, due in part to the accountability provisions of the No Child Left Behind legislation (Mandinach, Honey, Light, & Brunner, 2008). At the same time, technological advances provide easy access to the tools needed to analyze educational data (Wayman, Stringfield, & Yakimowski, 2004).

Data-driven decision making offers schools the opportunity to improve their decision making capacity by basing educational decisions on data rather than perceptions about students. Researchers have written about several different, but related, processes that can be used to model the use of data in schools (Boudett,

City, & Murnane, 2007; Kowalski, Lasley, II, & Mahoney, 2008; Mandinach, Honey, Light, & Brunner, 2008). Each of these models contains certain basic elements, such as identifying problematic issues at the school, using data to understand the issues better, proposing solutions, creating measurable objectives to track progress, measuring progress, and reevaluating the issues at the school in the light of the new data. Data-driven decision-making procedures are designed to provide a systematic means for schools to improve outcomes for their students (Boudett, City, & Murnane, 2007; Kowalski, Lasley, II, & Mahoney, 2008; Mandinach, Honey, Light, & Brunner, 2008).

Despite the potential benefits of data analysis, schools have been slow to use the wealth of data in their possession to modify existing practices and direct teaching and learning (Wayman & Stringfield, 2006). Researchers have identified a number of barriers that inhibit the use of data by schools. These barriers include questions about whether data can truly guide instruction in an effective manner; school cultures that offer resistance to using data; fears that data, including results of formative assessments, will be used to evaluate rather than improve teachers and schools; and insufficient capacity in the management and use of data (Earl & Katz, 2006; Lachat & Smith, 2005). The capacity issues exist even though the skills required to make sound use of educational data are quite straightforward in most cases (Glenn & Creighton, 2008).

This paper discusses the efforts of science departments in two secondary schools to implement data-driven decision making at the department level. It covers the first year of the process and identifies the successes and barriers to success experienced by the two departments.

Method

The researcher took a case study approach to understanding the first year of implementation of data-driven decision making at two science departments in public schools in the southern part of the United States. Case studies involve in-depth studies of a small number of cases in order to gain a deeper understanding of a phenomenon (Yin, 2009). The two science departments were selected because each department chair had received professional development in the area of data-driven decision making and was attempting to implement the process in the department. The schools acted independently in order to address their individual needs.

The primary data source consisted of focused interviews with teachers at the sites, including interviews with both department

chairs. The focused interviews concentrated on the topics that were important for this study (Yin, 2009). The interviews were designed to elicit information from the participants regarding their perceptions of the first year of the implementation. In particular, the interviews focused on how the teachers reacted to the new uses of data, the areas of improvement identified by the teachers, the plans for addressing the issues, the success of the plans, and the barriers to success. The secondary sources of data were field visits and documentary evidence, including results of formative and summative evaluations. These sources of data provided triangulation to assess the accuracy of the interview data and showed both the starting point of the analysis within each department and the progress that had been made through the year.

Results

The findings from the two schools shared several similarities, as well as some important differences. One of the key similarities relates to the fact that each department chair had participated in professional development activities related to data-driven decision making and brought what was learned to the department in an effort to improve outcomes for children. A second similarity is both departments focused their efforts on improving scores and pass rates on their respective standardized tests. The rest of this section summarizes the findings from the two departments.

Washington

Washington School¹ serves grades 7-8 in a mid-Atlantic suburban district. The midsized school serves a diverse student population, with no racial/ethnic group comprising more than forty percent of the students in the school. The

¹ The school names in this paper are pseudonyms.

science department at the school consists of seven teachers.

The Washington School Science department began its data-driven decision making process with a hypothesis about why the scores and pass rates of its students on the state's eighth grade science assessment were good, but not great (between eighty-five and ninety percent passing each year). The state assessment covered material that spanned three years of the state's science standards. The accepted wisdom of the department, which became the hypothesis, was the scores were brought down because students scored lower on material covered in the lower grades, especially in sixth grade. Sixth grade was regarded as a problem area because students at Washington attended sixth grade science in elementary school and entered Washington in the seventh grade.

The hypothesis testing followed a very straightforward procedure. The department chair disaggregated the test scores by the subscore for each of the curricular areas covered on the test. The mean subscores for each area were compared with one another, using three years' worth of test score data. The pass rates for each of the strands were also compared over a three-year period. The results showed no difference between test scores on areas covered in the year the test was taken versus areas covered earlier. The mean scores were within four percent of each other in each year, while the pass rates for each test area hovered around eighty-five percent. The department's analysis showed that the initial hypothesis did not explain the test score results.

The department held discussions about how to change to improve instruction, but decided that the standardized test was not the most useful assessment for that purpose. The issues that they had with the test included these:

- It is given at the end of the year, which is too late to influence instruction.
- It covers more than one year of material, which means feedback on

teaching in the lower grades comes over a year after the fact.

- The test covers too much ground in too little depth to guide instruction.

The department decided that it needed to implement frequent preassessments and shorter, more focused formative assessments to guide instruction. The school administration supported and encouraged this effort. The department was instructed by the administration to move at a reasonable pace, without trying to take on more than it could handle effectively. The formative assessment approach employed by the department involved using a mixture of school district formative assessments and assessments created by members of the department.

The first year of the data-driven decision-making process was regarded as a qualified success by the department members. The overall response to the initial data analysis was positive because the analysis provided new information about their students. The teachers were very enthusiastic to examine the data and were shocked to learn that the hypothesis did not hold up to a data-based analysis. The eye-opening use of data sparked enthusiasm in the group and a commitment to improve across the board.

The implementation of changes was not as successful as the initial analysis. The implementation ran into several barriers to success noted by the staff in the interviews. First, despite the agreement that formative assessments were better data sources for instructional improvement, some teachers resisted using them due to the time commitment needed for formative assessment, especially in the short run. These teachers described the overly broad nature of the state's science curriculum, which forced them to cover many topics very quickly, leading to a concern that adding formative assessments would slow down teaching the material.

The second set of barriers related to misunderstandings of the uses of formative assessment. An important concern was the fear that the assessments would be used as an evaluation tool by administrators. Interestingly, another issue was the tendency of some teachers to assign a grade to every assignment, thereby using the formative assessment for evaluative purposes with regard to their students. The use of assessment to guide, rather than evaluate, instruction appeared to be a shift that some teachers found difficult to make.

The third category of barriers was resistance to change on the part of various teachers. This issue was exacerbated by the perceived lack of benefits from the preassessments. The teachers hoped that the preassessments would show that students already knew various concepts in the curriculum, so that the teachers could spend less time on those concepts and more time on others. This hoped-for benefit carried the promise of reducing the need to cover some of the broad nature of the curriculum. Unfortunately for the teachers, the preassessments showed that the majority of the students needed to cover the entire curriculum, thereby eliminating the time-saving promise of the preassessments.

Lincoln

Lincoln School serves grades 9-12 in a southern exurban district. The fairly large school serves a diverse student population, with no racial/ethnic group comprising more than fifty percent of the students in the school. The science department at the school consists of twenty teachers.

Lincoln School began its data-driven decision-making process without any hypothesis regarding what would be found. The department chair wanted the faculty to approach the process with as much of an open mind as possible. A variety of summary graphs were prepared disaggregating the state end-of-course test score data by subgroup and by

subtopic over a three-year period. The teachers analyzed the data and reached a consensus that the problem they wished to address involved the disparity between children receiving special education services and those not receiving them. The primary concern among the teachers was the pass rate differential, with children receiving special education services passing at a rate of roughly fifty percent, compared with the pass rate of more than eighty percent for their peers.

After making this determination, the teachers attempted to identify the underlying causes of the disparities. They identified four areas that they thought contributed to the disparate outcomes: prior knowledge, graphing skills, understanding of the scientific process, and reading comprehension. The teachers devised strategies to improve the performance of children receiving special education services in these areas. However, the teachers only implemented one of the strategies: an increased emphasis on graphing. They incorporated graphs into more of the lessons in order to try to improve the skill of the students in reading graphs and making inferences based upon them. As with Washington, the teachers at Lincoln displayed great enthusiasm when engaging in the initial data analysis and realized that the test score data was insufficient to inform instructional change. The school in general had begun implementing formative assessments several years earlier, so the teacher resistance to conducting formative assessment was essentially nonexistent.

The formative assessment process was the more difficult step at Lincoln, as at Washington. Unlike the administrators at Washington, the Lincoln principal touted common formative assessments as being a way for administrators to keep tabs on student progress. The emphasis on administrative oversight catalyzed teacher concerns about formative assessment being used for evaluation rather than to guide instruction. The department chair explained that the fear became so pervasive that one teacher

appeared to be doctoring his results in order to look better in the eyes of the administration. The apparently falsified assessments from one classroom made the entire process almost unworkable in the eyes of the department chair.

Some teachers in the department also tended to use formative assessment data for its predictive value rather than for instructional improvement. One teacher in particular was renowned in the department for his ability to predict standardized test scores for each of his students based on the formative assessment results. His explanation of the accurate predictions involved him seeing the scores from the assessments, which were conducted several weeks before the standardized tests, and, with experience, estimating that students tended to improve by X number of points during the interval between the formative assessment and the state assessment. The teacher made no indication, however, that he used the formative assessment results to guide his instruction in order to make the gain greater than X.

As with Washington, Lincoln teachers consider data-driven decision making to be of mixed value, for similar reasons. The teachers responded very well to seeing the gaps and resolved to reduce the achievement gap that they identified. Similarly to the Washington teachers, the Lincoln teachers experienced greater difficulty with addressing the issue that they identified.

The main barriers to success at Lincoln were consistent with those at Washington. The first barrier related to the time that needed to be spent on the assessments. Some teachers doubted that the initial time investment would pay off down the road. As with the Washington teachers, they were unwilling to give up content and feared that the time devoted to common formative assessments would necessitate a reduction in content.

The fear of misusing the formative assessments for evaluation was also seen at Lincoln. The primary evaluation-related issue at Lincoln involved the use of formative

assessment data by administrators. As discussed above, the school administrators touted this use of the data. Also, one teacher falsified results to appear more favorably in the eyes of the administration. Unlike at Washington, the Lincoln teachers did not use formative assessments for grading purposes.

The third barrier was the resistance to change on the part of certain teachers. As at Washington, the perceived lack of benefits of the preassessments reinforced the resistance. An additional factor was the lack of evidence of improvement on the part of the students receiving special education services, even in the area of graphing.

A barrier that was unique to Lincoln was the lack of familiarity with the tools (e.g., Excel) that would facilitate the teachers' understanding of the data. The department chair had received training in this area, but other teachers had not and were not comfortable with using technological tools to guide instruction.

Discussion

The first year of data-driven decision making in the science department analyzed in this study shared many common features. The differences between the experiences of the departments were relatively minor in comparison. Both departments experienced barriers to successful implementation that mirrored those discussed in the research literature in the field.

The most fundamental similarity between the two groups consisted of their relative strength in analyzing test score data to determine weak areas in the instruction program, combined with their relative weakness in knowing how to use test score, common formative assessment, and/or other data to rectify those weaknesses. One might analogize between the early experiences of these departments and the process that one undergoes when learning a language. Language learners usually receive oral and written language from others at higher levels than they

are able to produce speech and writing themselves (Basham & Fathman, 2008). Similarly, the teachers in these science departments were able to interpret weaknesses in students more effectively than they were able to craft solutions for change.

The two departments shared other strengths in common. Each had at least one faculty member who possessed the ability to produce visual data summaries that could be used by the other teachers to examine the student results. The teachers at both schools also dared to examine the student outcomes critically, in an effort to improve teaching and learning. Each school also reported enthusiasm for the process of analyzing the data, determining an area for improvement, and striving to reach a consensus regarding how to address the issue.

The barriers encountered by these schools related closely to those discussed in the research literature, as summarized earlier. The following paragraphs will place the barriers found at the schools in the context of the barriers found by other researchers.

Questions Regarding whether Data Can Guide Instruction

At first glance, neither science department appeared to have questions about whether data could guide instruction. The faculty at each school studied the data, identified a weakness, and created a plan to address the weakness. These practices are consistent with the belief that data can guide instruction.

On the other hand, certain teachers at each school expressed doubts regarding whether the initial time investment in learning how to practice data-driven decision making and the ongoing investment in practicing it were worthwhile. Even the teachers who were more willing to use data expressed concerns about narrowing the curriculum. One way to narrow the curriculum involves identifying “Power Standards,” which are the most important standards that are used across the curriculum,

and focusing instruction on them (Ainsworth, 2007). One teacher at Washington expressed her opinion that a concept such as Power Standards would not be effective in science because the state curriculum is too broad.

These concerns were amplified by the lack of tangible results in the first year of the process. Neither school found their preassessments to be a time-saving device because the preassessments generally showed that the students needed to learn the entire breadth of the curriculum. Also, neither school saw appreciable changes in student outcomes, not even from the focused effort on graphing at Lincoln.

Fears That Data Will Be Used to Evaluate Teachers

Teachers at both schools were very upfront about their concerns that common formative assessments would be used to evaluate teachers. The administration at Washington attempted to soothe the concern, while the Lincoln administrators validated it, perhaps inadvertently. The fear of data being used for evaluation was a very important barrier at these schools. In addition, some of the Washington teachers fell into the trap of using the data to evaluate students rather than guide instruction, which reduces the effectiveness of formative assessment.

Insufficient Capacity in the Management and Use of Data

As discussed above, each department possessed strengths and weaknesses related to data management and use. Each department chair had the skills needed to present the data to the other teachers in a format that was user-friendly. The insufficient capacity related mainly to knowledge and skill in the area of using their findings to improve results, though some teachers at Lincoln lacked the data management

capacity to implement formative assessment effectively.

An important issue related to this factor was the failure to make the most out of some of the formative assessment data. At Washington, this was evidenced by certain teachers being unable to break the habit of assigning a grade to every assessment, including formative assessments. This practice represents a misuse of formative assessment data. At Lincoln, perhaps the clearest example involved the teacher who used formative assessments to predict state assessment results rather than to improve outcomes. It should be noted that an assessment with such high predictive value would make for a valid formative assessment to guide improvement on the state test itself.

Recommendations

The following recommendations to reduce barriers are based on the findings of this study. An extremely important finding involved the relative strength of both sets of teachers in identifying weakness and/or achievement gaps and the relative weakness in resolving the identified issue(s). This finding is not surprising since the latter seems to be the more difficult endeavor. Given this finding, professional development in data-driven decision making should emphasize using assessment data to guide instruction. Such activities should include sophisticated examples of using data to guide instruction (such as what Washington could do to improve instruction when outcomes are good but not great across the board) in addition to the usual straightforward ones (such as if Teacher A is strong in teaching genetics, she should share her techniques with the rest of the faculty). Ideally, such professional development would involve analyzing the actual data from the teachers.

A second finding concerned the importance of having expertise in data use and analysis among a core group of faculty members. Part of the reason for the success Washington and

Lincoln had in identifying problem areas can be attributed to the department chairs having skills in this area. It is important to identify people on staff with such skills in order to draw on their expertise.

A third finding related to the issue of whether the time spent on formative assessments was worth it. The level of buy-in to data-driven decision making can be increased if teachers think that they will benefit from it. The implementation of the reforms in these departments was hindered by the lack of early, tangible evidence of improvement. If possible, the data team should attempt to guide the teachers at the school toward initial issues that can be solved in a straightforward manner, which is easier said than done in many circumstances.

A fourth recommendation involves having administrators make extra efforts in word and deed to show that formative assessment will not be used for teacher evaluation. Teachers tend to fear that any assessment will be used for judging their performance rather than (or in addition to) improving student outcomes. Remarks similar to the one made by the Lincoln principal about how administrators will access the data to track learning should be avoided in order to build the trust needed for teachers to lose their fear regarding the underlying purpose of formative assessment.

The early stages of adopting a new way of working involve accepting risks and overcoming obstacles. This paper describes some of the successes and barriers that were experienced by the science departments in these two schools. It is hoped that reading these experiences can assist others in overcoming the barriers to implementing data driven decision making in schools.

References

- Ainsworth, L. (2007). Common formative assessments: The centerpiece of an integrated standards-based assessment

- system. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 79-102). Bloomington, IN: Solution Tree.
- Basham, C., & Fathman, A. K. (2008). The latent speaker: Attaining adult fluency in an endangered language. *International Journal of Bilingual Education and Bilingualism*, 11(5), 577-597.
- Boudett, K. P., City, E. A., & Murnane, R. J. (2007). Introduction. In K. P. Boudett, E. A. City, & R. J. Murnane (Eds.), *Data wise: A step-by-step guide to using assessment results to improve teaching and learning* (pp. 1-8). Cambridge, MA: Harvard Education Press.
- Earl, L. M., & Katz, S. (2006). *Leading schools in a data-rich world: Harnessing data for school improvement*. Thousand Oaks, CA: Corwin Press.
- Glenn, W. & Creighton, T. (2008). University teaching that improves school leadership: Basic techniques to help principals use data effectively. *NCPEA Education Leadership Review* 9(2) 129-142.
- Kowalski, T. J., Lasley, II, T. J., & Mahoney, J. W. (2008). *Data-driven decisions and school leadership: Best practices for school improvement*. Boston: Pearson/Allyn and Bacon.
- Lachat, M. A. & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education For Students Placed at Risk*, 10(3), 333–349.
- Mandinach, E., Honey, M., Light, D., & Brunner, C. (2008). A conceptual framework for data-driven decision making. In E. Mandinach & M. Honey, Eds. *Data-driven school improvement: Linking data and learning* (pp. 13-31). New York: Teachers College Press.
- Wayman, J. C., & Stringfield, S. (2006). Data use for school improvement: School practices and research perspectives. *American Journal of Education*, 112(4), 463-468.
- Wayman, J. C., Stringfield, S., & Yakimowski, M. (2004). *Software enabling school improvement through the analysis of student data* (Report No. 67). Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed At Risk.
- Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage.